Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 24 (2022), Issue 6 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

XSL•FO
RenderX

Commentary

Short Paper

Viewpoints

Original Papers

XSL•FO
RenderX

XSL•FO
RenderX

XSL•FO
RenderX

XSL•FO
RenderX

Letters to the Editor

Corrigenda and Addendas

XSL•FO
RenderX

Review

Strategies for Implementing Occupational eMental Health Interventions: Scoping Review

Renaldo M Bernard[1], PhD; Claudia Toppo[2], MSc; Alberto Raggi[2], PhD; Marleen de Mul[3], PhD; Carlota de Miquel[4,5], MSc; Maria Teresa Pugliese[2], MSc; Christina M van der Feltz-Cornelis[6], MD, PhD; Ana Ortiz-Tallo[5,7], MSc; Luis Salvador-Carulla[8], PhD; Sue Lukersmith[8], PhD; Leona Hakkaart-van Roijen[3], PhD; Dorota Merecz-Kot[9], PhD; Kaja Staszewska[10], MA; Carla Sabariego[1,11,12], PhD

[1]Swiss Paraplegic Research, Nottwil, Switzerland

[2]Neurology, Public Health and Disability Unit, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milano, Italy

[3]Erasmus School of Health Policy & Management, Erasmus University, Rotterdam, Netherlands

[4]Research, Innovation and Teaching Unit, Parc Sanitari Sant Joan de Déu, Universitat de Barcelona, Sant Boi de Llobregat, Spain

[5]Instituto de Salud Carlos III, Centro de Investigación Biomédica en Red de Salud Mental, Madrid, Spain

[6]Department of Health Sciences, Hull York Medical School, University of York, York, United Kingdom

[7]Department of Psychiatry, School of Medicine, Universidad Autónoma de Madrid, Madrid, Spain

[8]Health Research Institute, Faculty of Health, University of Canberra, Canberra, Australia

[9]Institute of Psychology, University of Lodz, Lodz, Poland

[10]Department of Health and Work Psychology, Nofer Institute of Occupational Medicine, Lodz, Poland

[11]Department of Health Sciences and Medicine, University of Lucerne, Lucerne, Switzerland

[12]Center for Rehabilitation in Global Health Systems, World Health Organization Collaborating Center, University of Lucerne, Lucerne, Switzerland

Corresponding Author:
Renaldo M Bernard, PhD
Swiss Paraplegic Research
Guido A. Zäch-Strasse 4
Nottwil, 6207
Switzerland
Phone: 41 419396654
Email: renaldo.bernard@paraplegie.ch

Abstract

Background: The implementation of eMental health interventions, especially in the workplace, is a complex process. Therefore, learning from existing implementation strategies is imperative to ensure improvements in the adoption, development, and scalability of occupational eMental health (OeMH) interventions. However, the implementation strategies used for these interventions are often undocumented or inadequately reported and have not been systematically gathered across implementations in a way that can serve as a much-needed guide for researchers.

Objective: The objective of this scoping review was to identify implementation strategies relevant to the uptake of OeMH interventions that target employees and detail the associated barriers and facilitation measures.

Methods: A scoping review was conducted. The descriptive synthesis was guided by the RE-AIM (reach, effectiveness, adoption, implementation, and maintenance) framework and the Consolidated Framework for Implementation Research.

Results: A total of 31 of 32,916 (0.09%) publications reporting the use of the web-, smartphone-, telephone-, and email-based OeMH interventions were included. In all, 98 implementation strategies, 114 barriers, and 131 facilitators were identified. The synthesis of barriers and facilitators produced 19 facilitation measures that provide initial recommendations for improving the implementation of OeMH interventions.

Conclusions: This scoping review represents one of the first steps in a research agenda aimed at improving the implementation of OeMH interventions by systematically selecting, shaping, evaluating, and reporting implementation strategies. There is a dire need for improved reporting of implementation strategies and combining common implementation frameworks with more technology-centric implementation frameworks to fully capture the complexities of eHealth implementation. Future research should investigate a wider range of common implementation outcomes for OeMH interventions that also focus on a wider set of

XSL•FO
RenderX

common mental health problems in the workplace. This scoping review's findings can be critically leveraged by discerning decision-makers to improve the reach, effectiveness, adoption, implementation, and maintenance of OeMH interventions.

(J Med Internet Res 2022;24(6):e34479) doi:10.2196/34479

KEYWORDS

implementation; mobile health; mHealth; mental health; eMental health; occupational health; barriers; facilitators; scoping review; mobile phone

Introduction

Background

Mental health problems experienced by the working population are a global public health issue. Worldwide, more than 210 million people, representing 70% of those affected by common mental health disorders (eg, anxiety and mood disorders) are employed [1]. Several risk factors, including working conditions, workplace culture, and the nature of work, have been linked to occupational mental health [2-4]. Public health emergencies, such as the COVID-19 pandemic, are linked to specific stressors, including the threat of infection, social distancing measures, stigma, and job insecurity, which considerably increase the prevalence of mental health problems in the working population [5].

Occupational eMental health (OeMH) interventions significantly improve mental health in work settings [6]. OeMH interventions use information and communication technology, including internet- and web-based services, mobile apps, and wearable technologies, to deliver knowledge and services such as psychoeducation, workplace health promotion, psychological and medical treatment, and return to work assistance to employees [7,8]. OeMH interventions have the potential to be more available, accessible, and scalable than traditional interventions [9,10], especially in public health emergencies, leading to physical-distancing policies to contain the spread of threatening conditions such as COVID-19.

However, implementing OeMH interventions is a complex process characterized by unique challenges involving adherence to new and crude regulatory frameworks, interoperability and compatibility with existing systems and procedures, threats to employees, organizational privacy and security, and associated costs [11]. Newly introduced working arrangements in response to public health emergencies, such as the COVID-19 pandemic, could also compound existing implementation challenges and persist after the pandemic ends. Carefully developing and planning implementation strategies, which can be defined as a method or technique used to enhance the adoption, execution plan, and sustainability of an intervention [12], is therefore essential to guarantee the sustainable uptake of OeMH interventions by employers and employees.

Nonetheless, it is difficult to establish a best practice for the implementation of OeMH interventions. Implementation strategies are often inadequately documented and seldom evaluated and published [12,13], especially in comparison with studies on the effectiveness of interventions. Even when reported, implementation strategies have been discussed within a general context, and researchers have called for more tailored implementation strategies that focus on specific contexts [14], for instance health care [15]. Context encompasses the environment, broad setting, and circumstances (eg, systems and structures) in which an intervention is implemented and its associated characteristics [16]. It is a key component of several widely adopted implementation frameworks, as evident in the Consolidated Framework for Implementation Research (CFIR) [17]. Currently, those implementing new OeMH interventions are likely insufficiently informed about the procedure, strengths, and weaknesses of poorly documented implementation strategies, or uninformed about many potentially useful facilitators in this context. Furthermore, replicating positive results from similar implementations or overcoming barriers encountered in similar contexts would be challenging to achieve [18,19].

Objectives

Therefore, a compilation of possible implementation strategies for OeMH interventions is critical to fostering improvements in their uptake and can serve as a reference for identifying and overcoming likely barriers and informing the future development of best practices. The objective of this scoping review was to identify implementation strategies relevant to the uptake of OeMH interventions that target employees and detail the associated barriers and facilitation measures. This scoping review would achieve these objectives by mapping the existing literature on the implementation of OeMH interventions and identifying gaps for future research. This work was conducted under the EMPOWER (European Platform to Promote Well-being and Health in the Workplace) project, funded by the European Commission, which investigates the impact of an eMental health platform aimed at preventing common mental health problems and reducing psychological distress in the workplace [20]. It is also one of the series of review papers on different aspects of the knowledge base related to the development of the EMPOWER platform.

Methods

Overview

A scoping review was conducted to identify implementation strategies relevant to the implementation of OeMH interventions and to describe related barriers and associated facilitation measures. The scoping review is an established method for assessing and mapping the extent of evidence to address and inform practice in a topic area [21-24]. The review proceeded through five stages as developed by Arksey and O'Malley [23], extended by Levac et al [22], and further modified by Westphaln et al [25] to accommodate a team-based approach: (1) identifying the research question; (2) identifying relevant

XSL•FO
RenderX

studies; (3) selecting studies; (4) charting the data; and (5) collating, summarizing, and reporting the results. Accordingly, this scoping review provides an overview of the existing evidence without a formal assessment of the methodological quality. It is conducted and reported in accordance with the widely adopted PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [26] to help ensure a high level of methodological rigor and reporting quality.

Search Strategy

Electronic bibliographic databases, including MEDLINE, Scopus, CINAHL Complete, PsycINFO, and Web of Science Core Collection, were searched to find eligible peer-reviewed and gray literature. Search terms were based on concepts related to mental health, digital tools, the workplace, and implementation strategies (Multimedia Appendix 1). The MEDLINE search strategy (Multimedia Appendix 1) was adapted for other databases using relevant syntax and keywords in consultation with all coauthors who are also experienced researchers in the area. Hand searching of the reference lists of included articles was also completed for further relevant literature not identified during the search of databases. Members of the EMPOWER Consortium (ie, mental health researchers, clinicians, and experts focusing on well-being in the workplace) were also requested to suggest potentially eligible references via email.

Eligibility Criteria

Publications were eligible for inclusion if they described implementation strategies (ie, according to Proctor et al [12]) or related barriers or facilitation measures relevant to the uptake of OeMH interventions targeting employees. For example, all other eligibility criteria being met, approaches with the following characteristics would be considered: aim to introduce and encourage continued use of an intervention; prescribe actions in support of the intervention (eg, adaptations, fiscal strategies, and testing); and ensure that interventions can deliver intended benefits to the relevant organization over time, for instance, creating routine organizational policies or best practices. OeMH interventions are broadly defined here as mental health information and services delivered by information and communication technologies to employees [7,8]. This definition is consistent with the definition of eHealth [27,28], as well as the broader term digital health [27]. Studies with employed participants aged ≥18 years, that were written in English. and published between January 2010 and May 2021 were considered. Primary research studies, systematic reviews, books, and gray literature (eg, conference proceedings, theses, government documents, and professional publications) were considered. Gray literature, such as commentaries, letters to editors, and editorials, were excluded.

Eligibility Assessment

A total of 10 researchers (AO-T, AR, CdM, CT, CMvdFC, DM-K, KS, MdM, MTP, and RMB), including psychologists, health scientists, and health economists, were involved in screening. To ensure consistency across researchers, they attended a web-based training workshop to practice the skills needed to reliably execute screening using the web and an app-based service Rayyan, Qatar Computing Research Institute [29]. A training set of 100 publications was screened by all workshop attendees. Screening decisions (ie, include, maybe, or exclude) were reviewed and discussed to clarify any misunderstandings and identify difficulties using Rayyan QCRI. Instructions not to use the natural language processing–, artificial intelligence (AI)-, and machine learning–based features offered in Rayyan QCRI as well as tips to overcome minor usability shortcomings were given. Screeners were randomly assigned a screening set, and a screener performed a second screening of 20% of titles, abstracts, and full texts, and 100% of the publications that received a *maybe* screening decision. All screenings were conducted independently to reduce the likelihood of reviewer bias [30] and inconsistencies in screening decisions were resolved in reconciliation meetings.

Data Extraction and Synthesis of Results

In all, 5 researchers (AR, CdM, CT, MdM, and RMB), including psychologists and health scientists, of the 10 (50%) screeners, were involved in data extraction and attended a web-based training workshop focused on developing consistency across researchers by practicing the skills needed to reliably execute data extraction using a web-based data extraction form. The form was reviewed and improved for clarity regarding the questions asked, user friendliness, and efficiency of data entry. For instance, it was clarified that single-component implementation strategies were to be extracted, and any bundling of strategies (ie, multifaceted strategies) in publications to address a goal were to be noted. Each researcher was randomly assigned an equal number of included records, and a researcher reviewed the extracted data for all the included publications.

A descriptive synthesis was performed, where identified implementation strategies, barriers, and facilitators were collated and later summarized. The synthesis was conducted by 3 (CT, MdM, and RMB, ie, psychologists and health scientists) of the 6 (50%) researchers involved in data extraction and guided by the RE-AIM (reach, effectiveness, adoption, implementation, and maintenance) framework [31-33] and the CFIR [17] and further informed by the Expert Recommendations for Implementing Change [34]. RE-AIM and CFIR were chosen as they are widely used frameworks in implementation research (IR) [35] and were deemed by the authors to be the most comprehensive of the recently reviewed implementation frameworks [35] and most applicable to our objectives. The RE-AIM was originally developed as a framework for reporting findings regarding health promotion and disease management interventions in various settings. RE-AIM is used here to highlight essential strategy components with respect to its five steps: reach—the number of people who are willing to participate in a given initiative; effectiveness—the impact of an intervention on important outcomes (eg, individualistic and economic); adoption—the number of people or organizations who are willing to initiate and deliver an intervention; implementation—fidelity of delivery for the intervention including adaptations, costs, and consistency of delivery; and maintenance—sustained delivery and effects of an intervention after the associated initiative has ended. The CFIR unifies implementation theories to help build a robust implementation

XSL•FO
RenderX

knowledge base across a wide range of studies, settings, contexts, and processes. The CFIR was used to provide a comprehensive view of multiple implementation contexts in which factors that might influence intervention implementation and effectiveness could be well detailed. Both frameworks determined the data for extraction: key publication characteristics, strategy definitions, key strategy implementation tasks, implementation processes, barriers and facilitators to strategy implementation, and any other data that holistically captured the complex and multilevel nature of strategy implementation were considered for data collection. Further synthesis of the identified barriers and facilitators produced recommendations for each relevant CFIR construct to improve the implementation of OeMH interventions.

Results

A total of 31 publications were included in this scoping review (Table 1). Figure 1 details the methodological process followed, and a detailed itemization of the presented findings is provided in Multimedia Appendices 2 and 3.

XSL•FO
RenderX

Table 1. Characteristics of included publications and interventions.

Citation and year of publication	Study aim and methods (n)	Country of implementation, industry, and participating organizations (n)	Intervention name, aim, and target conditions	Digital technologies used
[36], 2020	To develop, implement, and evaluate the intervention; survey (503), interviews (19), and focus groups (32)	United Kingdom; human health and social work activities; 7	Healthier Outcomes at Work Social Work Project; improve and manage; workplace stress and mental well-being	Smartphone app
[37],[a] 2020	To describe the intervention's implementation; protocol—pilot randomized controlled trial (106)	China; human health and social work activities; 1	Step-by-Step F; improve; depressive symptoms and anxiety symptoms	Web-based and smartphone app
[38],[a] 2020	To describe the evaluation of the intervention's implementation; protocol—focus groups (N/R[b])	Germany; agriculture, forestry, and fishing; N/A[c]	With us in balance; prevent; stress-related disorders, anxiety disorders, mood disorders, substance-related and addictive disorders, insomnia, and chronic pain	Web-based and telephone
[39], 2020	To examine perspectives on the role and legitimacy of the intervention; interviews (32) and focus group (14)	Sweden; N/R; N/A	mWorks; support; common mental disorders	Smartphone app
[40],[a] 2020	To conduct preliminary evaluation of the intervention; pilot—usability study (81)	Australia; N/S[d]; N/R	Anchored app; assess, improve, and monitor; depression, workplace stress, and mental well-being	Smartphone app
[41],[a] 2020	To rapidly develop and evaluate the intervention; stakeholder consultation groups (97), peer review panel (10), and intervention fidelity and implementation testing (55)	United Kingdom; human health and social work activities; N/R	Psychological Well-being in Healthcare Workers: Mitigating the Impacts of COVID-19; support and manage; workplace stress and mental well-being	Web-based
[42], 2019	To evaluate the feasibility, outcome, and acceptability of the intervention; proof-of-concept—survey (33)	United Kingdom; public administration and defense and compulsory social security; 2	Self-confidence webinar program; improve; mood disorders and depression	Web-based
[43], 2019	To evaluate engagement with the intervention; survey (149)	United States; public administration and defense and compulsory social security; 20	Stress Reduction Training for 9-1-1 Telecommunicators; improve and promote; workplace stress	Web-based
[44], 2019	To conduct formative evaluation of the intervention; interviews (24)	New Zealand; public administration and defense and compulsory social security; N/R	N/R; improve; stigma and discrimination	Web-based
[45], 2018	To evaluate adherence to the intervention; randomized controlled study (563)	Sweden; education; 21	N/R; improve and promote; workplace stress, occupational health, and sleep quality	Web-based
[46], 2018	To evaluate the helpfulness of the intervention; web-based survey (22) and focus groups (2)	United States; human health and social work activities; 1	Paving the Path to Mindfulness Website; improve; burnout and workplace stress	Web-based
[47], 2018	To evaluate acceptance and barriers to the uptake of OeMH[e] interventions; survey (3294)	N/A; N/A; N/A	N/A; manage; work-related distress	N/S
[48],[a] 2018	To evaluate the implementation strategy used; controlled trial (221)	The Netherlands; human health and social work activities; 1	Stress Prevention@Work; improve and prevent; workplace stress	Web-based
[49],[a] 2018	To evaluate the effectiveness of the implementation strategy used; follow-up controlled trial (252)	The Netherlands; human health and social work activities; 1	Stress Prevention@Work (SP@W); assess, improve, and prevent; workplace stress	Web-based
[50],[a] 2018	To identify key correlates of intention to use OeMH interventions; survey (1364)	China; human health and social work activities; N/A	N/A; N/A; mental health conditions	Web-based and smartphone app

XSL•FO
RenderX

Citation and year of publication	Study aim and methods (n)	Country of implementation, industry, and participating organizations (n)	Intervention name, aim, and target conditions	Digital technologies used
[51],[a] 2018	To evaluate use of OeMH; log data and survey (1284)	Sweden; N/R; 6	N/R; improve, monitor, promote, and support; workplace stress and mental well-being	Web-based
[52], 2018	To develop and pilot-test the usability, acceptability, feasibility, and preliminary effectiveness of the intervention; prototype testing (21) and effectiveness and feasibility pilot study (84)	Australia; agriculture, forestry and fishing, manufacturing, and logistics; 3	HeadGear; improve; depressive symptoms	Smartphone app
[53],[a] 2018	To identify facilitators and barriers to engagement with OeMH interventions; interviews (18)	United Kingdom; information and communication, public administration and defense, education, and other service activities; 6	WorkGuru; improve; workplace stress	Web-based
[54],[a] 2017	To conduct process evaluation of the intervention; survey, log data, interviews, and observations (132)	The Netherlands; N/R; 2	eHealth module embedded in collaborative occupational health care; improve and monitor; mental well-being and return to work	Web-based
[55], 2017	To compare engagement with(out) a discussion group; pilot—3-arm randomized controlled trial (84)	United Kingdom; information and communication, public administration and defense, compulsory social security, education, and third sector organization; 6	WorkGuru; educate, improve, and monitor; workplace stress and nonworkplace stress	Web-based
[56], 2016	To investigate the influence of guidance formats on adherence of the intervention; pooled data from randomized controlled trials (395)	Germany; N/R; N/R	GET.ON Stress; improve and manage; workplace stress	Smartphone app
[57], 2016	To investigate men's preferences for OeMH interventions' design features; cross-sectional survey (841)	Canada; N/A; N/A	N/A; N/A; workplace stress and major depression	N/A
[58], 2016	To describe the development, implementation, and outcomes of; survey (1333)	United States; human health and social work activities; 1	Sleep Smart; improve and promote; poor sleep health	Email
[8],[a] 2016	To describe approaches to and perspectives on OeMH interventions; N/A (N/A)	N/A; N/S; N/A	N/A; N/A; N/A	N/S
[59], 2016	To evaluate the potential effectiveness of the intervention and the effect of an online facilitated discussion group on engagement; protocol—3-arm randomized controlled trial (90)	United Kingdom; N/R; N/A	WorkGuru; educate, improve, and monitor; Workplace stress and nonworkplace stress	Web-based
[60], 2015	To describe the development the intervention; individual (34) and focus group (18) feedback sessions	United States; public administration and defense, compulsory social security, and human health and social work activities; N/R	Coming Home and Moving Forward; improve; stress-related disorders and substance-related and addictive disorders	Web-based
[61], 2014	To investigate users' views on two different technologies for an OeMH intervention; survey within randomized controlled trial (637)	United Kingdom; transportation and storage, information and communication, and human health and social work activities; 3	Mood GYM; improve; mood disorders	Web-based
[62],[a] 2014	To contrast the role of differing managerial levels during the implementation of an OeMH; interviews (29)	Sweden; information and communication; public administration and defense; compulsory social security; education; and arts, entertainment, and recreation; 9	N/R; assess, improve, monitor, and promote; mental well-being	Web-based

XSL•FO
RenderX

Citation and year of publication	Study aim and methods (n)	Country of implementation, industry, and participating organizations (n)	Intervention name, aim, and target conditions	Digital technologies used
[63],[a] 2014	To assess the feasibility of the intervention and explore barriers and /facilitators for the implementation of the intervention; process evaluation alongside a randomized controlled trial (116)	The Netherlands; financial and insurance activities; professional, scientific, and technical activities; public administration and defense; compulsory social security; and education; 6	Happy Work; improve and prevent; depressive symptoms	Web-based
[64],[a] 2013	To describe the development and implementation of the intervention; N/A (N/A)	International; N/R; N/R	HealthWatch; manage, prevent, and promote; mental well-being	Web-based
[65], 2010	To investigate determinants of high use of the intervention; randomized controlled intervention (303)	Sweden; information and communication and arts, entertainment, and recreation; N/R	N/R; assess, monitor, and promote; workplace stress	Web-based

[a]Focused on implementation.

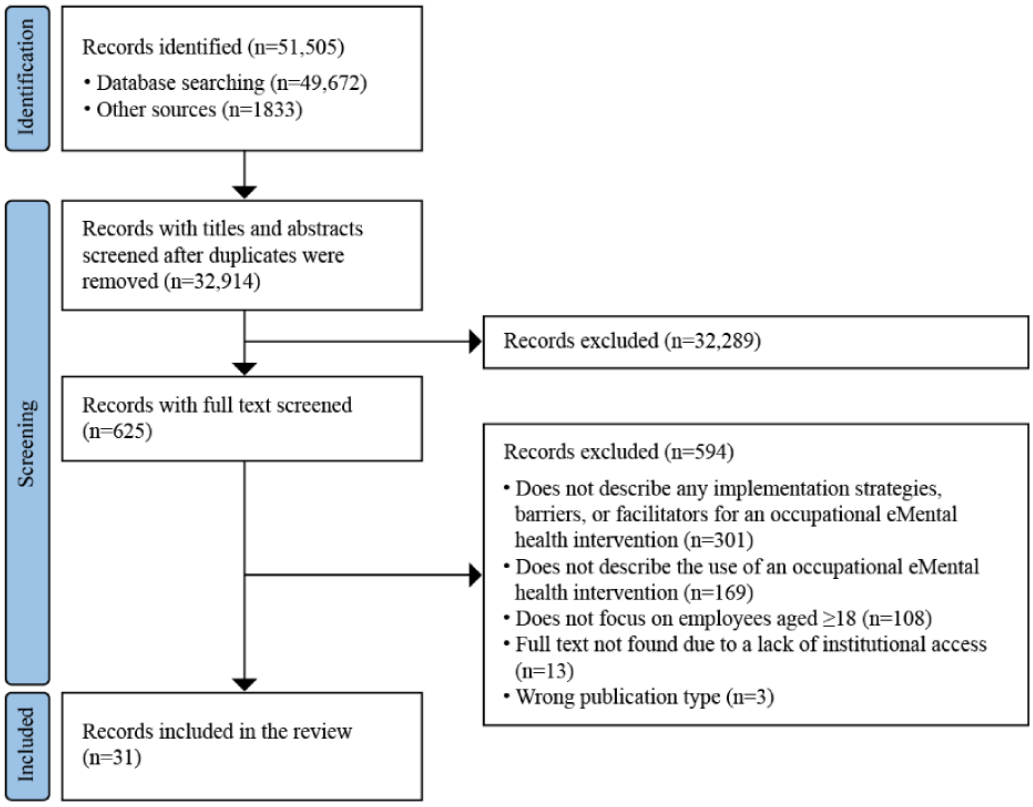
[b]N/R: not reported.

[c]N/A: not applicable.

[d]N/S: not specified.

[e]OeMH: occupational eMental health.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the review search, selection, and inclusion process.



Publication Characteristics

The 31 included publications comprised 28 journal articles [36-45,47,48,50-57,59-63], 2 book chapters [8,64], and a doctoral dissertation [46] (Tables 1 and 2). Most (25/31, 81%) [8,36-47,49-59] of the included articles were published between 2015 and 2021 and were mainly primary studies (23/31, 74%) [36,39,41-51,53,54,56-58,60-63,65]. Of the 31 included publications on OeMH interventions, 14 (45%) focused on their implementation [8,37,38,40,41,49-51,53,54,62-64] and 17 (55%) did not focus on their implementation but had results or noted implications related to their implementation [36,39,42-47,52,55-61,65].

XSL•FO
RenderX

Table 2. Summary of study characteristics (N=31).

Characteristics and citations	Frequency, n (%)
Publication type	
Book chapter [8,64]	2 (6)
Doctoral dissertation [46]	1 (3)
Journal article [36-45,47,48,50-57,59-63]	28 (90)
Publication year	
2010 [65]	1 (3)
2013 [64]	1 (3)
2014 [61-63]	3 (10)
2015 [60]	1 (3)
2016 [8,56-59]	5 (16)
2017 [54,55]	2 (6)
2018 [45-53]	9 (29)
2019 [42-44]	3 (10)
2020 [36-41]	6 (19)
Study type	
Narrative literature review [8,65]	2 (6)
Pilot [40,52,55]	3 (10)
Primary study [36,39,41-51,53,54,56-58,60-63,65]	23 (74)
Protocol [37,38,59]	3 (10)

Intervention Characteristics

A total of 24 interventions were reported in 27 studies [36-46,49,51-56,58-65] (Table 3). These interventions were largely web-based (n=16, 67%%) [41-46,48,49,51,53-55,60-65] and most aimed to improve (n=19, 79%) [36,37,40,42-46,49,51-56,58-63] and, to a lesser extent, educate users about mental health problems. Most interventions have focused on stress-related disorders and symptoms (n=17, 71%) [8,36-47,49-51,53-63,65], but a wide range of mental health problems (eg, burnout, anxiety disorders, and substance-related disorders) have also been covered to some extent. Where reported [36,38,41,43-46,49,51-53,55,56,58-65], these interventions (n=19, 79%) largely targeted employees in professional occupations (eg, teachers and physicians). Most of these interventions were made available in specific countries, mainly in Europe (n=15, 63%%) [36,38,39,41,42,45,49,51,53-56,59,61-65], except for one that was available internationally (n=1, 4%) [64]. Standardized information about these 24 interventions, including year of launch, language, number of employees and employers interested in and who adopted the app, organizational size, and internal policies, was not clearly reported where relevant and could not be accurately extracted in detail.

XSL•FO
RenderX

Table 3. Summary of intervention characteristics (N=24).

Characteristics and citations	Frequency, n (%)
Technology	
Smartphone [36,39,40,52,56]	5 (21)
Web [41-46,48,49,51,53-55,60-65]	16 (67)
Web and smartphone [37]	1 (4)
Web and telephone [38]	1 (4)
Email [58]	1 (4)
Aim	
Assess [40,49,62,65]	4 (17)
Educate [55,59]	1 (4)
Improve [36,37,40,42-46,49,51-56,58-63]	19 (79)
Manage [36,41,47,56,64]	5 (21)
Monitor [40,51,54,55,59,62,65]	6 (25)
Prevent [38,49,63,64]	4 (17)
Promote [43,45,51,58,62,64,65]	7 (29)
Support [39,41,51]	3 (13)
Target mental health problem	
Anxiety disorders and symptoms [37,38]	2 (8)
Burnout [46]	1 (4)
Chronic pain [38]	1 (4)
Common mental disorders [39]	1 (4)
Mood disorders and symptoms [37,38,40,42,52,57,61,63]	8 (33)
Return to work [54]	1 (4)
Sleep problems [38,45,58]	3 (13)
Substance-related and addictive disorders [38,60]	2 (8)
Stigma and discrimination [44]	1 (4)
Stress-related disorders and symptoms [8,36-47,49-51,53-63,65]	17 (71)
Well-being problems [36,40,41,51,54,62,64]	7 (29)
Country of implementation	
Australia [40,52]	2 (8)
Canada [43]	1 (4)
China [37]	1 (4)
Germany [38,56]	2 (8)
International [64]	1 (4)
The Netherlands [49,54,63]	3 (13)
New Zealand [44]	1 (4)
Sweden [39,45,51,62,65]	5 (21)
United Kingdom [36,41,42,53,55,59,61]	5 (21)
United States [43,46,58,60]	4 (17)
Target occupational groups	
Armed forces occupations [44,60]	2 (8)
Clerical support worker [45,55,61,62]	4 (17)
Elementary occupations (eg, cleaners and laborers) [45]	1 (4)

XSL•FO
RenderX

Characteristics and citations	Frequency, n (%)
Managers (eg, chief executive officer) [41,55,61]	3 (13)
Not reported [37,39,40,42,54]	5 (21)
Plant and machine operators and assemblers [61]	1 (4)
Professionals (eg, teachers and physicians) [41,43,45,46,49,55,58,61-63]	10 (42)
Service and sales workers [45,61,65]	3 (13)
Social workers (ie, specifically child and family social workers) [36]	1 (4)
Skilled agricultural, forestry, and fishery workers [38,61]	2 (8)
Technicians and associate professionals [36,41,55,61,62,65]	6 (25)

Implementation Strategies

Overview

Overall, 98 examples of implementation strategies were identified (Multimedia Appendix 2). Table 4 categorizes these strategies into 17 discrete implementation strategies and maps them onto relevant RE-AIM domains based on the perceived intent of the themes. Each discrete implementation strategy is reported with the percentage and absolute number of defining strategy examples in relation to the 98 examples. Although evaluating effectiveness strategies is beyond the scope of this review, relevant effectiveness data could not be presented as they were largely absent or incomplete.

A total of 36 examples of implementation strategies were extracted from publications that focused on the implementation of OeMH interventions, and 62 from publications that reported results or noted implications related to their implementation. There were no notable differences other than the larger number of examples extracted from the latter group so strategy examples would not be reported separately. Most strategy examples were organized under implementation (61/98, 62%), followed by reach (27/98, 28%), effectiveness (19/98), adoption (17/98, 17%), and maintenance (8/98, 8%). A couple of strategy examples were organized into multiple domains (6/98, 6%). The following sections provide a descriptive summary of the strategy examples categorized in each RE-AIM domain.

XSL•FO
RenderX

Table 4. Discrete implementation strategies mapped to relevant RE-AIM (reach, effectiveness, adoption, implementation, and maintenance) domains (N=98).

Discrete implementation strategies—proportion of strategy examples; n (%)	Example strategy	Relevant RE-AIM domains
Develop and organize implementation quality monitoring systems and act on insights in a timely manner where feasible; 17 (17)	Improve maintenance and adherence through the timely presentation of findings from monthly user feedback surveys where after follow-up actions can be immediately applied to the intervention [64]	Effectiveness, implementation, and maintenance
Assess for readiness and tailor strategies to address identified barriers and benefit from facilitators; 13 (13)	Filled knowledge gaps surrounding the effectiveness of eMental health interventions in the workplace by conducting systematic reviews on relevant topics [52]	Reach, effectiveness, implementation, and maintenance
Use mass media to increase reach; 9 (9)	Users were recruited by sharing information about the intervention through advertisements distributed via email and the organizations' intranet and magazine [59]	Reach
Capture local knowledge from implementation sites and involve users early in the implementation and intervention development effort; 8 (8)	A consultation process was carried out with users, clinical psychologists, psychiatrists, information technology professionals, and design and user experience specialists to ensure the app's content and design appealed to a broad range of workers from different industries [40]	Reach, effectiveness, adoption, implementation, and maintenance
Promote adaptability in the intervention to meet local needs without compromising fidelity; 8 (8)	Interventions were improved and adapted to each participating organization based on user feedback [36]	Reach, effectiveness, adoption, implementation, and maintenance
Send reminders; 7 (7)	Automatic email reminders were sent based on user-determined intervals and user inactivity [45]	Implementation
Provide support to users during the intervention; 6 (6)	Users were able to contact the intervention coach at any time to ask for feedback, additional help, or advice and the coach would respond within 24 hours [59]	Implementation
Conduct educational meetings; 5 (5)	Senior and middle management–led introductory seminars with employees that aimed to explain the intervention, secure acceptance, provide answers to questions, and inspire their participation [64]	Reach, adoption, and implementation
Provide incentives; 5 (5)	Users received a certificate of completion and the training was recognized as continuing education toward the renewal of their professional certification [43]	Reach and implementation
Identify and prepare organizational champions who will dedicate themselves to supporting, marketing, and driving the implementation; 4 (4)	Identification of champions at the implementation site facilitated organizational and employee buy-in [46]	Reach, adoption, implementation, and maintenance
Involve senior management; 4 (4)	The program was developed as a quality improvement project by the hospital and all research procedures (ie, retrospectively reviewing these outcomes) were approved by the institutional review board at the hospital [58]	Reach, adoption, implementation, and maintenance
Provide opportunities for users to obtain feedback on progress; 4 (4)	Participants received immediate and automatic tailored feedback and could monitor their own responses and trends over time [45]	Implementation
Stage implementation scale-up; 4 (4)	Conducted a pilot study aimed at assessing the usability, feasibility, acceptability, and preliminary effects of an app-based intervention designed to target depressive symptoms in a stressed working population [55]	Effectiveness and implementation
Customize recruitment activities to enhance reach; 3 (3)	When recruitment efforts did not attract enough participants, executives with the largest workforces in the region and industry were contacted directly via telephone and offered enrollment [45]	Reach and adoption
Develop and distribute educational materials; 3 (3)	All participants who returned the consent form received an email welcoming them to the study and explaining how to log in and use their personal webpage for the stress management program [65]	Reach, adoption, and implementation
Provide immediate opportunities to demonstrate commitment; 3 (3)	Management representatives were offered spots to enroll their organizations immediately after educational meetings about the intervention or to enroll at a later time [45]	Reach and adoption

XSL•FO
RenderX

Discrete implementation strategies—proportion of strategy examples; n (%)	Example strategy	Relevant RE-AIM domains
Use advisory boards and workgroups to provide input and advice on implementation and improvements; 3 (3)	Systematic feedback was sought from researchers, expert clinicians, and veterans on the program and its content [60]	Implementation

Reach

Mass media services were mainly used to increase reach (9/27, 33%). Examples include email [36,43,46,52,58,59,63], industry publications [43,59], targeted web-based advertisements (eg, Facebook) [40,52], and the organizations' intranet [59,61,63]. The provision of attractive incentives for participation (5/27, 19%) included monetary remuneration [60], vouchers [40], points for employee reward schemes [58], educational credits for professional certifications [43], and additional medical benefits [65]. Other strategies included engaging potential users through educational meetings [62] and materials [65] (2/27, 7%), employees tasked with the responsibility of supporting the implementation of the intervention [46,61] (2/27, 7%), and well-timed opportunities to commit [64,65] (2/27, 7%). Local barriers to increasing reach among target users were identified through consultations with implementation sites, eligible users, and literature [52,61,64] (3/27, 11%), and recruitment activities were later modified to avoid or overcome these barriers where possible [63] (4/27, 15%).

Effectiveness

Strategies to improve the effectiveness of the intervention mainly relied on insights obtained from a diverse group of professionals in relevant fields, representatives of implementation sites, target users, and intervention use data (12/19, 63%). This insight was captured through stakeholder consultations [41], steering group interviews and focus groups with target users [36], peer review panels [41], and user experience research [36,37,41,52,53,63] throughout the implementation process. Several strategies adopted an incremental approach to implementation [52,55,59] (3/19, 16%), fostered adaptability in the intervention to adequately meet the local needs [60] (2/19, 11%), or implemented measures to avoid or mitigate identified barriers that could negatively impact the effectiveness of the particular intervention [49,52] (2/19, 11%).

Adoption

Sharing and discussing details about the proposed intervention with decision-makers was the most commonly used adoption strategy. This involved conducting educational meetings [45,62,64,65] (4/17, 24%) and distributing educational materials about the intervention [45] (1/17, 6%). Engaging senior management and others from the organization to identify necessary adaptations for intervention to succeed in the organization was also common. These strategies involved organizational stakeholders early in the intervention development process [49,64] (2/17, 12%) to adapt the intervention to meet special organizational needs without compromising fidelity [64] (1/17, 6%) and address other identified barriers [39] (1/17, 6%). Some strategies also identified staff members who could dedicate themselves to supporting, marketing, and driving the implementation within the organization, as this was expected to increase the likelihood of success [46,49] (2/17, 12%). The provision of immediate opportunities for decision-makers to confirm their commitment to adopt the intervention was also used [45] (1/17, 6%).

Implementation

Implementation strategies focused on adapting interventions and customizing the implementation process to implementation settings, monitoring the consistency of delivery, and providing various forms of support as needed. Implementers underwent training, subscribed to a common protocol, and had their work reviewed to help ensure fidelity. Some implementation strategies were continuously monitored using both qualitative and quantitative methods, including surveys, implementation reviews, process evaluations, and other similar methods [36-38,41,49,54,56,61,63,64] (16/61, 26%). A diverse group of stakeholders were involved in the assessments across the included studies. These assessments focused on measuring effectiveness, acceptability, and engagement [36-38,41,49,54,56,61,63,64] (16/61, 26%). Findings were regularly applied quickly to overcome identified barriers and improve ongoing implementation processes [36,60,63,64] (10/61, 16%). Some support options included a reminder feature (7/61, 11%) where users could set their own reminder notifications [46] and be notified when their participation level was too low [45,56,63] or when new updates became available [43,46].

Maintenance

Maintenance strategies involved changes at the organizational level, where accommodating work conditions [43,55] (2/8, 25%) and support staff [58] (1/8, 13%) were sometimes arranged. Embedding interventions within existing employee programs was also expected to help sustain the use of the intervention [58] (1/8, 13%). Special monitoring measures (eg, postintervention acceptability surveys and opportunities for monthly user feedback) were also established to provide insight into how benefits to users could be sustained after the initiative had officially ended [37,64] (2/8, 25%).

Barriers and Facilitators

Overview

The included publications reported 114 barriers and 131 facilitation measures (Multimedia Appendix 3), and 28 barriers were accompanied by facilitation measures. There were no notable differences between barriers and facilitators extracted from publications that focused on the implementation of OeMH interventions (108/217, 49.8%) or that reported results or noted implications related to their implementation (109/217, 50.2%) so these will be reported together. Examples of barriers and facilitators organized by the relevant CFIR domains and associated constructs are provided in the corresponding tables. Most of the 217 identified barriers and facilitation measures

XSL•FO
RenderX

were related to key attributes of interventions that influence successful implementation (103/217, 47.5%), followed by the inner setting of the organization (87/217, 40.1%), individual characteristics of target users (25/217, 11.5%), and the outer setting of the organization (2/217, 0.9%). The highest number of barriers were categorized under the inner setting (54/114, 47.4%), followed by intervention characteristics (35/114, 30.7%), individual characteristics of target users (22/114, 19.3%), and the outer setting of the organization (2/114, 1.8%) domains. The highest number of facilitators were categorized under intervention characteristics (77/131, 58.8%), followed by inner setting (44/131, 33.6%), individual characteristics of target users (9/131, 6.9%), and outer setting of the organization (1/131, 0.8%) domains.

Intervention Characteristics

Numerous barriers and facilitators were identified regarding how the interventions were bundled, presented, and assembled (ie, design quality and packaging) (Table 5). Participants from several studies considered web-based platforms to be an impersonal medium (eg, no face-to-face contact or human interaction) [53,57,61], and some saw its use as inappropriate for helping with sensitive topics such as mental health problems [44]. Several usability issues (eg, poor accessibility, technical issues, unclear navigational elements and user interface, and overly effortful tasks) have also emerged as barriers [40,52,53,60,61,64]. Accordingly, ensuring good usability [8,39,40,53,57,64] and considering individual factors (eg, high impulsivity benefits from continuous motivational components) [39,45,57,65] in the design were also often reported as facilitators.

The stakeholders' perceptions of the evidence supporting the effectiveness of the proposed occupational mental interventions were influenced by several factors. The barriers included between-group contamination due to limited randomization at the individual level, unrepresentativeness of samples used for the general workforce, use of new or adapted measures with low reliability [55], and type 1 errors [58]. Identified facilitators focused on the including diverse samples (eg, including underrepresented industries and occupations) [55], collecting comparable demographic data [55], including comprehensive engagement measures [55], presenting interventions based on credible information highly relevant to target employees [57], using control conditions when evaluating effectiveness [42,58], providing evidence from similar interventions that demonstrate effectiveness [65], and conducting comprehensive and ongoing process evaluations to inform implementation [51,63].

Table 5. Examples of barriers and facilitators organized under the intervention characteristics Consolidated Framework for Implementation Research domain (N=217).

Relevant associated construct—proportion of barriers and facilitators; n (%) and brief description	Example of identified barriers	Example of identified facilitators
Evidence strength and quality; 15 (6.9); stakeholders' perceptions of the quality and validity of evidence supporting the belief that the intervention will have desired outcomes	Using newly created or adapted measures demonstrating low reliability negatively impacts the strength of findings [55]	Providing evidence from other programs and interventions could be a strategy (oral presentations or reading materials) to demonstrate likely effectiveness [65]
Relative advantage; 2 (0.9); stakeholders' perception of the advantage of implementing the intervention versus an alternative solution	Possible low motivation from employers and organization in their employees return to work as they came from small- to medium-sized companies that had insurance for the costs of sickness absence [54]	The lack of a previous existing intervention for well-being in the organization, except for the intranet, which was difficult to use, so the app resulted to be a huge advantage for employees [36]
Adaptability; 4 (1.8); the degree to which an intervention can be adapted, tailored, refined, or reinvented to meet local needs	Materials presented in a modular format that had to be completed start to finish in a single sitting or in a set order [61]	Possibility to use the program at their own pace [60]
Design quality and packaging; 80 (36.9); perceived excellence in how the intervention is bundled, presented, and assembled	Usability was affected by unclear navigational elements and user interface [40]	Improving usability based on participant and expert feedback [40]

Outer Setting

Strict external policies and failure of interventions to meet patient needs erected several barriers to the implementation of OeMH interventions (Table 6). For example, strict legislation and policies regarding privacy and confidentiality were highlighted as potential reasons for the reduced adoption of interventions based on innovative technologies [39]. Moreover, failure to maintain employees' confidentiality during these programs was believed to discourage the use of interventions for fear of being vulnerable to privacy breaches by employers [59]. The sole facilitation measure identified for this CFIR domain also addresses this point by urging implementers to find ways to maintain employee confidentiality [59].

XSL•FO
RenderX

Table 6. Examples of barriers and facilitators organized under the outer setting Consolidated Framework for Implementation Research domain (N=217).

Relevant associated construct—proportion of barriers and facilitators; n (%) and brief description	Example of identified barriers	Example of identified facilitators
External policy and incentives; 1 (0.5); a broad construct that includes external strategies to spread interventions including policy and regulations (governmental or other central entity), external mandates, recommendations and guidelines, pay for performance, collaboratives, and public or benchmark reporting	The surrounding legislation and policy regulation of privacy and confidentiality may make it difficult to use innovative technology [39]	—[a]
Patient needs and resources; 1 (0.5); the extent to which patient needs, as well as barriers and facilitators to meet those needs, are accurately known and prioritized by the organization	Reluctancy of the potential participants in participating for fear of demonstrating vulnerability [45]	Maintaining confidentiality between employee and employer [45]

[a]No facilitator reported.

Inner Setting

Many publications have identified the lack of resources dedicated to implementation as a major barrier (Table 7). For example, there is a lack of time for employees to use the intervention [40,48,49,52,53,59,61], funds to meet additional costs [39], unreliable systems that lead to data loss [58], inflexible participation times [42], lack of workspaces to avoid office distractions and private spaces [53] when completing interventions [61], low technology (eg, computers and email) adoption by the organization [64], little support from the app or implementor [54], and insufficient resources for piloting [62]. Some interventions were also inadequately adjusted to organizational processes [36,49,54,63] and insufficiently tailored to the work situation and culture [42,48,54,58,63]. Organizational restructuring has also been identified as a barrier to successful implementation and should be considered during implementation planning [42,48,63].

Several facilitators have also been identified. For example, it was recommended for employers to arrange dedicated time for employees to participate in the intervention [59]; to allow employees flexibility regarding the time, place, and pace when completing the intervention [53]; to offer an option for employees to use the intervention in a private workspace [53]; to provide recordings of any live sessions with feedback options [42]; and to encourage employee access to or ownership of technology (eg, smartphone) in use [50]. Intervention creators can also support employers with recruitment [55], by obtaining support from a dedicated organizational support group for implementation [58], providing lower-cost intervention options (eg, email based) [58], using reliable data storage methods [58], and demonstrating cost-effectiveness of the proposed intervention [64].

XSL•FO
RenderX

Table 7. Examples of barriers and facilitators organized under the inner setting Consolidated Framework for Implementation Research domain (N=217).

Relevant associated construct—proportion of barriers and facilitators; n (%) and brief description	Example of identified barriers	Example of identified facilitators
Structural characteristics; 4 (1.8%); the social architecture, age, maturity, and size of an organization	Personnel shortage, turnover, and organizational restructuring hindered the use of the strategy considerably [49]	Changes in the organizations should be considered (in light of resulting delays and communication problems) when planning intervention studies [42]
Networks and communications; 4 (1.8%); the nature and quality of webs of social networks and the nature and quality of formal and informal communications within an organization	Restrictive internet security settings was a barrier for accessing the intervention [42]	Conduct onsite testing before implementation [42]
Implementation climate; 17 (7.8); the absorptive capacity for change, shared receptivity of involved individuals to an intervention, and the extent to which use of that intervention will be rewarded, supported, and expected within their organization	Alignment with other stakeholders was absent and resulted in poor adherence to the recommended roles and tasks [62]	Embedding the intervention in a well-established wellness program to benefit from existing infrastructure to promote the intervention; users benefiting from incentive programs [58]
Tension for change; 1 (0.5); the degree to which stakeholders perceive the current situation as intolerable or needing change	Some stakeholders may be reluctant to implement new technology as it might threaten their ability to keep their job [39]	—[a]
Compatibility; 21 (9.7); the degree of tangible fit between meaning and values attached to the intervention by involved individuals; how those align with individuals' own norms, values, and perceived risks and needs; and how the intervention fits with existing workflows and systems	It was not possible for employees to contact their occupational physician themselves by telephone outside their regular consultations. This could have caused difficulty when an employee struggled with a module in Return@Work and wanted to ask the occupational physician for advice [54]	Alignment to relevant stakeholders is also important and can be attained by offering ongoing support to leaders at all organizational levels during an implementation [62]
Organizational incentives and rewards; 2 (0.9); extrinsic incentives such as goal-sharing awards, performance reviews, promotions, and raises in salary and less tangible incentives such as increased stature or respect	Complimentary gifts (eg, measuring tapes to be used by users with diabetes) with logos and information stimulate discussions and act as reminders [64]	—
Readiness for implementation; 6 (2.8); tangible and immediate indicators of organizational commitment to its decision to implement an intervention	Ensuring fidelity as coaches could not provide good feedback without supervision [63]	Consult review boards and consider these issues early in the data planning process [58]
Leadership engagement; 7 (3.2); commitment, involvement, and accountability of leaders and managers with the implementation	Senior management was not engaged and too much responsibility for implementation was given to the team members who did not prioritize these activities [49]	Adherence is better when managers are active and engaged [64]
Available resources; 25 (11.5); the level of resources dedicated for implementation and ongoing operations including money, training, education, physical space, and time	The intervention required all participants to allocate the same time slot and competed with other time commitments [42]	Supporting statement from the employers which will suggest to all employees who participate in the study that they will have 1 hour per week over the 8-week period to complete the program [59]
Access to knowledge and information; 2 (0.9); ease of access to digestible information and knowledge about the intervention and how to incorporate it into work tasks	Email messages from the decision aid supported the occupational physicians when guiding employees. The email gave them sufficient information and the layout was visually attractive [54]	

[a]No facilitator reported.

Characteristics of Individuals

Barriers were related to either the employer or the individual (Table 8). Employer-related barriers included the perception of low organizational commitment to addressing issues targeted by the proposed intervention [49], perceived stigma associated with intervention adoption [57], and a lack of privacy (eg, sharing information disclosed within the intervention with employers) [60]. Individual-related barriers included a general lack of motivation and interest in using the intervention [40,53], no opportunities to interact with others during the intervention [57], poor consistency in using the intervention as directed [60], poor digital skills [8,41,48,49], difficulty relating to content [60], low work ability [47], and reduction in engagement and adoption due to symptoms associated with medical conditions [52,53]. Proposed facilitators include willingness to seek mental health support [50], prior experience using an eHealth intervention and interventions that are freely accessible [47], low technical skill requirement (eg, no authentication) [41], and content that is available in multiple media formats (eg, printed versions) [41,57].

XSL•FO
RenderX

Table 8. Examples of barriers and facilitators organized under the characteristics of individuals Consolidated Framework for Implementation Research domain (N=217).

Relevant associated construct—proportion of barriers and facilitators; n (%) and brief description	Example of identified barriers	Example of identified facilitators
Knowledge and beliefs about the intervention; 7 (3.2); individuals' attitudes toward and the value placed on the intervention as well as familiarity with facts, truths, and principles related to the intervention	Skepticism toward the independence of the project from the organization [36]	Maintaining confidentiality between employee and employer [59]
Self-efficacy; 12 (5.5); individual belief in their own capabilities to execute courses of action to achieve implementation goals	Lack of computer skills in team members [49]	The package developed in a free-to-access and simple format that does not require logging in to a system or any specific technical expertise [41]
Other personal attributes; 6 (2.8); a broad construct to include other personal traits such as tolerance of ambiguity, intellectual ability, motivation, values, competence, capacity, and learning style	Barriers reported by participants at high risk for a major depressive episode included perceived stigma, lack of interaction with others that is characteristic of eMental health, lack of time, and lack of knowledge [57]	Willingness to seek professional mental health services [50]

Summary of Facilitation Measures

The identified facilitation measures were further synthesized and organized by the associated CFIR construct (Table 9).

XSL•FO
RenderX

Table 9. Summary of potential facilitation measures organized by associated Consolidated Framework for Implementation Research (CFIR) construct.

Associated CFIR construct	Facilitation measure
Evidence strength and quality	Strategies must provide evidence of effectiveness regarding the proposed or similar interventions in similar contexts featuring a representative sample of employees and a control group, where feasible, using valid and reliable measures.
Relative advantage	Strategies must be perceived to provide an advantage over the implementation of an alternative or no solution.
Adaptability	Strategies must allow flexibility on intervention completion times, the pace of progression, access options, and the format of provided materials.
Design quality and packaging	Strategies must ensure that the design of the intervention is based on an explicit understanding of users, their tasks, and environments and provides guidance (eg, reminders, knowledge base, progress tracking, and feedback); considers opportunities to integrate intervention features with organizational processes; creates personalized, informative, and nonstigmatizing content that encourages user participation; provides user adaptable content and tasks (ie, increased user control); allows access via additional modalities (eg, ability to print content) and formats (eg, video and audio); includes formative and summative usability testing and accessibility evaluations; highlights a strict approach to privacy and data security; and considers a multichannel recruitment strategy.
External policy and incentives	Strategies must identify and comply with applicable privacy legislation and policy regulations.
Structural characteristics	Strategies must consider the capacity of stakeholders to complete assigned tasks and account for turnover and other restructuring activities.
Networks and communications	Strategies must involve all stakeholders, include onsite testing of required technology, and establish clear communication procedures at the planning stage.
Implementation climate	Strategies must be cohesive and compatible with the organization's culture (eg, high turnover and highly active working environment), ensure that interventions can be used in distraction-free environments (ie, free from excessive noise), account for prior negative experiences with similar interventions, secure support from senior management for strategy implementation, and leverage existing programs by embedding interventions into them.
Tension for change	Strategies must consider the impact of implementation on-the-job security of stakeholders and how that affects their perception of proposed changes.
Compatibility	Strategies must adequately reflect the implementation needs of the organization and its existing processes and policies; be aligned with stakeholders at different organizational levels; provide adequate separation between work and working with the intervention; and avoid stigmatization, especially of employees with mental health conditions.
Organizational incentives and rewards	Strategies should offer incentives for using the intervention and consider incorporating gamification components to offer these incentives.
Readiness for implementation	Strategies must ensure that stakeholders are involved in strategy development, aware of the strategy and their role in it, equipped with the necessary tools and access, and adequately trained to implement the strategy.
Leadership engagement	Strategies must secure support from all stakeholders, especially an active and engaged senior management who strongly sanctions and advocates for the intervention.
Available resources	Strategies must provide organizational support for implementation, intervention support for users, dedicated time and private spaces for completing interventions in the workplace, less time-intensive interventions, alternative options to live-participation activities (eg, live webinar recording), low-cost technology-based options (eg, email) for interventions, reliable cloud data storage, access from varying device types, and implementation cost estimates with demonstrated cost-effectiveness.
Access to knowledge and information	Strategies must provide information that sets realistic expectations about the intervention and how to implement it.
Knowledge and beliefs about the intervention	Strategies must clearly articulate the role of the organization in the development of the intervention and address privacy and stigmatization concerns associated with using mental health interventions.
Self-efficacy	Strategies must accommodate users whose performance is affected by symptoms (eg, lack of motivation) associated with their health conditions (eg, depression) and a lack of confidence using technology.
Individual identification with organization	Strategies must consider users' perception of and level of commitment to the organization.
Other personal attributes	Strategies must address a lack of motivation (eg, due to symptoms associated with health conditions) to adopt and consistently use interventions and to seek help.

XSL•FO
RenderX

Discussion

Principal Findings and Comparison With Prior Work

The 31 included publications revealed 98 implementation strategies used when implementing OeMH interventions, 114 barriers, and 131 facilitators. The findings support observations [12,13] that the reporting of implementation strategies used for eHealth interventions is largely incomplete, nonsystematic, and unstructured. Nonetheless, the findings provide valuable insights into what is known and where knowledge gaps lie in the area.

Implementation Strategies

The OeMH knowledge base does not provide definitive answers regarding the implementation strategies to adopt and when and how it is most effective and efficient to adopt them. For example, the efficacy and cost-effectiveness of using innovative methods such as web-based targeted advertising compared with traditional methods (eg, posters) to increase reach is unclear [66,67], despite the former's success in being more time-efficient [67] and effective at recruiting hard-to-reach populations [67,68]. Those responsible for implementation must use their judgment about which of the provided strategies would be most appropriate for their circumstances. These findings support the notion that the implementation of eHealth technology (eg, eMental health [eMH] interventions) is often narrowly seen as a postdevelopment activity rather than being a crucial part of the development process [69]. Nonetheless, this could be partly a consequence of many included studies not specifically or comprehensively investigating implementation and therefore not reporting other details regarding implementation. Alternatively, publication restrictions [70] (eg, strict word limits) and the multidisciplinary nature of digital health research [71] may prioritize other study information over details regarding implementation when reporting on digital health interventions.

Barriers and Facilitators

Similar to findings related to medical devices [72], the findings here also suggest that usability [73] appears to be the main design consideration in the evaluation of OeMH interventions, with little consideration given to other critical elements of the user experience. Findings regarding the CFIR inner setting domain highlight the need for researchers to articulate potential facilitators, including those that may have failed in one implementation context, as they might work in other contexts. Existing research [74,75] addresses many of the barriers (eg, associated with symptoms associated with mental health problems and limited digital literacy skills) categorized under the CFIR characteristics of individual domains and could provide an easy opportunity to improve implementation if given more consideration during the planning phase. Factors external to the organization (eg, external policies, partners, and competition) are known to greatly hinder or support the successful implementation of technology [76-78] but have been largely undocumented or overlooked by the included publications.

Recency of Work and Coverage of Technologies

Similar to recent eMH reviews focusing on college students [79] and user engagement [80], this review also reported an increase since 2015 in eMH intervention studies meeting broad inclusion criteria. Recent reviews [79,80] also found that the eMH interventions described in the included studies were primarily web-based despite the added benefits of mobile apps that are coded for a specific mobile operating system such as iOS and Android (eg, faster, functionality-rich, and offline access) [81]. This is perhaps because web-based interventions likely cost less to develop and could be accessed via more devices if they were developed in a responsive way [81]. Emerging technologies, including AI, were considered in our search strategy, but were not used by the OeMH interventions described in the included studies. Nonetheless, this knowledge area is expected to increasingly feature the use of emerging technologies in the near future as the focus extends beyond nascent explorations of their applications for mental health and investigates the optimization of their implementation as well [82].

Implications and Recommendations for Practice and Future Research

Based on the findings of the scoping review, four practical recommendations could be considered to avoid and mitigate the identified barriers and improve the implementation of OeMH interventions:

1. Strategies must demonstrate a relative advantage over alternative solutions and promote flexibility in the delivery of interventions based on an explicit understanding of users, their tasks, and environments.
2. Strategies must promote the active engagement of organizational leadership, assess organizational readiness, and ensure compatibility with the organization's technological infrastructure and culture, in addition to providing desirable incentives and the necessary resources (eg, time and information about the intervention) for users to use the intervention as directed.
3. Strategies must ensure transparency regarding the intervention and implications of use and help users build confidence in their ability to benefit from the intervention.
4. Strategies must identify and ensure that interventions comply with applicable privacy legislation and policy regulations.

Future IR should continue with the broad aim of understanding what, why, and how OeMH interventions work under real-world conditions, and how to improve their implementation. The findings do not support the prioritization of any one aim over others. However, findings show that IR principles [83] such as the importance of context (eg, industry, size, and policies) and the people using the research need more attention for OeMH interventions. For example, surprisingly few findings were relevant to CFIR contextual domains (eg, outer setting), which speak to governmental regulations similar to COVID-19–related policies that have a strong influence on working arrangements. In addition, the general lack of detailed, systematic, and standardized reporting on proposed digital health interventions (eg, CONSORT-EHEALTH [Consolidated Standards of

XSL•FO
RenderX

Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth]—Expanded CONSORT figure) [32,84-86] and the implementation strategies used to achieve these outcomes (eg, Standards for Reporting Implementation Studies) [19,70] need to be remedied for IR to be properly used in this area. Reporting could benefit from subscribing to technology-centric frameworks (eg, the mobile health evidence reporting and assessment checklist [71] and the integrated technology implementation model [76-78]) that are more comprehensive in capturing key technology implementation factors (eg, accreditation, regulation, technology vendors, individual adoption factors, and interfacing systems). This should allow future studies to replicate and develop theories based on assessments of the implementation strategies used. In addition, any encountered or anticipated barriers and corresponding remedies that might be useful in avoiding these barriers or reducing their negative impact on implementation should also be reported. The development of an OeMH implementation checklist that includes comprehensive reporting guidelines and other prompts to ensure consistency and completeness when implementing these interventions would be beneficial. Future IR should also focus on investigating a wider range of common implementation outcomes (eg, cost-effectiveness and sustainability) [87] facilitated by implementation strategies for OeMH interventions that also target more common mental health problems in the workplace (eg, anxiety, substance use, and addiction). Issues regarding lack of digital access and digital inequity are an ongoing challenge [88], although not prominently featured in the results, and should be considered to avoid OeMH interventions contributing to any disparities. This study should also investigate how implementation strategies for OeMH interventions could benefit from emerging technologies. For instance, AI can use usage data to complement existing methods to better identify people who are at a high risk of mental health problems, support health decision-making, and offer resources that meet users' individual health needs [89]. This could have a profound positive impact on implementation through improvements in the effectiveness and maintenance of interventions.

Limitations

Search results were limited to publications in English, and a publication date restriction was imposed from 2010 onwards; however, given the broad search strategy, it is not anticipated that many, if any, potentially eligible publications were missed as a result. The term *eMental* was coined in 2002 [90], merely 8 years before this review's year restriction, and a recent review of 50 publications about OeMH interventions [6] included 11 publications that were published before 2010 and none were eligible for inclusion in this study. In addition, despite our exhaustive search strategy, 6 publications from 2010 to 2015 compared with 25 from the subsequent 5-year period were eventually included. Incomplete reporting also made it challenging to detail strategies (eg, their effectiveness), barriers, facilitators, and contextual data (eg, industry, organizational size, and employee level) from the included publications and to synthesize these data later. Nonetheless, all researchers involved in data extraction completed the training specifically for this review, followed the same thorough approach, and the extracted data were reviewed at least once by a second researcher. Interrater reliability was not calculated, and reasons for disagreement in screening decisions were not reported, which might have affected the reproducibility of this study [91]. However, this does not compromise the consistency and accuracy of the screening. Moreover, two 2-hour workshops were conducted with training sessions, and reconciliation meetings were consequently held when there were inconsistencies in screening decisions.

Although multiple implementation strategies can legitimately contribute to multiple RE-AIM domains, adopting a framework with more specificity could potentially be useful for the identification of more targeted strategies. Common implementation models (eg, RE-AIM and CFIR) predate the current development of eHealth, and concerns about their inability to fully capture the complexities of eHealth implementation have been raised [69] and persist [92] despite some recent updates [32,86] and clarifications [33]. Nonetheless, these generic frameworks are useful for guiding data extraction and as tools for making valuable comparisons with other types of interventions. Similar to other scoping reviews, this review reports on the nature and features of the literature on the topic of focus and does not attempt to present a view regarding the appropriateness of the used methods and the strength or quality of evidence. Similarly, the provision of more detailed recommendations would have been premature and potentially misleading, as this was unsupported by the data collected. Further research is needed to determine valid facilitators and how they should be used in the process of OeMH development and delivery on a case-by-case basis while considering contextual factors such as industry, organizational size, employee level, and internal and external policies. Nevertheless, these recommendations could still be particularly relevant for OeMH interventions in comparison with similar interventions in different contexts. Consequently, readers should be mindful that the review cannot determine whether the included studies provide robust or generalizable findings.

Conclusions

This scoping review represents one of the first steps in a research agenda aimed at improving the implementation of OeMH interventions by systematically selecting, shaping, evaluating, and reporting implementation strategies. It has identified 98 implementation strategies, 114 barriers, and 131 facilitation measures related to the implementation of these interventions. A synthesis of these findings offers 19 recommendations that provide initial guidance on how to improve the implementation of OeMH interventions. This scoping review also highlighted the need to combine common implementation models (eg, RE-AIM and CFIR) with more technology-centric frameworks (eg, integrated technology implementation model and the mobile health evidence reporting and assessment checklist) to fully capture the complexities of eHealth implementation. Despite yielding less detailed insight than hoped, owing to incomplete reporting and the adoption of incomprehensive frameworks by the included publications, this scoping review's findings can still be critically leveraged by discerning decision-makers to improve the reach, effectiveness, adoption, implementation, and maintenance of OeMH interventions.

XSL•FO
RenderX

Acknowledgments

The research leading to these results received funding from the European Union Horizon 2020 Research and Innovation Programme under grant 848180 and the National Health and Medical Research Council of Australia under grant APP1195937. CdM received funding in the form of a predoctoral grant from Generalitat de Catalunya (PIF-Salut grant, code SLT017/20/000138).

The authors especially thank Dr Beatriz Olaya, Dr Christophe Vanroelen, Ellen Vorstenbosch, Dr Desirée Gutiérrez Marín, and other colleagues from the European Platform to Promote Well-being and Health in the Workplace Consortium for their generous feedback on the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search concepts and terms and MEDLINE search strategy.
[DOCX File , 16 KB - jmir_v24i6e34479_app1.docx]

Multimedia Appendix 2
Identified examples of implementation strategies organized by relevant reach, effectiveness, adoption, implementation, and maintenance domains.
[XLSX File (Microsoft Excel File), 18 KB - jmir_v24i6e34479_app2.xlsx]

Multimedia Appendix 3
Identified barriers and facilitators organized by relevant Consolidated Framework for Implementation Research domain and associated construct.
[XLSX File (Microsoft Excel File), 25 KB - jmir_v24i6e34479_app3.xlsx]

References

1. Depression and other common mental disorders: global health estimates. World Health Organization. 2017. URL: https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf [accessed 2022-05-18]
2. Young KP, Kolcz DL, O'Sullivan DM, Ferrand J, Fried J, Robinson K. Health care workers' mental health and quality of life during COVID-19: results from a mid-pandemic, national survey. Psychiatr Serv 2021 Feb 01;72(2):122-128. [doi: 10.1176/appi.ps.202000424] [Medline: 33267652]
3. Cahill J, Cullen P, Anwer S, Wilson S, Gaynor K. Pilot Work Related Stress (WRS), effects on wellbeing and mental health, and coping methods. Int J Aerospace Psychol 2021 Jan 14;31(2):87-109. [doi: 10.1080/24721840.2020.1858714]
4. Sampson H, Ellis N. Stepping up: the need for proactive employer investment in safeguarding seafarers' mental health and wellbeing. Maritime Policy Manag 2020 Dec 30;48(8):1069-1081. [doi: 10.1080/03088839.2020.1867918]
5. Hamouche S. COVID-19 and employees' mental health: stressors, moderators and agenda for organizational actions. Emerald Open Res 2020 Apr 20;2:15. [doi: 10.35241/emeraldopenres.13550.1]
6. Phillips EA, Gordeev VS, Schreyögg J. Effectiveness of occupational e-mental health interventions: a systematic review and meta-analysis of randomized controlled trials. Scand J Work Environ Health 2019 Nov 01;45(6):560-576 [FREE Full text] [doi: 10.5271/sjweh.3839] [Medline: 31184758]
7. Riper H, Andersson G, Christensen H, Cuijpers P, Lange A, Eysenbach G. Theme issue on e-mental health: a growing field in internet research. J Med Internet Res 2010 Dec 19;12(5):e74 [FREE Full text] [doi: 10.2196/jmir.1713] [Medline: 21169177]
8. Lehr D, Geraedts A, Perrson Asplund R, Khadjesari Z, Heber E, de Bloom J, et al. Occupational e-mental health: current approaches and promising perspectives for promoting mental health in workers. In: Wiencke M, Cacace M, Fischer S, editors. Healthy at Work: Interdisciplinary Perspectives. Cham, Switzerland: Springer; Aug 27, 2016:257-281.
9. Ha SW, Kim J. Designing a scalable, accessible, and effective mobile app based solution for common mental health problems. Int J Human Comput Interact 2020 Apr 26;36(14):1354-1367. [doi: 10.1080/10447318.2020.1750792]
10. Rodriguez-Villa E, Naslund J, Keshavan M, Patel V, Torous J. Making mental health more accessible in light of COVID-19: scalable digital health with digital navigators in low and middle-income countries. Asian J Psychiatr 2020 Dec;54:102433. [doi: 10.1016/j.ajp.2020.102433] [Medline: 33271713]
11. Purtova N, Kosta E, Koops BJ. Laws and regulations for digital health. In: Fricker SA, Thümmler C, Gavras A, editors. Requirements Engineering for Digital Health. Cham, Switzerland: Springer; 2014:47-74.
12. Proctor EK, Powell BJ, McMillen JC. Implementation strategies: recommendations for specifying and reporting. Implement Sci 2013 Dec 01;8:139 [FREE Full text] [doi: 10.1186/1748-5908-8-139] [Medline: 24289295]

XSL•FO
RenderX

13. Powell BJ, McMillen JC, Proctor EK, Carpenter CR, Griffey RT, Bunger AC, et al. A compilation of strategies for implementing clinical innovations in health and mental health. Med Care Res Rev 2012 Apr;69(2):123-157 [FREE Full text] [doi: 10.1177/1077558711430690] [Medline: 22203646]
14. Connolly SL, Hogan TP, Shimada SL, Miller CJ. Leveraging implementation science to understand factors influencing sustained use of mental health apps: a narrative review. J Technol Behav Sci (forthcoming) 2020 Sep 07:1-13 [FREE Full text] [doi: 10.1007/s41347-020-00165-4] [Medline: 32923580]
15. Graham AK, Lattie EG, Powell BJ, Lyon AR, Smith JD, Schueller SM, et al. Implementation strategies for digital mental health interventions in health care settings. Am Psychol 2020 Nov;75(8):1080-1092 [FREE Full text] [doi: 10.1037/amp0000686] [Medline: 33252946]
16. Dryden-Palmer KD, Parshuram CS, Berta WB. Context, complexity and process in the implementation of evidence-based innovation: a realist informed review. BMC Health Serv Res 2020 Feb 03;20(1):81 [FREE Full text] [doi: 10.1186/s12913-020-4935-y] [Medline: 32013977]
17. 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]
18. Varsi C, Solberg Nes L, Kristjansdottir OB, Kelders SM, Stenberg U, Zangi HA, et al. Implementation strategies to enhance the implementation of eHealth programs for patients with chronic illnesses: realist systematic review. J Med Internet Res 2019 Sep 27;21(9):e14255 [FREE Full text] [doi: 10.2196/14255] [Medline: 31573934]
19. Pinnock H, Barwick M, Carpenter CR, Eldridge S, Grandes G, Griffiths CJ, StaRI Group. Standards for Reporting Implementation Studies (StaRI): explanation and elaboration document. BMJ Open 2017 Apr 03;7(4):e013318 [FREE Full text] [doi: 10.1136/bmjopen-2016-013318] [Medline: 28373250]
20. EMPOWER: The European platform to promote wellbeing and health in the workplace. EMPOWER Consortium. 2021. URL: https://empower-project.eu/ [accessed 2022-05-18]
21. Colquhoun HL, Levac D, O'Brien KK, Straus S, Tricco AC, Perrier L, et al. Scoping reviews: time for clarity in definition, methods, and reporting. J Clin Epidemiol 2014 Dec;67(12):1291-1294. [doi: 10.1016/j.jclinepi.2014.03.013] [Medline: 25034198]
22. 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]
23. 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]
24. Khalil H, Peters M, Godfrey CM, McInerney P, Soares CB, Parker D. An evidence-based approach to scoping reviews. Worldviews Evid Based Nurs 2016 Apr;13(2):118-123. [doi: 10.1111/wvn.12144] [Medline: 26821833]
25. Westphaln KK, Regoeczi W, Masotya M, Vazquez-Westphaln B, Lounsbury K, McDavid L, et al. From Arksey and O'Malley and beyond: customizations to enhance a team-based, mixed approach to scoping review methodology. MethodsX 2021 May 7;8:101375 [FREE Full text] [doi: 10.1016/j.mex.2021.101375] [Medline: 34430271]
26. 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]
27. Global strategy on digital health 2020-2025. World Health Organization. 2021 Aug 18. URL: https://www.who.int/publications/i/item/9789240020924 [accessed 2022-05-10]
28. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
29. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
30. Denison HJ, Dodds RM, Ntani G, Cooper R, Cooper C, Sayer AA, et al. How to get started with a systematic review in epidemiology: an introductory guide for early career researchers. Arch Public Health 2013 Aug 07;71(1):21 [FREE Full text] [doi: 10.1186/0778-7367-71-21] [Medline: 23919540]
31. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]
32. Glasgow RE, Harden SM, Gaglio B, Rabin B, Smith ML, Porter GC, et al. RE-AIM planning and evaluation framework: adapting to new science and practice with a 20-year review. Front Public Health 2019 Mar 29;7:64 [FREE Full text] [doi: 10.3389/fpubh.2019.00064] [Medline: 30984733]
33. Holtrop JS, Estabrooks PA, Gaglio B, Harden SM, Kessler RS, King DK, et al. Understanding and applying the RE-AIM framework: clarifications and resources. J Clin Transl Sci 2021 May 14;5(1):e126 [FREE Full text] [doi: 10.1017/cts.2021.789] [Medline: 34367671]
34. 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]

XSL•FO
RenderX

35. Strifler L, Cardoso R, McGowan J, Cogo E, Nincic V, Khan PA, et al. Scoping review identifies significant number of knowledge translation theories, models, and frameworks with limited use. J Clin Epidemiol 2018 Aug;100:92-102. [doi: 10.1016/j.jclinepi.2018.04.008] [Medline: 29660481]
36. Ravalier JM, Wainwright E, Smyth N, Clabburn O, Wegrzynek P, Loon M. Co-creating and evaluating an app-based well-being intervention: the HOW (healthier outcomes at work) social work project. Int J Environ Res Public Health 2020 Nov 24;17(23):8730 [FREE Full text] [doi: 10.3390/ijerph17238730] [Medline: 33255460]
37. Liem A, Garabiles MR, Pakingan KA, Chen W, Lam AI, Burchert S, et al. A digital mental health intervention to reduce depressive symptoms among overseas Filipino workers: protocol for a pilot hybrid type 1 effectiveness-implementation randomized controlled trial. Implement Sci Commun 2020 Oct 31;1:96 [FREE Full text] [doi: 10.1186/s43058-020-00072-y] [Medline: 33145495]
38. Freund J, Titzler I, Thielecke J, Braun L, Baumeister H, Berking M, et al. Implementing Internet- and tele-based interventions to prevent mental health disorders in farmers, foresters and gardeners (ImplementIT): study protocol for the multi-level evaluation of a nationwide project. BMC Psychiatry 2020 Aug 27;20(1):424 [FREE Full text] [doi: 10.1186/s12888-020-02800-z] [Medline: 32854660]
39. Engdahl P, Svedberg P, Lexén A, Bejerholm U. Role of a digital return-to-work solution for individuals with common mental disorders: qualitative study of the perspectives of three stakeholder groups. JMIR Form Res 2020 Sep 16;4(9):e15625 [FREE Full text] [doi: 10.2196/15625] [Medline: 32936089]
40. 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]
41. Blake H, Bermingham F, Johnson G, Tabner A. Mitigating the psychological impact of COVID-19 on healthcare workers: a digital learning package. Int J Environ Res Public Health 2020 Apr 26;17(9):2997 [FREE Full text] [doi: 10.3390/ijerph17092997] [Medline: 32357424]
42. Wan Mohd Yunus WM, Musiat P, Brown JS. Evaluating the feasibility of an innovative self-confidence webinar intervention for depression in the workplace: a proof-of-concept study. JMIR Ment Health 2019 Apr 26;6(4):e11401 [FREE Full text] [doi: 10.2196/11401] [Medline: 31025943]
43. Kerr DC, Ornelas IJ, Lilly MM, Calhoun R, Meischke H. Participant engagement in and perspectives on a web-based mindfulness intervention for 9-1-1 telecommunicators: multimethod study. J Med Internet Res 2019 Jun 19;21(6):e13449 [FREE Full text] [doi: 10.2196/13449] [Medline: 31219045]
44. Davey S, Gordon S, Tester R. Addressing police discrimination regarding mental distress using a service user-led and interpersonal contact/education based 'e-Learning'. Police Pract Res 2019 Nov 07;22(1):426-442. [doi: 10.1080/15614263.2019.1689128]
45. Villaume K, Tafvelin S, Hasson D. Health-relevant personality traits in relation to adherence to a web-based occupational health promotion and stress management intervention. Int J Workplace Health Manag 2018 Jun 04;11(3):143-158. [doi: 10.1108/ijwhm-11-2017-0092]
46. Rush KE. Paving the path to mindfulness: implementation of a program to reduce stress and burnout in inpatient psychiatric nurses. University of North Carolina Digital Repository. 2018. URL: https://cdr.lib.unc.edu/concern/dissertations/w37637581?locale=en [accessed 2022-05-18]
47. Hennemann S, Witthöft M, Bethge M, Spanier K, Beutel ME, Zwerenz R. Acceptance and barriers to access of occupational e-mental health: cross-sectional findings from a health-risk population of employees. Int Arch Occup Environ Health 2018 Apr;91(3):305-316. [doi: 10.1007/s00420-017-1280-5] [Medline: 29189895]
48. Havermans BM, Boot CR, Brouwers EP, Houtman IL, Anema JR, van der Beek AJ. Process evaluation of a digital platform-based implementation strategy aimed at work stress prevention in a health care organization. J Occup Environ Med 2018 Sep;60(9):e484-e491. [doi: 10.1097/JOM.0000000000001402] [Medline: 30199413]
49. Havermans BM, Boot CR, Brouwers EP, Houtman IL, Heerkens YF, Zijlstra-Vlasveld MC, et al. Effectiveness of a digital platform-based implementation strategy to prevent work stress in a healthcare organization: a 12-month follow-up controlled trial. Scand J Work Environ Health 2018 Nov 01;44(6):613-621 [FREE Full text] [doi: 10.5271/sjweh.3758] [Medline: 30033477]
50. Hall BJ, Shi W, Garabiles MR, Chan EW. Correlates of expected eMental Health intervention uptake among Filipino domestic workers in China. Glob Ment Health (Camb) 2018 Apr;5:e33 [FREE Full text] [doi: 10.1017/gmh.2018.25] [Medline: 30455968]
51. Frykman M, Lundmark R, von Thiele Schwarz U, Villaume K, Hasson H. Line managers' influence on employee usage of a web-based system for occupational health management. Int J Workplace Health Manag 2018 Aug 06;11(4):193-209. [doi: 10.1108/ijwhm-12-2017-0104]
52. Deady M, Johnston D, Milne D, Glozier N, Peters D, Calvo R, et al. Preliminary effectiveness of a smartphone app to reduce depressive symptoms in the workplace: feasibility and acceptability study. JMIR Mhealth Uhealth 2018 Dec 04;6(12):e11661 [FREE Full text] [doi: 10.2196/11661] [Medline: 30514694]

XSL•FO
RenderX

53. Carolan S, de Visser RO. Employees' perspectives on the facilitators and barriers to engaging with digital mental health interventions in the workplace: qualitative study. JMIR Ment Health 2018 Jan 19;5(1):e8 [FREE Full text] [doi: 10.2196/mental.9146] [Medline: 29351900]
54. Volker D, Zijlstra-Vlasveld MC, Brouwers EP, van der Feltz-Cornelis DF. Process evaluation of a blended Web-based intervention on return to work for sick-listed employees with common mental health problems in the occupational health setting. J Occup Rehabil 2017 Jun;27(2):186-194 [FREE Full text] [doi: 10.1007/s10926-016-9643-4] [Medline: 27150734]
55. Carolan S, Harris PR, Greenwood K, Cavanagh K. Increasing engagement with an occupational digital stress management program through the use of an online facilitated discussion group: results of a pilot randomised controlled trial. Internet Interv 2017 Dec;10:1-11 [FREE Full text] [doi: 10.1016/j.invent.2017.08.001] [Medline: 30135747]
56. Zarski AC, Lehr D, Berking M, Riper H, Cuijpers P, Ebert DD. Adherence to Internet-based mobile-supported stress management: a pooled analysis of individual participant data from three randomized controlled trials. J Med Internet Res 2016 Jun 29;18(6):e146 [FREE Full text] [doi: 10.2196/jmir.4493] [Medline: 27357528]
57. Wang J, Lam RW, Ho K, Attridge M, Lashewicz BM, Patten SB, et al. Preferred features of e-mental health programs for prevention of major depression in male workers: results from a Canadian national survey. J Med Internet Res 2016 Jun 06;18(6):e132 [FREE Full text] [doi: 10.2196/jmir.5685] [Medline: 27267782]
58. Smith KC, Wallace DP. Improving the sleep of children's hospital employees through an email-based sleep wellness program. Clin Pract Pediatric Psychol 2016 Sep;4(3):291-305. [doi: 10.1037/cpp0000152]
59. Carolan S, Harris PR, Greenwood K, Cavanagh K. Increasing engagement with, and effectiveness of, an online CBT-based stress management intervention for employees through the use of an online facilitated bulletin board: study protocol for a pilot randomised controlled trial. Trials 2016 Dec 15;17(1):598 [FREE Full text] [doi: 10.1186/s13063-016-1733-2] [Medline: 27978858]
60. Possemato K, Acosta MC, Fuentes J, Lantinga LJ, Marsch LA, Maisto SA, et al. A Web-based self-management program for recent combat veterans with PTSD and substance misuse: program development and veteran feedback. Cogn Behav Pract 2015 Aug 01;22(3):345-358 [FREE Full text] [doi: 10.1016/j.cbpra.2014.03.005] [Medline: 26120269]
61. Schneider J, Sarrami Foroushani P, Grime P, Thornicroft G. Acceptability of online self-help to people with depression: users' views of MoodGYM versus informational websites. J Med Internet Res 2014 Mar 28;16(3):e90 [FREE Full text] [doi: 10.2196/jmir.2871] [Medline: 24681717]
62. Hasson H, Villaume K, von Thiele Schwarz U, Palm K. Managing implementation: roles of line managers, senior managers, and human resource professionals in an occupational health intervention. J Occup Environ Med 2014 Jan;56(1):58-65. [doi: 10.1097/JOM.0000000000000020] [Medline: 24351889]
63. Geraedts AS, Kleiboer AM, Wiezer NM, Cuijpers P, van Mechelen W, Anema JR. Feasibility of a worker-directed Web-based intervention for employees with depressive symptoms. Internet Interv 2014 Jul;1(3):132-140. [doi: 10.1016/j.invent.2014.07.001]
64. Hasson D, Villaume K. An automated and systematic Web-based intervention for stress management and organizational health promotion. In: Bauer GF, Jenny GJ, editors. Salutogenic Organizations and Change. Dordrecht, The Netherlands: Springer; 2013:217-237.
65. Hasson H, Brown C, Hasson D. Factors associated with high use of a workplace web-based stress management program in a randomized controlled intervention study. Health Educ Res 2010 Aug;25(4):596-607. [doi: 10.1093/her/cyq005] [Medline: 20150531]
66. Reagan L, Nowlin SY, Birdsall SB, Gabbay J, Vorderstrasse A, Johnson C, et al. Integrative review of recruitment of research participants through Facebook. Nurs Res 2019;68(6):423-432. [doi: 10.1097/NNR.0000000000000385] [Medline: 31693547]
67. Kayrouz R, Dear BF, Karin E, Titov N. Facebook as an effective recruitment strategy for mental health research of hard to reach populations. Internet Interv 2016 May;4:1-10 [FREE Full text] [doi: 10.1016/j.invent.2016.01.001] [Medline: 30135786]
68. 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]
69. van Gemert-Pijnen L, Kelders SM, Kip H, Sanderman R. eHealth Research, Theory and Development: A Multidisciplinary Approach. London, UK: Routledge; 2018:247-270.
70. Pinnock H, Barwick M, Carpenter CR, Eldridge S, Grandes G, Griffiths CJ, StaRI Group. Standards for Reporting Implementation Studies (StaRI) statement. BMJ 2017 Mar 06;356:i6795 [FREE Full text] [doi: 10.1136/bmj.i6795] [Medline: 28264797]
71. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]
72. Bitkina OV, Kim HK, Park J. Usability and user experience of medical devices: an overview of the current state, analysis methodologies, and future challenges. Int J Ind Ergon 2020 Mar;76:102932. [doi: 10.1016/j.ergon.2020.102932]

XSL•FO
RenderX

73. Ergonomics of human-system interaction — Part 304: user performance test methods for electronic visual displays - ISO 9241-304:2008. International Organization for Standardization. 2018. URL: https://www.iso.org/standard/40099.html [accessed 2022-05-18]
74. Bernard R, Sabariego C, Cieza A. Difficulties encountered by people with depression and anxiety on the Web: qualitative study and Web-based expert survey. J Med Internet Res 2019 Oct 31;21(10):e12514 [FREE Full text] [doi: 10.2196/12514] [Medline: 31674915]
75. Bernard R, Sabariego C, Cieza A. Barriers and facilitation measures related to people with mental disorders when using the web: a systematic review. J Med Internet Res 2016 Jun 09;18(6):e157 [FREE Full text] [doi: 10.2196/jmir.5442] [Medline: 27282115]
76. Schoville RR. Discovery of implementation factors that lead to technology adoption in long-term care. J Gerontol Nurs 2017 Oct 01;43(10):21-26. [doi: 10.3928/00989134-20170914-06] [Medline: 28945269]
77. Schoville R, Titler MG. Integrated technology implementation model: examination and enhancements. Comput Inform Nurs 2020 Nov;38(11):579-589. [doi: 10.1097/CIN.0000000000000632] [Medline: 32520784]
78. Schoville RR, Titler MG. Guiding healthcare technology implementation: a new integrated technology implementation model. Comput Inform Nurs 2015 Mar;33(3):99-E1. [doi: 10.1097/CIN.0000000000000130] [Medline: 25799235]
79. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. J Med Internet Res 2019 Jul 22;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]
80. 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]
81. Panhale M. Beginning Hybrid Mobile Application Development. New York, NY, USA: Springer; 2016:15-20.
82. Baños RM, Herrero R, Vara MD. What is the current and future status of digital mental health interventions? Span J Psychol 2022 Feb 02;25:e5. [doi: 10.1017/SJP.2022.2] [Medline: 35105398]
83. Peters DH, Adam T, Alonge O, Agyepong IA, Tran N. Implementation research: what it is and how to do it. BMJ 2013 Nov 20;347:f6753. [doi: 10.1136/bmj.f6753] [Medline: 24259324]
84. 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]
85. Eysenbach G. CONSORT-EHEALTH: implementation of a checklist for authors and editors to improve reporting of Web-based and mobile randomized controlled trials. In: Studies in Health Technology and Informatics. Amsterdam, The Netherlands: IOS Press; 2013:657-661.
86. Glasgow RE, Huebschmann AG, Brownson RC. Expanding the CONSORT figure: increasing transparency in reporting on external validity. Am J Prev Med 2018 Sep;55(3):422-430. [doi: 10.1016/j.amepre.2018.04.044] [Medline: 30033029]
87. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
88. Crawford A, Serhal E. Digital health equity and COVID-19: the innovation curve cannot reinforce the social gradient of health. J Med Internet Res 2020 Jun 02;22(6):e19361 [FREE Full text] [doi: 10.2196/19361] [Medline: 32452816]
89. Ebert DD, Harrer M, Apolinário-Hagen J, Baumeister H. Digital interventions for mental disorders: key features, efficacy, and potential for artificial intelligence applications. Adv Exp Med Biol 2019;1192:583-627. [doi: 10.1007/978-981-32-9721-0_29] [Medline: 31705515]
90. Christensen H, Griffiths K, Evans K. E-Mental Health in Australia: Implications of the Internet and Related Technologies for Policy. Canberra, Australia: Commonwealth Department of Health and Ageing; 2002.
91. Belur J, Tompson L, Thornton A, Simon M. Interrater reliability in systematic review methodology: exploring variation in coder decision-making. Sociol Methods Res 2018 Sep 24;50(2):837-865. [doi: 10.1177/0049124118799372]
92. Heinsch M, Wyllie J, Carlson J, Wells H, Tickner C, Kay-Lambkin F. Theories informing eHealth implementation: systematic review and typology classification. J Med Internet Res 2021 May 31;23(5):e18500 [FREE Full text] [doi: 10.2196/18500] [Medline: 34057427]

Abbreviations

AI: artificial intelligence
CFIR: Consolidated Framework for Implementation Research
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
eMH: eMental health
EMPOWER: European Platform to Promote Well-being and Health in the Workplace
IR: implementation research

XSL•FO
RenderX

OeMH: Occupational eMental health
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RE-AIM: reach, effectiveness, adoption, implementation, and maintenance

Edited by R Kukafka; submitted 28.10.21; peer-reviewed by M Deady, K Wall, C Guinemer, I Mircheva, C Tickner, C Sinha; comments to author 27.12.21; revised version received 02.03.22; accepted 30.03.22; published 01.06.22.

Please cite as:
Bernard RM, Toppo C, Raggi A, de Mul M, de Miquel C, Pugliese MT, van der Feltz-Cornelis CM, Ortiz-Tallo A, Salvador-Carulla L, Lukersmith S, Hakkaart-van Roijen L, Merecz-Kot D, Staszewska K, Sabariego C
Strategies for Implementing Occupational eMental Health Interventions: Scoping Review
J Med Internet Res 2022;24(6):e34479
URL: https://www.jmir.org/2022/6/e34479
doi:10.2196/34479
PMID:35648457

©Renaldo M Bernard, Claudia Toppo, Alberto Raggi, Marleen de Mul, Carlota de Miquel, Maria Teresa Pugliese, Christina M van der Feltz-Cornelis, Ana Ortiz-Tallo, Luis Salvador-Carulla, Sue Lukersmith, Leona Hakkaart-van Roijen, Dorota Merecz-Kot, Kaja Staszewska, Carla Sabariego. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 01.06.2022. 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

Machine Learning in Health Promotion and Behavioral Change: Scoping Review

Yong Shian Goh[1*], MN, PhD; Jenna Qing Yun Ow Yong[1*], BPsy (Hons); Bernice Qian Hui Chee[2*], BSocSci (Hons); Jonathan Han Loong Kuek[3*], BPsy (Hons); Cyrus Su Hui Ho[4*], MBBS, MRCPsych

[1]Alice Lee Centre for Nursing Studies, National University of Singapore, Singapore, Singapore

[2]Faculty of Arts and Social Sciences, National University of Singapore, Singapore, Singapore

[3]Susan Wakil School of Nursing, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

[4]Department of Psychological Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[*]all authors contributed equally

Corresponding Author:
Yong Shian Goh, MN, PhD
Alice Lee Centre for Nursing Studies
National University of Singapore
MD11, Clinical Research Centre, Level 2
10 Medical Drive
Singapore, 117597
Singapore
Phone: 65 93896825
Email: nurgys@nus.edu.sg

Abstract

Background: Despite health behavioral change interventions targeting modifiable lifestyle factors underlying chronic diseases, dropouts and nonadherence of individuals have remained high. The rapid development of machine learning (ML) in recent years, alongside its ability to provide readily available personalized experience for users, holds much potential for success in health promotion and behavioral change interventions.

Objective: The aim of this paper is to provide an overview of the existing research on ML applications and harness their potential in health promotion and behavioral change interventions.

Methods: A scoping review was conducted based on the 5-stage framework by Arksey and O'Malley and the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Scoping Reviews) guidelines. A total of 9 databases (the Cochrane Library, CINAHL, Embase, Ovid, ProQuest, PsycInfo, PubMed, Scopus, and Web of Science) were searched from inception to February 2021, without limits on the dates and types of publications. Studies were included in the review if they had incorporated ML in any health promotion or behavioral change interventions, had studied at least one group of participants, and had been published in English. Publication-related information (author, year, aim, and findings), area of health promotion, user data analyzed, type of ML used, challenges encountered, and future research were extracted from each study.

Results: A total of 29 articles were included in this review. Three themes were generated, which are as follows: (1) enablers, which is the adoption of information technology for optimizing systemic operation; (2) challenges, which comprises the various hurdles and limitations presented in the articles; and (3) future directions, which explores prospective strategies in health promotion through ML.

Conclusions: The challenges pertained to not only the time- and resource-consuming nature of ML-based applications, but also the burden on users for data input and the degree of personalization. Future works may consider designs that correspondingly mitigate these challenges in areas that receive limited attention, such as smoking and mental health.

(J Med Internet Res 2022;24(6):e35831) doi:10.2196/35831

KEYWORDS

machine learning; health promotion; health behavioral changes; artificial intelligence

XSL•FO
RenderX

Introduction

Chronic diseases account substantially for morbidity [1] and mortality [2] worldwide. The occurrence of such diseases can be attributed to behavioral risk factors such as smoking, poor nutrition, alcohol consumption, and the lack of physical activity [3]. Despite the implementation of numerous lifestyle-based health behavioral interventions targeting modifiable risk factors such as obesity, stress management, and sedentary habits, dropouts and nonadherence of individuals to such recommendations have remained high [4]. Given the complex interplay [5] of factors such as financial and psychosocial, as well as vagueness of recommendations [6,7], it is challenging for one to understand the reasons underlying nonadherence. Of these factors, the psychosocial ones affecting adherence are manifold, where the difficulty is in changing lifestyles, attitudes, and beliefs of the individual coupled with feelings of guilt. Moreover, the feelings of hopelessness and isolation expressed by the individual further worsen by the lack of resources (eg, lack of support, food, time for behavioral changes, and treatment-related information). Against this background, it is evident that individuals face unique obstacles to their adherence to behavioral changes. To address such obstacles, behavioral change interventions need to be both multifaceted and personalized [6].

Machine learning (ML) techniques such as artificial intelligence (AI) and natural language processing over the past decade have resulted in advancements in learning algorithms, increased availability of online data, and low developmental costs [8]. The field of ML builds upon advanced statistical, computational, and probabilistic techniques to construct systems that automatically learn from data sets and require limited (ie, supervised) or no (ie, unsupervised) human input to yield accurate predictions and insights [9]. ML has been incorporated in various health processes such as detecting and diagnosing conditions [10], assessing and monitoring population health [11], providing prognoses and predicting treatment outcomes [12], and improving health research and clinical administration [13].

Alongside advances in mobile and wearable sensor technologies, AI monitoring systems, and telecare services [14], ML has gained popularity given its ability to analyze information from vast and complex data sets to provide readily available personalized experiences for users [8]. This ability to deliver tailored interventions may address the said problem of nonadherence [6], offering promising potential in health promotion and behavioral change interventions. Considering the rapid advances in ML over the last decade, this review aims to provide an overview of the existing research on machine learning applications and harness its potential in health promotion and behavioral change interventions.

Methods

Overview

Given the relative novelty of this area of research, a scoping review was undertaken to provide an overview of the literature. This review adopted the 5-stage framework by Arksey and O'Malley [15] and reported according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Scoping Reviews) guidelines [16], which is as follows: (1) identifying the research question; (2) identifying relevant studies; (3) selecting the studies; (4) charting the data; and (5) collating, summarizing, and reporting the results. As a scoping review aims to map the extent and nature of studies in the literature rather assessing their quality [15], the team did not perform a formal quality appraisal on the studies thus included.

Identifying the Research Question

The main research question formulated to guide this review was as follows: "How have machine learning technologies been used as a strategy for health promotion and behavioral change interventions?"

Identifying Suitable Studies

A search strategy was formulated to identify studies on ML techniques to promote behavioral changes and physical or mental health. A total of 9 databases were searched from inception to February 2021, aimed at encompassing not only multiple disciplines (Scopus and Web of Science), but also specific disciplines, including biomedical (the Cochrane Library, Embase, PubMed, Ovid, and ProQuest), nursing and allied health (CINAHL), and psychology (PsycInfo). To maximize the number of articles generated, no limits were applied to the dates and types of publications in order to maintain a comprehensive and updated search [17]. Relevant keywords and Medical Subject Headings (MeSH) terms were used, including "Learning, Machine," "Behavior Control," "Healthcare," and "Mental Health" (Multimedia Appendix 1). Where appropriate, the keywords were truncated, and Boolean terms were added to maximize the retrieval of all relevant articles. End references of the studies were also hand searched for relevant articles [18].

Selecting the Studies

Primary studies (including user preliminary testing or preliminary studies) were eligible for title and abstract screening if they had incorporated ML in promoting health or behavioral changes and were published in English (since the team lacked access to interpreters). Since this review aimed to examine real-life implications and experiences of incorporating ML techniques in health applications, articles were included only if they had studied at least a group of participants. Accordingly, those with no such reported results such as conference abstracts, proposals, and newspaper columns were excluded. Additionally, articles were excluded if they had examined other areas of health, such as the detection or diagnosis of conditions, population-health monitoring, prognosis and prediction of treatment outcomes, and research and clinical administration. Lastly, as ML techniques might overlap with statistical approaches [19], the 3 reviewers exercised discretion in determining those articles that had deployed ML techniques for inclusion.

All records retrieved from database searches were uploaded onto EndNote X9 (Clarivate Analytics), followed by the electronic removal of duplicates. During the screening of titles

XSL•FO
RenderX

and abstracts, the remaining articles were independently reviewed by 2 reviewers (JOY and BC). Articles not meeting the eligibility criteria were removed, while those deemed suitable by at least one reviewer were downloaded for further examination. The opinion of a third reviewer (YSG) was sought, and disagreements between the reviewers were consensually resolved.

Charting the Data

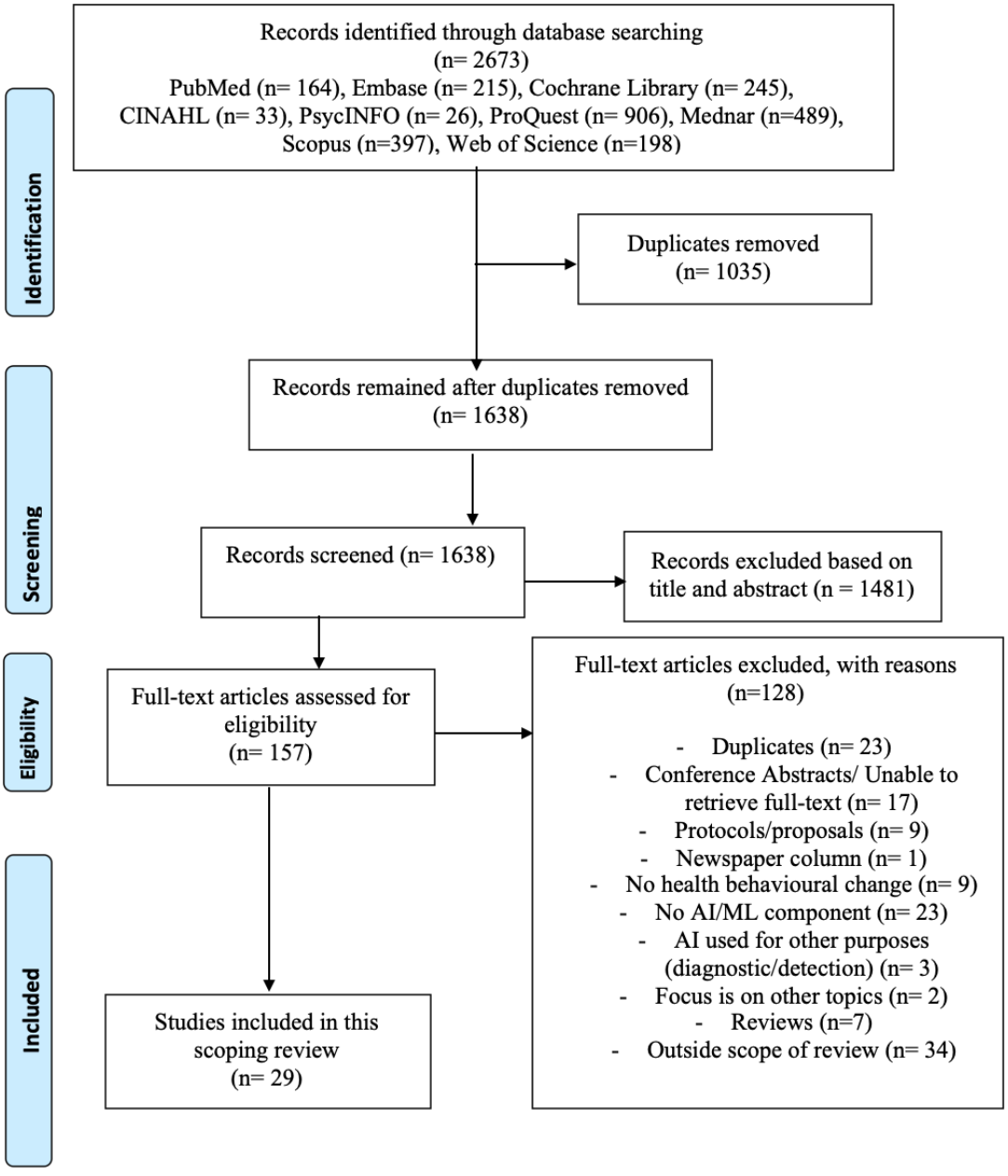
For each study, data extraction was conducted by one reviewer (JOY) and independently verified by a second (YSG). The following information was tabulated: author, year of publication, aim, main findings, area of health promotion, user data analyzed, type of ML used, challenges encountered, and future research. Such information served to inform the subsequent formation of themes in this review.

Results

Article Selection

The literature search concluded in February 2021 and yielded 2673 search results. The removal of 1035 duplicates was followed by the screening of the titles and abstracts, during which another 1481 articles were removed. Of the remaining 157 articles included in full-text screening, 128 (82%) were removed with reasons such as duplicated articles, not using ML as main components, or not focusing on health behavioral change, leaving 29 (18.4%) for inclusion in this review (Figure 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram [16].



XSL•FO
RenderX

Characteristics of the Articles

The included articles were published between 2001 and 2020. The domains of health promotion presented in the articles were diverse, which are as follows: physical activity (9/29, 31%); smoking cessation (4/29, 14%); physical activity and dietary habits (4/29, 14%); stress detection or management (2/29, 7%); weight management (2/29, 7%); physical activity and sun-protection behavior (1/29, 3%); general self-health management (1/29, 3%); self-management for mental health (1/29, 3%); self-health management for hypertension (1/29, 3%); self-health management for sickle cell disease (1/29, 3%); help-seeking behavior for heart attack (1/29, 3%); diabetes prevention (1/29, 3%); and mindfulness meditation (1/29, 3%; Multimedia Appendix 2). Three themes emerged from the findings, which are as follows: (1) enablers, which it is the adoption of information technology for optimizing systemic operation; (2) challenges, which are the discussion of various hurdles and limitations presented in the articles; and (3) future directions, which explore prospective strategies in health promotion through ML.

Themes

Theme 1: Enablers

Two key areas—the type of ML or AI technology used and data analyzed—have been identified as crucial to the implementation of the applications. Several studies used supervised ML techniques and algorithms in their designs [20-23]; in these studies, labelled data sets such as support vector machines [20], Markov logic network models [21], Naïve Bayes classifiers [23], and decision tree algorithms [22] were used to train algorithms for the purpose of classification or prediction. One study used unsupervised ML algorithms such as the Euclidean distance similarity algorithm [24], which identified hidden patterns in unlabeled data sets to recommend a best-fit message for users. Another study employed the multiarmed bandit model [25], a type of reinforcement ML technique that maximized cumulative reward to help users reach their behavioral goals. However, most studies either have not specified the type of ML or have stated only the use of rule-based algorithms [26-39].

Among these studies, several incorporated AI virtual agents such as embodied conversational agents or relational agents [22,31,34,38,40], while others incorporated chatbots in their designs [28,39,41,42]. The deployment of virtual agents allowed feedback and data output from ML algorithms to be presented to users and was typically designed to mimic human-like appearances or facial expressions. The data analyzed in the 29 studies included self-reports [20,24,26,29,32-37,39,43-48], GPS or wearable sensor data [20,23,25,29,31,36,39,45,46], application-generated data such as conversations with chatbots or users' interactions in applications [21-24,27-29,31,34,35,38-42,45,46], and clinical data and health records [21,49].

Theme 2: Challenges

Two subthemes were generated to cluster the different challenges identified from the studies, which are (1) application- or ML-related challenges and (2) methodological challenges from current studies. Additionally, 4/29 (14%) studies [24,28,31,37] have not discussed any challenges and were thus excluded from this theme.

Application- or ML-Related challenges

Many studies have ascertained difficulties specific to their application or ML models. One such challenge was related to usability. In some studies, usability has been considered beforehand and thus would usually not complicate subsequent testing [28,44,45]. However, in 1/29 (3%) study targeting physical activity [29], problems with usability were noted because of space constraints in the participants' homes and the need to switch between various digital systems in their homes. As the study involved older adults who might have had difficulties navigating complicated digital devices, the system's usability warranted improvement.

Another challenge concerned the degree of personalization in health-behavior applications, as underlined in 4/29 (14%) studies [25,27,45,47]. A lack of personalization resulted in application-generated suggestions that were not tailored to the user [25]; it also resulted in the repetition of questions [27]. In addition, overpersonalization, intruding into the individual's personal life, might breach confidentiality and privacy [45]. A similar challenge was the degree of automation in the applications, as highlighted by Block et al [49] and Traficante [47]. While the pilot test of their intervention demonstrated success in reducing diabetes-related factors, they suggested the caveat that the addition of human support might be more beneficial for specific users. Thus, the degrees of personalization and of automation in these health-behavior applications warrant prudent consideration of the balance between maximizing user benefits and minimizing drawbacks.

The accuracy of self-reported outcome measurements represented another hurdle, as reflected in 6/29 (21%) studies [34,35,39,43,44,48]. While some adopted a mixture of objective and self-reported outcomes [35,39,44], others collected only self-reported ones [34,43,48]. Such outcomes typically attempted to characterize health behaviors that were difficult to quantify, such as physical activity and dietary intake, for which the accuracy might thus be of concern. Furthermore, 1/29 (3%) study [20] reported the obtrusiveness of body sensors as a downside for its objective outcome measurements; to address this, the authors reduced the number of such sensors and monitored only the participants' computer- and posture-based behavioral patterns.

Challenges also arose from specific components of the system, such as robot-programming [27], training ML models used in the applications [40,42], and tracking users' lifestyles [32,36]. In robot-programming, the principal difficulty lay in ensuring human-like attributes in various robotic modules [27]. Such aspects included speech (intonation and speed), appearance (breathing, fidgeting, eye color, and face tracking), and fluency of interactions with users (comprehension of speech) [27]. Additionally, certain human aspects such as summarizing and reflecting upon the subjects' meanings could not be replicated by the robot [27].

In training ML models used in the applications, the principal difficulty lay in the need for substantial input, as identified in

XSL•FO
RenderX

2/29 (7%) studies [40,42]. In both studies, their natural language models enabled chatbots or virtual coaches to comprehend users' utterances and to respond appropriately. Given the prerequisite of considerable training for conversational fluency, users might be frustrated with the inability of the models to understand them [40]. To mitigate this, multiple-choice–based options, which were more effective than free-form natural language input, were added [40].

Lastly, in tracking users' lifestyles, the principal difficulty lies in the inconvenience for the users, as highlighted in 2/29 (7%) studies [32,36]. The use of ecological momentary assessment with repeated logging to obtain the users' data (eg, physical activity and dietary intake), was found to be both time-consuming and inconvenient [32]. Thus, Maimone et al [32] connected their tracking system to wearable devices with sensors and external functions (eg, weather forecasting and maps of bus stops). This not only minimized the burden on the users, but also added contextualized details to their input. Rahmani et al [36] took a step further by creating a personal chronicle ("Personicle") of the users' daily activities, integrating data from multiple streams such as sensor data, smartphone apps (eg, calendar, to-do lists, and social media), and ambient sounds.

Methodological Challenges From Current Studies

Apart from application- or ML-related challenges, methodological ones were also encountered. Of the 25 reviewed studies, 15 (60%) [21-23,29,33,35,38,39,41,43-48] cited challenges such as restricted generalizability due to small sample sizes, limited data collection, inadequate demographic representation, and the choice of a laboratory setting (ie, lack of ecological validity). Others included short testing periods [22,25,34,49] and limited outcomes due to the use of unrefined prototypes or algorithms [26,35,41,45,46]. These challenges were unsurprising since many of the included works were pilot, feasibility, or usability studies. Developers were often faced with limited resources, resulting in small-scale investigations with prototypes either of low fidelity or pending refinement. In such studies, late-stage participants tended to have better experiences since the cumulative addition of user data would enable both ongoing adjustments by the developers and continuous systemic improvements by the ML algorithms [41].

Additionally, difficulties were observed in ensuring the quality of collected data [20,39], in evaluative comparisons due to the lack of a control group [39,43,44,48], and in determining causality [39]. In ensuring data quality, the main difficulty pertained to errors in data input such as sensor failures or transmission faults [20] and in data entry by users [39]. For example, in the study by Stein and Brooks [39], data of physical activity collected through sensors could be altered by the users while data of dietary intake were manually entered into the app by them. Such data collection might have compromised the data quality when incomplete or inaccurate input occurred due to the users' oversight. Lastly, the difficulty in determining causality has been reported by Stein and Brooks [39] due to the potential presence of confounding factors such as the subjects' simultaneous engagement in other weight-loss interventions.

Theme 3: Future Directions

This review has provided valuable insights into ML-related research, including improvements on how ML-based technologies were being proposed for future health-promotion applications. The majority of the studies [20-22,24-29,31-34,36-49] have suggested the effectiveness of ML in health promotion and behavioral changes. However, some studies were unable to provide any statistical differences in their outcomes when comparing ML-based and typical applications [23,35]. Therefore, to address some of the said challenges and, accordingly, to present areas for development for future works, the following two subthemes were generated: (1) future studies and (2) future application of ML. Two studies [24,28] have not provided suggestions for any future research, which resulted in their exclusion from this theme.

Future Studies

This subtheme examines two areas that may be addressed—methodological designs and potential areas of research—to inform future studies on ML. In terms of methodological designs, many studies have underscored the need for longitudinal investigations to allow for more data collection for ML models and to determine whether the intervention-induced behavioral changes are sustained over prolonged periods [20,25,30,33,44,49]. This is especially true in order to rule out the novelty effect of using ML as an intervention for health behavioral change. Furthermore, additional buffer time was recommended for participants to accustom themselves to the app before the study [27]. Additionally, potential confounding variables should be accounted for by screening participants beforehand [39] and by broadening demographic representation, especially for age, ethnicity, and socioeconomic status [26,27,33,38,39,41,48,49]. Lastly, all stakeholders' perspectives should be considered to holistically appraise the acceptability and usability of the application [41,47].

In terms of potential areas of research, suggestions for future studies are numerous. Studies could consider examining the effects of various measures, such as the following: notification content and purpose on users' receptivity and response rates [23]; complex graphics such as games and chatbots in users' engagement during behavioral changes [46]; and the use of digital services as active participants in users' interactions [46]. Furthermore, research may be undertaken for tailoring variables to address smokers in the precontemplation stage [37], for comparing a simultaneous intervening mode with a sequential mode for physical activity and fat-intake interventions [43], for examining impacts of the emotivity of virtual agents [22], and for adding the element of user control to ML-generated suggestions [25]. Lastly, studies could also consider testing their app on other health behaviors [21,23,26,42].

Future Application of ML

Apart from addressing the identified challenges, several opportunities for future works have been highlighted for further developing and refining ML in health promotion and behavioral changes. Firstly, given the challenges of users' inconvenience and human- and sensor-related errors [20,34,35,39,43,44,48], effective data-input modalities that require minimal users' effort

XSL•FO
RenderX

and encourage continuous engagement warrant greater attention [32]. Future works may accordingly consider developing systems that detect sensor failures [20] and those that account for erroneous self-reported measures [34]. Secondly, validated outcome measures and surveys have been found to exhibit low sensitivity in detecting subtle day-to-day behavioral changes, as exemplified by the short-term interventions in one study [34]. Future works may accordingly consider further examining the effectiveness of these interventions. Thirdly, ML-based applications offer the potential to be integrated into health care delivery systems [21]. However, care must be taken to ensure the reliability and safety of the information thus provided to users [21]. Lastly, Kulyk et al [31] have emphasized the need for guidelines and standardization for evaluating health technologies. They suggested multidisciplinary approaches and independent evaluators to examine the effects of interventions at all developmental stages of the app, not only to encourage continued engagement and motivation, but also to meet the target users' needs.

Discussion

Principal Findings

This review has provided an insightful overview of the use of ML technologies in health promotion and behavioral changes, alongside a discussion of the challenges and potential opportunities for future works. Barring 4 studies [22,37,43,47], most others have been conducted over the past decade, indicating a recent growing interest in this topic. Additionally, the majority of the interventions in this review involved physical activity, whereas those specific to certain illness or those targeting smoking or mental health were relatively scarce. Hence, such interventions involving physical activity may be of interest for future works.

One noteworthy finding across the reviewed studies was the amount of time required to develop their ML-based applications. Some of the studies [20-25] specified the methods employed, including supervised ML, unsupervised ML, and reinforcement ML techniques. These differed in their use of supervised (labelled) data sets, which needed additional human intervention, and unsupervised (unlabeled) data sets. Unsupervised ML, coupled with the incorporation of virtual agents, embodied conversational agents, and human-like robots, required substantial time and resources to program natural language processing and to train the ML models [27,40,42]. Given the need for such length of time, the methodological implication was that most of the reviewed studies involved prototype research and pilot studies.

Furthermore, the complexity and comprehensiveness of ML-based applications grew with increasing amounts of data fed to their algorithms. The practical implication was that the experience differed between early-stage participants (who might experience a more rudimentary system) and their late-stage counterparts (who experienced a more refined one) [41]. These findings indicated that, to appraise the impacts of such ML-based applications more prudently and confidently, more studies examining them in a later, more mature, developmental stage would be warranted. Accordingly, developers might consider structured data input such as multiple-choice–based options [40] to aid the ML systems during the training of the application to refine and improve the accuracy of the feedback. These unique aspects of ML-based applications would deserve consideration in future studies incorporating ML techniques in their designs.

Another notable finding concerned hurdles in the design of the studies that might compromise their results, namely the lack of control groups [39,43,44,48], limited periods or samples for data collection [21-23,29,33,35,38,39,41,43-48], and potential confounders such as participation in other similar interventions [39]. To improve study designs and minimize the risk of confounders, the deployment of a waiting list control group and the disclosure of participation in other similar interventions should be instituted as part of the study methodology. Additionally, for short-term pilot and prototype studies with limited data collection, follow-up longitudinal studies might provide not only more data that would benefit the training of the ML system, but also insights into potential lasting effects of the intervention. Lastly, against such a background of these hurdles, health technologies would benefit from a set of guidelines and standardization, as advocated by Kulyk et al [31], for constant evaluation at all developmental stages to maximize the users' potential to meet the target health behaviors.

Two other aspects merited discussion, which are as follows: (1) the burden on users to provide accurate data for tracking their lifestyles through self-reporting and the use of sensors; and (2) the challenges surrounding personalization. Although systems detecting sensor failures [20] and erroneous self-reported measures [34] might enhance the accuracy of behavioral measurement outcomes, further work in this area of research would be needed, especially with the increase in sensor usage in recent years [50]. Moreover, methods to reduce sensor failures through hardware enhancements could improve the quality of data. Additionally, both under- and overpersonalization would lead to overly generic feedback for users and breaches of privacy; this aspect thus warranted developers' prudent and sensitive handling to strike a balance between providing customization and upholding privacy. Taken together, the challenges identified by the reviewed studies have underscored the need for a design that is nonobtrusive, requires minimal users' input, encourages continuous engagement, and enables personalization while not encroaching upon privacy. Alongside such efforts, greater investment in combining multiple ML and AI techniques for synergy should likewise be pursued.

Limitations

This review offers a timely overview on the emerging field of ML in health technologies, alongside potential referential value for future works. Nonetheless, several limitations are noteworthy. Firstly, the findings may not be generalizable since this review has included studies in only English due to the lack of access to interpreters. Additionally, the predefined scope of this review meant that studies examining other areas of health, such as detecting and diagnosing medical conditions, were excluded. Secondly, despite the use of numerous combinations to attempt to encapsulate the concept of ML in our search strategy for this review, the wide variety of ML techniques,

XSL•FO
RenderX

alongside their overlap with statistical approaches [19], might have resulted in an omission of eligible studies. Finally, despite the demonstrable viability of ML techniques in health promotion and behavioral changes and their potential for integration into health care systems, most of the reviewed works have been pilot studies or prototype research, for which more investigations would thus be warranted to determine the clinical utility of the interventions.

Conclusions

This review has examined the use of ML in health promotion and behavioral changes by mapping its associated challenges and highlighting potential areas for future works. The findings have collectively demonstrated that the challenges pertained to not only the time- and resource-consuming nature of ML-based applications, but also problems such as the burden on users for data input and the degree of personalization. Future works may consider designs that correspondingly mitigate these challenges in areas such as mental health promotion where the use of ML remains limited.

Authors' Contributions

All authors contributed equally from conceptualization to data collection, data analysis, and the eventual drafting of the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Keywords.
[DOCX File , 21 KB - jmir_v24i6e35831_app1.docx]

Multimedia Appendix 2
Characteristics of the included studies.
[DOCX File , 45 KB - jmir_v24i6e35831_app2.docx]

References

1. Buttorff C, Ruder T, Bauman M. Multiple Chronic Conditions in the United States. Santa Monica, CA: RAND Corporation; 2017. [doi: 10.7249/TL221]
2. GBD 2015 MortalityCauses of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016 Oct 08;388(10053):1459-1544 [FREE Full text] [doi: 10.1016/S0140-6736(16)31012-1] [Medline: 27733281]
3. Di Renzo L, Gualtieri P, De Lorenzo A. Diet, Nutrition and Chronic Degenerative Diseases. Nutrients 2021 Apr 20;13(4):1372 [FREE Full text] [doi: 10.3390/nu13041372] [Medline: 33923865]
4. Middleton KR, Anton SD, Perri MG. Long-Term Adherence to Health Behavior Change. Am J Lifestyle Med 2013;7(6):395-404 [FREE Full text] [doi: 10.1177/1559827613488867] [Medline: 27547170]
5. Jaam M, Awaisu A, Ibrahim MI, Kheir N. Synthesizing and Appraising the Quality of the Evidence on Factors Associated with Medication Adherence in Diabetes: A Systematic Review of Systematic Reviews. Value Health Reg Issues 2017 Sep;13:82-91. [doi: 10.1016/j.vhri.2017.09.001] [Medline: 29073997]
6. Haskard-Zolnierek K, Martin L, DiMatteo M, Williams S. AdherenceHealth Behavior Change in the Context of Mental Health Challenges. In: The Oxford Handbook of Health Communication, Behavior Change, and Treatment Adherence. New York, US: Oxford University Press; 2013.
7. Pesantes MA, Tetens A, Valle AD, Miranda JJ. "It is Not Easy Living with This Illness": A Syndemic Approach to Medication Adherence and Lifestyle Change among Low-income Diabetes Patients in Lima, Peru. Human Organization 2019 Mar;78(1):85-96. [doi: 10.17730/0018-7259.78.1.85]
8. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science 2015 Jul 17;349(6245):255-260. [doi: 10.1126/science.aaa8415] [Medline: 26185243]
9. Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. Psychol Med 2019 Jul;49(9):1426-1448. [doi: 10.1017/S0033291719000151] [Medline: 30744717]
10. Faedda GL, Ohashi K, Hernandez M, McGreenery CE, Grant MC, Baroni A, et al. Actigraph measures discriminate pediatric bipolar disorder from attention-deficit/hyperactivity disorder and typically developing controls. J Child Psychol Psychiatry 2016 Jun 22;57(6):706-716 [FREE Full text] [doi: 10.1111/jcpp.12520] [Medline: 26799153]

XSL•FO
RenderX

11. Chary M, Genes N, Giraud-Carrier C, Hanson C, Nelson LS, Manini AF. Epidemiology from Tweets: Estimating Misuse of Prescription Opioids in the USA from Social Media. J Med Toxicol 2017 Dec 22;13(4):278-286 [FREE Full text] [doi: 10.1007/s13181-017-0625-5] [Medline: 28831738]
12. Ye Z, Rae CL, Nombela C, Ham T, Rittman T, Jones PS, et al. Predicting beneficial effects of atomoxetine and citalopram on response inhibition in Parkinson's disease with clinical and neuroimaging measures. Hum Brain Mapp 2016 Mar 12;37(3):1026-1037 [FREE Full text] [doi: 10.1002/hbm.23087] [Medline: 26757216]
13. Hu B, Villazón-Terrazas B. Building a Mental Health Knowledge Model to Facilitate Decision Support. 2016 Presented at: Knowledge Management and Acquisition for Intelligent Systems: 14th Pacific Rim Knowledge Acquisition Workshop, PKAW 2016; August 22-23, 2016; Phuket, Thailand p. 198-212. [doi: 10.1007/978-3-319-42706-5_15]
14. Sapci AH, Sapci HA. Innovative Assisted Living Tools, Remote Monitoring Technologies, Artificial Intelligence-Driven Solutions, and Robotic Systems for Aging Societies: Systematic Review. JMIR Aging 2019 Nov 29;2(2):e15429 [FREE Full text] [doi: 10.2196/15429] [Medline: 31782740]
15. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
16. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Sep 04;169(7):467. [doi: 10.7326/M18-0850]
17. Tam WWS, Lo KKH, Khalechelvam P, Seah J, Goh SYS. Is the information of systematic reviews published in nursing journals up-to-date? a cross-sectional study. BMC Med Res Methodol 2017 Nov 25;17(1):151 [FREE Full text] [doi: 10.1186/s12874-017-0432-3] [Medline: 29178832]
18. Whittemore R, Knafl K. The integrative review: updated methodology. J Adv Nurs 2005 Dec;52(5):546-553. [doi: 10.1111/j.1365-2648.2005.03621.x] [Medline: 16268861]
19. Bi Q, Goodman K, Kaminsky J, Lessler J. What is Machine Learning? A Primer for the Epidemiologist. Am J Epidemiol 2019 Dec 31;188(12):2222-2239. [doi: 10.1093/aje/kwz189] [Medline: 31509183]
20. Alberdi A, Aztiria A, Basarab A, Cook DJ. Using smart offices to predict occupational stress. International Journal of Industrial Ergonomics 2018 Sep;67:13-26. [doi: 10.1016/j.ergon.2018.04.005]
21. Chen S, Guo X, Ju X. The design of personalized artificial intelligence diagnosis and the treatment of health management systems simulating the role of general practitioners. 2018 Presented at: International Conference on Smart Health (ICSH); July 133, 2018; Wuhan, China p. 26-40. [doi: https://doi.org/10.1007/978-3-030-03649-2_3]
22. Silverman B, Holmes J, Kimmel S, Branas C, Ivins D, Weaver R, et al. Modeling emotion and behavior in animated personas to facilitate human behavior change: the case of the HEART-SENSE game. Health Care Manag Sci 2001 Sep;4(3):213-228. [doi: 10.1023/a:1011448916375] [Medline: 11519847]
23. Morrison LG, Hargood C, Pejovic V, Geraghty AWA, Lloyd S, Goodman N, et al. The Effect of Timing and Frequency of Push Notifications on Usage of a Smartphone-Based Stress Management Intervention: An Exploratory Trial. PLoS ONE 2017 Jan 3;12(1):e0169162. [doi: 10.1371/journal.pone.0169162]
24. Ahsan G, Addo I, Ahamed S, Petereit D, Kanekar S, Burhansstipanov L, et al. Toward an mHealth Intervention for Smoking Cessation. Proc COMPSAC 2013:345-350 [FREE Full text] [doi: 10.1109/COMPSACW.2013.61] [Medline: 24172662]
25. Rabbi M, Pfammatter A, Zhang M, Spring B, Choudhury T. Automated personalized feedback for physical activity and dietary behavior change with mobile phones: a randomized controlled trial on adults. JMIR Mhealth Uhealth 2015 May 14;3(2):e42 [FREE Full text] [doi: 10.2196/mhealth.4160] [Medline: 25977197]
26. Brigham J, Javitz H, Krasnow R, Jack L, Swan G. A Tailoring Algorithm to Optimize Behavior Change. 2014 Presented at: 47th Hawaii International Conference on System Sciences; 10 March 2014; Waikoloa, HI, USA. [doi: 10.1109/HICSS.2014.334]
27. Galvão Gomes da Silva J, Kavanagh DJ, Belpaeme T, Taylor L, Beeson K, Andrade J. Experiences of a Motivational Interview Delivered by a Robot: Qualitative Study. J Med Internet Res 2018 May 03;20(5):e116 [FREE Full text] [doi: 10.2196/jmir.7737] [Medline: 29724701]
28. Issom D, Rochat J, Hartvigsen G, Lovis C. Preliminary Evaluation of a mHealth Coaching Conversational Artificial Intelligence for the Self-Care Management of People with Sickle-Cell Disease. Stud Health Technol Inform 2020 Jun 16;270:1361-1362. [doi: 10.3233/SHTI200442] [Medline: 32570659]
29. Jimison H, Hagler S, Kurillo G, Bajcsy R, Pavel M. Remote health coaching for interactive exercise with older adults in a home environment. 2015 Presented at: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); August 25-29, 2015; Milan, Italy. [doi: 10.1109/embc.2015.7319633]
30. King A, Holder MG, Ahmed RA. Errors as allies: error management training in health professions education. BMJ Qual Saf 2013 Jun;22(6):516-519. [doi: 10.1136/bmjqs-2012-000945] [Medline: 23293120]
31. Kulyk O, Op DAR, Klaasseny R, van GL. Personalized virtual coaching for lifestyle support: Principles for design and evaluation. International Journal on Advances in Life Sciences 2014;6(3 & 4):300-309.
32. Maimone R, Guerini M, Dragoni M, Bailoni T, Eccher C. PerKApp: A general purpose persuasion architecture for healthy lifestyles. J Biomed Inform 2018 Jun;82:70-87 [FREE Full text] [doi: 10.1016/j.jbi.2018.04.010] [Medline: 29729482]

XSL•FO
RenderX

33. Martin SS, Feldman DI, Blumenthal RS, Jones SR, Post WS, McKibben RA, et al. mActive: A Randomized Clinical Trial of an Automated mHealth Intervention for Physical Activity Promotion. J Am Heart Assoc 2015 Nov 09;4(11):e002239 [FREE Full text] [doi: 10.1161/JAHA.115.002239] [Medline: 26553211]
34. Mohan S. Exploring the Role of Common Model of Cognition in Designing Adaptive Coaching Interactions for Health Behavior Change. URL: https://arxiv.org/pdf/1910.07728.pdf [accessed 2022-05-17]
35. 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]
36. Rahmani A, Lai J, Jafarlou S, Asal Y, Rivera A, Labbaf S. Personal Mental Health Navigator: Harnessing the Power of Data, Personal Models, and Health Cybernetics to Promote Psychological Well-being. Cornell University Library, arXiv.org 2020:1-24.
37. Schumann A, John U, Ulbricht S, Rüge J, Bischof G, Meyer C. Computer-generated tailored feedback letters for smoking cessation: theoretical and empirical variability of tailoring. Int J Med Inform 2008 Nov;77(11):715-722. [doi: 10.1016/j.ijmedinf.2008.03.001] [Medline: 18417417]
38. Sillice MA, Morokoff PJ, Ferszt G, Bickmore T, Bock BC, Lantini R, et al. Using Relational Agents to Promote Exercise and Sun Protection: Assessment of Participants' Experiences With Two Interventions. J Med Internet Res 2018 Feb 07;20(2):e48 [FREE Full text] [doi: 10.2196/jmir.7640] [Medline: 29415873]
39. Stein N, Brooks K. A Fully Automated Conversational Artificial Intelligence for Weight Loss: Longitudinal Observational Study Among Overweight and Obese Adults. JMIR Diabetes 2017 Nov 01;2(2):e28 [FREE Full text] [doi: 10.2196/diabetes.8590] [Medline: 30291087]
40. Hudlicka E. Virtual training and coaching of health behavior: example from mindfulness meditation training. Patient Educ Couns 2013 Aug;92(2):160-166 [FREE Full text] [doi: 10.1016/j.pec.2013.05.007] [Medline: 23809167]
41. Stephens TN, Joerin A, Rauws M, Werk LN. Feasibility of pediatric obesity and prediabetes treatment support through Tess, the AI behavioral coaching chatbot. Transl Behav Med 2019 May 16;9(3):440-447. [doi: 10.1093/tbm/ibz043] [Medline: 31094445]
42. Almusharraf F. Motivating Smokers to Quit Through a Computer-Based Conversational System. University of Toronto. 2019. URL: https://tspace.library.utoronto.ca/bitstream/1807/94055/1/Almusharraf_Fahad_201903_MAS_thesis.pdf [accessed 2022-05-17]
43. Vandelanotte C, De Bourdeaudhuij I, Brug J. Two-year follow-up of sequential and simultaneous interactive computer-tailored interventions for increasing physical activity and decreasing fat intake. ann. behav. med 2007 Jun;33(2):213-219. [doi: 10.1007/bf02879903]
44. 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]
45. Monteiro-Guerra F, Signorelli GR, Tadas S, Dorronzoro Zubiete E, Rivera Romero O, Fernandez-Luque L, et al. A Personalized Physical Activity Coaching App for Breast Cancer Survivors: Design Process and Early Prototype Testing. JMIR Mhealth Uhealth 2020 Jul 15;8(7):e17552 [FREE Full text] [doi: 10.2196/17552] [Medline: 32673271]
46. Nurmi J, Knittle K, Ginchev T, Khattak F, Helf C, Zwickl P, et al. Engaging Users in the Behavior Change Process With Digitalized Motivational Interviewing and Gamification: Development and Feasibility Testing of the Precious App. JMIR Mhealth Uhealth 2020 Jan 30;8(1):e12884 [FREE Full text] [doi: 10.2196/12884] [Medline: 32003750]
47. Traficante R. Computer -based interventions, health behavior change, and ethics. Dissertation. Salve Regina University. 2004. URL: https://digitalcommons.salve.edu/dissertations/AAI3133900 [accessed 2022-05-18]
48. King AC, Hekler EB, Grieco LA, Winter SJ, Sheats JL, Buman MP, et al. Harnessing different motivational frames via mobile phones to promote daily physical activity and reduce sedentary behavior in aging adults. PLoS One 2013;8(4):e62613 [FREE Full text] [doi: 10.1371/journal.pone.0062613] [Medline: 23638127]
49. Block G, Azar KM, Romanelli RJ, Block TJ, Hopkins D, Carpenter HA, et al. Diabetes Prevention and Weight Loss with a Fully Automated Behavioral Intervention by Email, Web, and Mobile Phone: A Randomized Controlled Trial Among Persons with Prediabetes. J Med Internet Res 2015 Oct 23;17(10):e240 [FREE Full text] [doi: 10.2196/jmir.4897] [Medline: 26499966]
50. Dei M, Aymerich J, Piotto M, Bruschi P, del Campo F, Serra-Graells F. CMOS Interfaces for Internet-of-Wearables Electrochemical Sensors: Trends and Challenges. Electronics 2019 Jan 31;8(2):150. [doi: 10.3390/electronics8020150]

Abbreviations

AI: artificial intelligence
MeSH: Medical Subject Headings
ML: machine learning
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Scoping Reviews

XSL•FO
RenderX

Edited by G Eysenbach; submitted 19.12.21; peer-reviewed by JA Benítez-Andrades, A Joseph; comments to author 22.03.22; revised version received 23.03.22; accepted 23.03.22; published 02.06.22.

Please cite as:
Goh YS, Ow Yong JQY, Chee BQH, Kuek JHL, Ho CSH
Machine Learning in Health Promotion and Behavioral Change: Scoping Review
J Med Internet Res 2022;24(6):e35831
URL: https://www.jmir.org/2022/6/e35831
doi:10.2196/35831
PMID:35653177

©Yong Shian Goh, Jenna Qing Yun Ow Yong, Bernice Qian Hui Chee, Jonathan Han Loong Kuek, Cyrus Su Hui Ho. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 02.06.2022. 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

Computerized Psychological Interventions in Veterans and Service Members: Systematic Review of Randomized Controlled Trials

Rahel Pearson[1], PhD; Emily Carl[2], MA; Suzannah K Creech[1,3], PhD

[1]Veterans Health Administration Veterans Integrated Service Network 17, Center of Excellence for Research on Returning War Veterans, Central Texas Veterans Affairs Healthcare System, Waco, TX, United States

[2]Department of Psychology, University of Texas, Austin, TX, United States

[3]Department of Psychiatry and Behavioral Sciences, Dell Medical School of the University of Texas, Austin, TX, United States

Corresponding Author:
Rahel Pearson, PhD
Veterans Health Administration Veterans Integrated Service Network 17
Center of Excellence for Research on Returning War Veterans
Central Texas Veterans Affairs Healthcare System
4800 Memorial Drive (151C)
Waco, TX, 76711
United States
Phone: 1 2542975155
Email: rahel.pearson@va.gov

Abstract

Background: Computerized psychological interventions can overcome logistical and psychosocial barriers to the use of mental health care in the Veterans Affairs and Department of Defense settings.

Objective: In this systematic review, we aim to outline the existing literature, with the goal of describing: the scope and quality of the available literature, intervention characteristics, study methods, study efficacy, and study limitations and potential directions for future research.

Methods: Systematic searches of two databases (PsycINFO and PubMed) using PRISMA (Preferred Reporting Item for Systematic Reviews and Meta-Analyses) guidelines were conducted from inception until November 15, 2020. The following inclusion criteria were used: the study was published in an English language peer-reviewed journal, participants were randomly allocated to a computerized psychological intervention or a control group (non–computerized psychological intervention active treatment or nonactive control group), an intervention in at least one treatment arm was primarily delivered through the computer or internet with or without additional support, participants were veterans or service members, and the study used validated measures to examine the effect of treatment on psychological outcomes.

Results: This review included 23 studies that met the predefined inclusion criteria. Most studies were at a high risk of bias. Targeted outcomes, participant characteristics, type of support delivered, adherence, and participant satisfaction were described. Most of the examined interventions (19/24, 79%) yielded positive results. Study limitations included participant characteristics limiting study inference, high rates of attrition, and an overreliance on self-reported outcomes.

Conclusions: Relatively few high-quality studies were identified, and more rigorous investigations are needed. Several recommendations for future research are discussed, including the adoption of methods that minimize attrition, optimize use, and allow for personalization of treatment.

(J Med Internet Res 2022;24(6):e30065) doi:10.2196/30065

KEYWORDS

computer; digital; internet; interventions; veterans; service members; review; mobile phone

Introduction

Most individuals with diagnosable mental health disorders do not have access to adequate care [1]. Computerized psychological interventions are well-positioned to address this treatment gap, as these interventions provide a cost-effective and easily accessible alternative to traditional face-to-face mental health care [2,3]. Computerized psychological interventions, often delivered through the internet, have grown

XSL•FO
RenderX

steadily in popularity over the past decade. Computerized psychological interventions may be effective for the treatment of a range of mental health disorders. Meta-analyses have supported the efficacy of computerized psychological interventions in the treatment of depression [4-6], anxiety [4,7], posttraumatic stress disorder (PTSD; [8]), substance use disorders (SUDs; [9,10]), and insomnia [11], with effect sizes ranging from small [5,7,9] to medium-large [4,6,11]. Various meta-analyses have outlined the therapeutic benefit of specific computer-delivered treatment modalities, such as cognitive behavioral therapy (CBT) [4,11], acceptance and commitment therapy [12], and mindfulness-based therapy [13]. There is evidence that computerized psychological interventions are as effective as face-to-face interventions [4,14], suggesting that a variety of mental health conditions can be addressed by computerized treatments with the potential for widespread dissemination of such therapies.

As the body of evidence supporting computerized psychological interventions is robust, it is not surprising that the Department of Veterans Affairs (VA) and the Department of Defense have increasingly implemented these interventions [15], with increasing studies examining computerized psychological interventions in current and former service members. To our knowledge, there has been no systematic review of the literature examining computerized psychological interventions in veterans and service members. Reviews of computerized psychological interventions in general community samples might not generalize to veterans and service members, given the unique considerations in terms of gender ratio, severity of comorbid conditions, socioeconomic factors, and situational or environmental exposures. The use of computerized psychological interventions in veterans and service members will likely continue to rise. The objective of this systematic review is to examine the literature on randomized controlled trials (RCTs) using computerized psychological interventions in veteran and military populations by describing study characteristics and summarizing the efficacy of these interventions. We also aim to examine the quality and limitations of the identified studies. Finally, recommendations for future research will be made.

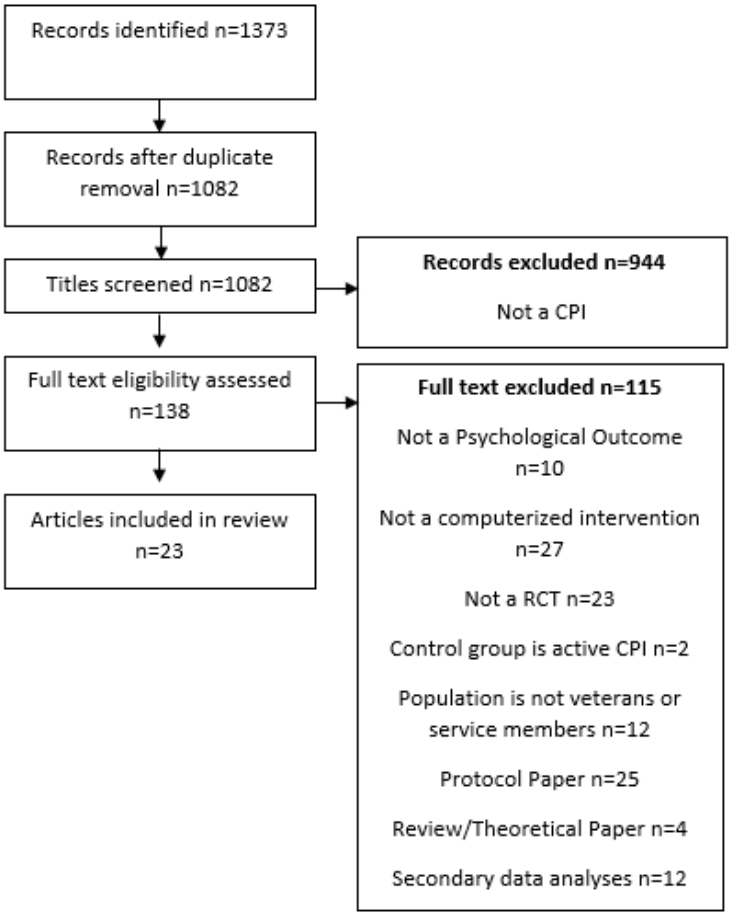
Methods

Study Selection and Data Collection

Systematic searches of two databases (PsycINFO and PubMed) were conducted from inception until November 15, 2020. The search terms are described in Multimedia Appendix 1 (adapted from Moore et al [16]). Duplicates were removed, and the reference lists of the included studies were examined for additional articles. A final list of included studies was circulated among colleagues with subject matter expertise to verify that no relevant papers were omitted. The following inclusion criteria were used: (1) the study was published in an English language peer-reviewed journal, (2) participants were randomly allocated to a computerized psychological intervention or a control group (non–computerized psychological intervention active treatment or nonactive control group), (3) an intervention in at least one treatment arm was primarily delivered through the computer or internet with or without additional support, (4) participants were military veterans or service members, and (5) the study used validated measures to examine the effect of treatment on psychological outcomes. Validated psychological outcome measures are those measures with demonstrated reliability and validity that quantify mood, well-being, emotion, affect, and/or psychosocial functioning. The outcomes examined were mental health disorders, as defined by the Diagnostic Statistical Manual, fifth edition (eg, SUDs and neurocognitive disorders), as well as psychosocial and behavioral correlates of these disorders (eg, romantic relationship dysfunction and anger). As commonly used VA smartphone apps are designed to be used in conjunction with traditional face-to-face treatment [17], they were not included. Papers were reviewed for inclusion at the title, abstract, and full paper levels. RP determined if studies met the inclusion criteria, and any ambiguity was discussed with SC until a final decision was reached. Using a standardized form, intervention characteristics, population characteristics, study design, methods, procedures, and outcomes were recorded. The number of papers identified, screened, and included is reported in Figure 1.

XSL•FO
RenderX

Figure 1. Flow chart of records identified through database screening.



Data Synthesis

Findings were grouped by the targeted psychological outcome and population examined (veterans and/or service members). These subgroups used similar methods, such as intervention techniques and recruitment tactics, allowing for a more direct comparison. Study outcomes were examined and visually presented in a harvest plot by these subgroups, as well as the control group (active or inactive), to provide a more nuanced understanding of the study results. If a study included both an active and inactive control group, the computerized psychological intervention was compared with the active control group to provide a more robust test of computerized psychological intervention efficacy. Overall, the studies were heterogeneous, and subgroups of similar studies were very small (≤3), making meta-analysis inappropriate. A narrative synthesis of the effects was conducted to describe the efficacy of interventions on primary outcome measures. Effect sizes (Cohen *d* of Hedges *g* if n<20) were extracted or calculated where possible. The authors of the included studies were contacted when insufficient data were provided in the article.

Quality Assessment

The Cochrane risk of bias tool [18] was used to determine the methodological quality of the included studies. The assessed features were sequence generation and allocation sequence concealment (selection bias), blinding of personnel (performance bias), blinding of outcome assessment (detection bias), incomplete outcome data (attrition bias), and selective outcome reporting (reporting bias). Two authors (RP and EC) independently rated the risk of bias as *high risk* (eg, sequence generation: randomization based on birthday), *low risk* (eg, sequence generation: block randomization with randomly varied blocks sizes), or *unclear risk* (eg, sequence generation: no information on randomization method). Overall bias was rated as *low* if ratings of bias were low in all domains or if bias was *unclear* in 1 domain, and this was unlikely to have biased the study outcome. Overall bias was rated as *moderate* if bias was high in 1 domain or *unclear* in 2 domains, and this was unlikely to have biased the study outcome. Overall bias was rated as *high* if bias was rated as *high* in 1 domain or *unclear* in 2 domains, and this was likely to have biased study outcomes. Similarly, overall bias was rated as *high* if bias was rated as *unclear* in ≥3 domains or *high* in ≥2 domains. Ratings were compared, and the 2 raters discussed discrepancies until a consensus was reached.

Results

Intervention Characteristics

Overview

Characteristics of the included studies are outlined in Multimedia Appendix 2 [19-42]. Most computerized psychological interventions targeted PTSD and SUDs. Most computerized psychological interventions were web-based, although some studies used software programs installed on study computers [19-23]. Intervention content was often

XSL•FO
RenderX

presented with graphics and interactive features that allowed tailored treatment content based on participant characteristics [19,24-32]. Computerized psychological interventions offered varying degrees of guidance in the 23 studies: 3 (13%) studies [24,29,33] provided no human contact during any study phase, 15 (65%) studies [19-23,25-27,31,34-39] provided administrative contact (eg, reminders, use support, and structured or semistructured assessment), and 5 (22%) studies [28,30,32,40,41] included therapeutic contact.

PTSD Interventions

Of the 24 computerized psychological interventions, 8 (33%) addressed PTSD and related symptoms. Of the 23 studies, 3 (13%) studies examined expressive writing interventions of different lengths [38-40] with PTSD symptomatology as the primary outcome, 3 (13%) studies [26,28,36] tested the efficacy of computerized psychological interventions that included a variety of CBT-based techniques, such as cognitive restructuring and behavioral modification. McLean et al [32] computerized prolonged exposure, an evidence-based treatment (EBT) for PTSD, and Larsen et al [37] examined the efficacy of cognitive training on PTSD symptoms.

SUD Interventions

Interventions addressing SUDs were primarily focused on drinking problems (3/23, 13% studies; 4/24, 17% interventions) [22,29,33] or drinking or substance use problems and PTSD [24,35]. One of the studies [25] tested the efficacy of a smoking cessation intervention. Interventions addressing only drinking problems were completed in one sitting [22,29,33]. All interventions included alcohol assessment feedback and psychoeducation on alcohol use. In addition, interventions presented peer-specific norms for alcohol use [22,29,33] and included motivational techniques [22,29]. Interventions addressing comorbid alcohol/substance use and PTSD were CBT based and significantly longer than interventions addressing alcohol use in isolation [24,35]. Both interventions integrated CBT components and motivational techniques aimed at trauma and substance or alcohol use. Acosta et al [35] also offered optional trauma exposure modules, although there was limited engagement with this content. Calhoun et al [25] tested the efficacy of a web-based smoking intervention, Quitnet, which included behavioral goal setting and social support components.

Other Interventions

Approximately 8% (2/24) of interventions targeted depression: Bedford et al [19] examined a 6-session problem-solving intervention, whereas Pfeiffer et al [30] examined Beating the Blues, a well-established CBT for depression intervention. CBT interventions were also used to treat anger [21] and insomnia [31] and decrease suicidality by targeting perceived burdensomeness and thwarted belongingness [23]. Cooper et al [20] examined a brain fitness program to ameliorate the deleterious effects of mild traumatic brain injury. Finally, 2 interventions included participants and their partners. Kahn et al [27] examined an intervention aimed at promoting postdeployment rehabilitation, which included mindfulness-based techniques and massage therapy instructions. Salivar et al [41] tested 2 interventions, combined for data analyses, for low-income couples who experienced relationship distress. Interventions integrated conflict management with improving communication, commitment, and positivity (intervention 1) or acceptance and implementing behavioral change (intervention 2).

Recruitment and Sample Size

A subset of trials recruited veterans from one [21,25,37,40] or more [22,30,34-36,38] VA facilities. Approximately 13% (3/23) of trials [22,30,35] contacted veterans who were likely to meet the study inclusion criteria based on information obtained from the VA electronic medical record system, and another sent out eligibility questionnaires and US $5 incentives to a large (N=15,686) random sample of veterans [39]. Several trials [19,24,26,27,33,34] recruited participants through targeted emails on the web (eg, emails sent out to members of veterans' associations) or through Facebook and other social media advertising. In addition, veterans were recruited from university campuses [19,34,37] and the community [23,34,37]. Approximately 30% (7/23) of studies recruited active duty personnel [20,28,29,31] from military installations and Department of Defense sites or a combination of active duty personnel and veterans [32,36,41]. Of the 23 studies, 4 (17%) studies included all or majority US army personnel [20,31,32,36], 1 (4%) study recruited from various military branches at different installations and sites [29], and 2 (9%) studies did not specify the military branches represented in their sample [28,41]. Pemberton et al [29] recruited a convenience sample at 8 military installations and was able to recruit the largest sample size (N=3070, compared with a sample range of N=40 to N=180 for other studies recruiting service members). However, the study authors noted several sample limitations, such as the low prevalence of drinking problems and high study attrition.

Web-only trials that recruited nationally through social media or by US mail were generally more successful in obtaining large sample sizes [26,27,33,38,39]. Sample sizes were generally reduced in pilot trials [34,37,38], if ≥1 study arm required intensive participant contact (eg, face-to-face treatment and magnetic resonance imaging [21,31,34]), or if eligibility criteria were stringent and/or required specialty assessment [20].

Adherence and Attrition

Adherence to the intervention was addressed in 87% (20/23) of examined studies, although only 9% (2/23) of studies clearly defined the level of engagement that differentiated adherence versus nonadherence (ie, >80% of sessions completed [37] or >5 sessions completed [30]). The remaining studies provided a quantitative assessment of participant engagement with the intervention, such as the time spent or sessions completed. Adherence, defined as completion of all study sessions, ranged from 25% to 100% (based on 12/23, 52% of studies with available data [19,22,24,29,32-36,38,40,41]).

Attrition, defined as a loss to follow-up, was reported in all studies. Across all studies, 36.25% (2994/8260) of participants were lost at the first follow-up time point, which often coincided with the posttreatment assessment. At the second follow-up

XSL•FO
RenderX

time point (average follow-up length was 16 weeks), 45.59% (3542/7770) participants across 87% (20/23) of studies were lost. When computerized psychological intervention attrition at the first follow-up time point was examined by support type, attrition was 51.78% (2311/4463) in the no support group (3/23, 13% studies; 4/24, 17% interventions), 17.29% (548/3170) in the administrative support group (15/23, 65% studies), and 21.5% (135/629) in the therapeutic support group (4/23, 17% studies).

Risk of Bias

Overview

The Cochrane risk of bias tool [18] assessment of bias consensus ratings for the included studies is found in Table 1.

Table 1. Consensus ratings for Cochrane assessment of bias.

Study	Random sequence generation	Allocation concealment	Blind outcome assessment	Incomplete data	Selective reporting	Overall risk of bias
Acosta et al [35]	Unclear	Unclear	Low	Low	Low	Moderate
Bedford et al [19]	Low	Low	Low	High	Unclear	High
Brief et al [24]	Unclear	Unclear	Low	High	Unclear	High
Calhoun et al [25]	Unclear	Low	Low	Unclear	Unclear	High
Clausen et al [34]	High	High	Low	High	Low	High
Cooper et al [20]	Unclear	Unclear	Low	High	Low	High
Cucciare et al [22]	Low	Unclear	Low	Low	Low	Low
Engel et al [36]	Low	Low	Low	Low	Unclear	Low
Hobfoll et al [26]	Low	Unclear	Low	High	Unclear	High
Kahn et al [27]	Low	Unclear	Low	Low	Low	Low
Krupnick et al [40]	Unclear	Unclear	Low	High	Unclear	High
Larsen et al [37]	Unclear	Unclear	Low	Unclear	Low	High
Litz et al [28]	Unclear	Unclear	Low	High	Unclear	High
McLean et al [32,42]	Unclear	Unclear	Low	High	Unclear	High
Pedersen et al [33]	Unclear	Unclear	Low	Low	Low	Moderate
Pemberton et al [29]	High	High	Low	High	Unclear	High
Pfeiffer et al [30]	Low	Unclear	Low	Unclear	Unclear	High
Possemato et al [38]	Unclear	Unclear	Low	High	Unclear	High
Salivar et al [41]	Unclear	Unclear	Low	Unclear	Unclear	High
Sayer et al [39]	Low	Unclear	Low	Low	Unclear	Moderate
Short et al [23]	Low	Unclear	Low	High	High	High
Taylor et al [31]	High	High	Low	Low	Low	High
Timmons et al [21]	Unclear	Unclear	Low	Unclear	Unclear	High

Sequence Generation

Of the 23 studies, 11 (48%) provided a detailed description of their process of sequence generation, 8 (25%) were rated as low risk for bias, and 3 (13%) were rated as high risk for bias. Studies that were at high risk for bias either discontinued randomization during the study because of technical difficulties [31] or an investigator-moving institution [34]. Pemberton et al [29] were unable to randomize as intended as internet speed limited the availability of interventions at certain study locations. Approximately 52% (12/23) of studies did not provide sufficient detail on their process of sequence generation (eg, specified that block randomization was used but failed to outline the process of selecting block size), and their risk of bias was rated as unclear.

Allocation Concealment

Of the 23 studies, 3 (13%) were rated as having a low risk for bias, 2 (9%) explicitly stated that study staff were blind to treatment allocation [19,25], and 1 (4%) used variable block size, blinding study staff to treatment allocation [36]. The 13% (3/23) of studies that were unable to randomize as intended (see *Sequence Generation* section; Taylor et al [29], Clausen et al [31], and Pemberton et al [34]) presumably had to unblind study staff to treatment allocation and were rated as having a high risk of bias. The remaining studies did not provide sufficient

XSL•FO
RenderX

details about allocation concealment and were rated as having an unclear risk of bias.

Blind Outcome Assessment

All but 17% (4/23) of studies used self-report or physiological measures exclusively to assess outcomes, and these studies were rated as low risk for bias. The studies that used clinician-guided or rated assessment measures [20,28,32,34] specified that raters were blind to the treatment arm and were thus rated as having a low risk of bias.

Incomplete Data

Of the 23 studies, 17 (74%) clearly defined treatment attrition and stated that they used intention-to-treat (ITT) analyses [19,20,22,24-31,35-40]; however, 2 (9%) studies did not include all randomized participants in ITT analyses [20,38] and 1 (4%) study failed to impute missing data [19]. In addition, 30% (7/23) of these studies [19,20,24,26,28,29,38] reported high attrition (>15% difference in missing data between treatment arms and/or >40% missing data overall [43]; as cited in Berge et al [44]), which introduces risk for bias regardless of the statistical methods used to attenuate this risk. The remaining studies either conducted completer analyses [21,23,34,42], had missingness that was not random [23,32,34], or failed to specify attrition rates by treatment arm [41].

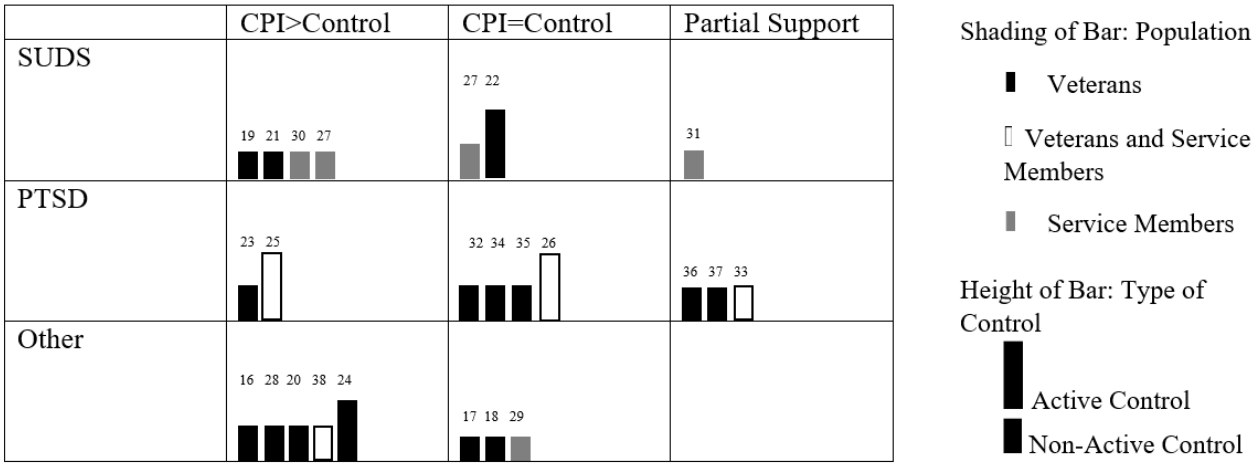
Selective Reporting

Approximately 35% (8/23) of trials preregistered their studies [20,22,27,31,34,35,37], and 30% (7/23) reported preregistered outcomes and were rated as having a low risk of bias. One of the studies [23] reported outcomes significantly different from those that were preregistered, and this study was rated as having a high risk for bias. One of the studies published a protocol paper outlining their methods and intended analyses and was rated as having a low risk for bias [33]. The remaining studies were not preregistered, and the risk for bias was rated as unclear.

Outcomes and Satisfaction

Figure 2 presents the outcome data visually in a harvest plot, and study effect sizes are included in Multimedia Appendix 2.

Figure 2. Harvest plot of study outcomes by intervention, population, and control group. PTSD: posttraumatic stress disorder; SUD: substance use disorder; TAU: treatment as usual.



Of the 23 studies, 13 (57%) reported positive results when comparing computerized psychological interventions to nonactive control groups or treatment as usual, 9 (39%) reported results that supported the specified hypotheses, and 4 (17%) reported partial support for the hypotheses. For example, Krupnick et al [40] reported that expressive writing (compared with treatment as usual) reduced PTSD hyperarousal symptoms; however, there was no significant movement in other PTSD domains. Sayer et al [39] found that expressive writing (compared with placebo writing) reduced physical complaints, anger, and general psychological distress but did not report a reduction in PTSD symptoms. Similarly, Acosta et al [35] and Engel et al [36] reported that gains in one domain (drinking and PTSD, respectively) failed to generalize to other outcomes.

Approximately 9% (2/23) of studies reported that the computerized psychological intervention under examination outperformed an active control treatment. Litz et al [28] found that self-management CBT (vs computerized supportive counseling) reduced daily measures of PTSD and depression, with 6-month follow-up reductions in depression, PTSD, and anxiety in the completer group. Kahn et al [27] found that their computerized psychological intervention led to improvements in various mental health–related outcomes when compared with residential treatment. Approximately 17% (4/23) of studies found that their computerized psychological intervention performed as well as the face-to-face treatment equivalent. Positive results found in computerized psychological interventions were comparable with those for in-person CBT for insomnia [31], in-person group anger inoculation training [21], therapist-led cognitive rehabilitation [20], and clinic-based smoking cessation care [25]. McLean et al [32] found that computerized prolonged exposure did not outperform non–trauma-focused face-to-face treatment when examining PTSD outcomes.

Approximately 17% (4/23) of studies failed to demonstrate significant treatment effects, with expressive writing [38], executive functioning training [34], and working memory training [37] for PTSD not outperforming placebo. Similarly, 50% (1/2) of the computerized psychological interventions for

XSL•FO
RenderX

alcohol use examined by Pemberton et al [29] did not outperform the waitlist control group.

Satisfaction data were provided in several studies. Overall, participants felt that they benefited from computerized psychological intervention engagement: with 82% [35], 93% [38], 96% [41], and 76.1% [39] of participants noting positive treatment effects. One of the studies found similar satisfaction rates between the computerized psychological intervention and the in-person treatment equivalent [21]. Although participants generally reported high treatment satisfaction and acceptability, there were reports of computerized psychological interventions being difficult to complete [34], difficult to understand [35], or impersonal and time consuming [40].

Discussion

Principal Results and Comparison With Previous Work

We aimed to systematically review the literature examining computerized psychological interventions in veteran and military populations. Across 23 studies, 24 interventions met the inclusion criteria and were reviewed. PTSD and SUDs were the most commonly targeted clinical difficulties, and other outcomes included anger, depression, insomnia, traumatic brain injury, relationship distress, suicidality, and readjustment difficulties. Interventions spanned a range of modalities and mostly focused on veterans, although a subset of studies recruited active military personnel. Approximately 8% (2/24) of interventions included romantic partners, whereas the other interventions followed an individual format. Most studies provided administrative support only, 13% (3/23) of studies provided no support, and 22% (5/23) of studies provided clinically meaningful support. Results were mostly positive; however, only 13% (3/23) of studies reporting positive results were rated as low risk for bias. Similarly, all studies that did not report significant treatment effects were rated as having a high risk of bias. Therefore, we caution against interpreting these results as unambiguous evidence of clinical effectiveness or ineffectiveness. Although it appears that computerized psychological interventions hold promise for the treatment of psychological difficulties in veterans and military service members, there is a need for more high-quality evidence to increase the confidence with which conclusions can be drawn.

Given the broad inclusion criteria that allowed great heterogeneity in intervention content and outcomes, a limited number of RCTs were identified. This is especially true in comparison with the number of RCTs that examine computerized psychological interventions in the general population. For example, Andrews et al [4] identified 53 RCTs that targeted depression and anxiety. Although the literature search returned a substantial number of pilot trials and process papers, limited RCTs were identified, indicating that the step from examining feasibility to establishing efficacy has not been decisively made. Furthermore, a high or unclear risk of bias across multiple features was common, with attrition bias being a concern for 48% (11/23) of the studies reviewed, greatly reducing the confidence with which inferences can be made. There are generally high rates of attrition in computerized psychological intervention trials [45], and veterans might be at a higher risk of attrition from treatment [46]. Although 74% (17/23) of the studies included in this review attempted to compensate for missing data by using ITT analyses, 30% (7/23) of studies had such a significant loss of data that high risk for bias was introduced regardless of the statistical methods used to attenuate this risk. Future research should attempt to reduce data loss by incorporating procedures associated with improved study retention. For example, clear study completion deadlines and prescheduled posttreatment assessments have been shown to reduce attrition [47], and adherence is improved when interventions are designed to include persuasive technology (ie, technology designed to include elements of social influence, such as praise, personalization, and social learning [48]). Therapeutic support (vs no support or administrative support) is also associated with improved retention in computerized psychological intervention trials [49], although it is not clear whether the benefits of including therapist contact outweigh the limitations placed on intervention scalability. Recently, evidence [50,51] has supported the efficacy of single-session web-based interventions, which can maximize recruitment while minimizing attrition. Of the studies reviewed here, Pedersen et al [33] reported low attrition rates and medium effect sizes for a single-session intervention administered in a help-seeking sample, demonstrating that these interventions can be successfully adapted for use in veterans.

Attrition is associated with insufficient statistical power, especially when the initial sample sizes are small, as was the case in many of the studies reviewed. Future research should anticipate high rates of attrition and set recruitment goals to ensure adequate statistical power for detecting treatment effects. In addition to dropout, low study use is another common concern in studies examining computerized psychological interventions. Notably, only 9% (2/23) of studies [30,37] defined study use (eg, minutes spent in the program or modules completed) as constituting adherence versus nonadherence. Similarly, only 30% (7/23) of studies described the relationship between use metrics and study outcomes. It is important to further consider these metrics in computerized psychological intervention research, as there is evidence that a dose–response relationship exists for computerized psychological interventions. For example, increases in modules completed [52] and more frequent use [53] are related to greater improvement at posttreatment. Examining study use and its relationship to outcome would also elucidate which intervention components are associated with change, allowing for the optimization of treatment.

Several studies have noted that sample characteristics limit the generalizability of the results. As is the case with much veteran and military research, samples tended to be mostly male. There is a need for computerized psychological interventions addressing the unique needs of female veterans, with recent evidence suggesting that these interventions are feasible, satisfactory, and potentially beneficial [54]. Male veterans who experienced military sexual trauma and transgender veterans are other subpopulations that might benefit from computerized psychological interventions because of high mental health disorder rates and numerous barriers to establishing care [55,56]. Further, some studies limited enrollment to post–9/11 war veterans. This cohort is younger and might be more computer

XSL•FO
RenderX

literate than the overall veteran population; however, given that a substantial proportion of veterans served in VA are ages >65 years [57], it is important to investigate the utility of computerized psychological interventions for this population. Although evidence suggests that computerized psychological interventions can provide additional care options for rural patients [58], there is a need to establish ways of extending coverage to veterans who live in rural areas without adequate internet infrastructure or transportation.

Several studies noted sample heterogeneity within treatment arms as a limitation. Although limiting sample variability can enhance the interpretability of findings, it also constrains generalizability. Alternatively, given a robust sample size, statistical methods can be leveraged to identify characteristics within heterogeneous samples associated with beneficial and/or adverse computerized psychological intervention treatment effects [59]. This would allow a stepped care approach, where minimally invasive and cost-effective treatments such as computerized psychological interventions are initially offered to those who are most likely to benefit, and resource-intensive face-to-face treatments are reserved for veterans or service members who require a higher level of care. Optimizing service delivery by identifying subsets of veterans or military personnel who are well-suited for computerized psychological intervention treatment is an important avenue of research, and initial explorations in this area are being reported [60].

VA is increasingly focusing on offering EBTs for mental health disorders. Although various interventions reviewed here included evidence-based practices (eg, exposure and motivational techniques), only 13% (3/23) of studies digitalized an EBT [30-32]. There is evidence that face-to-face EBTs used at VA can be successfully offered in a computerized format, as there have been positive trials of web-based exposure-based trauma treatment [61] and CBT for chronic pain [62] in general community samples. Similarly, there is an evidence base for web-based CBT [63] and acceptance and commitment therapy [64] for depression. Future research should continue to focus on digitalizing those treatments that have proven efficacy.

Limitations

This review has several limitations. First, owing to small sample sizes and high rates of potential bias, conclusions that computerized psychological interventions are potentially beneficial for veterans and service members are tentative. Lack of reporting was common across studies, resulting in many *unclear* bias ratings, which introduces uncertainty in our overall assessment of bias. Second, findings might not generalize to the broader veteran population, as many of the reviewed studies limited enrollment to post–9/11 veterans. Third, only English language articles were included, and all the studies reviewed were conducted in North America with US veterans or service members, and results cannot be extrapolated to other contexts. Fourth, only 1 author (RP) evaluated the results of the search; having an additional reviewer of these results would have strengthened the methodology. Finally, the methods of the studies included in this review are diverse, which in some instances complicates direct comparisons. The study heterogeneity also precluded conducting a meta-analysis, which would be the most robust way of assessing intervention efficacy. As the literature examining computerized psychological interventions in veterans and service members grows, efforts should be made to synthesize results by conducting a meta-analysis, which would provide evidence for intervention utility in this population.

Conclusions

Computerized psychological interventions are uniquely positioned to optimize treatment access and use for service members and veterans. These interventions could be integrated into a stepped care framework and reduce the burden on the health care system while increasing engagement with mental health services in this vulnerable population. Despite increased research interest, significant work remains in the development and evaluation of computerized psychological interventions targeted at veterans and service members. Although initial outcomes suggest that computerized psychological interventions are potentially beneficial for this population, much of the available research is at high risk for bias and fails to fully incorporate known evidence-based practices. There is an opportunity to design treatments that minimize threats to internal validity (eg, attrition and limited engagement) by including strategies to increase user motivation and by distilling treatments to include the most active intervention components. As veterans and service members report complex mental health challenges, as well as perceived and actual barriers in obtaining adequate care, there is an acute need to address the limitations of the existing literature.

Acknowledgments

This work was supported by the resources of the Central Texas Veterans Health Care System (Temple, Texas) and the Veterans Integrated Service Network 17 Center of Excellence for Research in Returning War Veterans (Waco, Texas). The contents of this manuscript do not represent the views of the US Department of Veterans Affairs or the United States Government.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms for review.
[DOCX File , 13 KB - jmir_v24i6e30065_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Characteristics of the studies included in the review.
[DOCX File , 18923 KB - jmir_v24i6e30065_app2.docx]

References

1. Kazdin AE. Addressing the treatment gap: a key challenge for extending evidence-based psychosocial interventions. Behav Res Ther 2017;88:7-18. [doi: 10.1016/j.brat.2016.06.004] [Medline: 28110678]
2. Webb CA, Rosso IM, Rauch SL. Internet-based cognitive-behavioral therapy for depression: current progress and future directions. Harv Rev Psychiatry 2017;25(3):114-122 [FREE Full text] [doi: 10.1097/HRP.0000000000000139] [Medline: 28475503]
3. Paganini S, Teigelkötter W, Buntrock C, Baumeister H. Economic evaluations of internet- and mobile-based interventions for the treatment and prevention of depression: a systematic review. J Affect Disord 2018;225:733-755. [doi: 10.1016/j.jad.2017.07.018] [Medline: 28922737]
4. 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;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
5. 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;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
6. Sztein DM, Koransky CE, Fegan L, Himelhoch S. Efficacy of cognitive behavioural therapy delivered over the Internet for depressive symptoms: a systematic review and meta-analysis. J Telemed Telecare 2018;24(8):527-539. [doi: 10.1177/1357633X17717402] [Medline: 28696153]
7. Cuijpers P, Marks I, van Straten A, Cavanagh K, Gega L, Andersson G. Computer-aided psychotherapy for anxiety disorders: a meta-analytic review. Cogn Behav Ther 2009;38(2):66-82. [doi: 10.1080/16506070802694776] [Medline: 20183688]
8. Lewis C, Roberts NP, Simon N, Bethell A, Bisson JI. Internet-delivered cognitive behavioural therapy for post-traumatic stress disorder: systematic review and meta-analysis. Acta Psychiatr Scand 2019;140(6):508-521. [doi: 10.1111/acps.13079] [Medline: 31359407]
9. Boumparis N, Karyotaki E, Schaub MP, Cuijpers P, Riper H. Internet interventions for adult illicit substance users: a meta-analysis. Addiction 2017;112(9):1521-1532 [FREE Full text] [doi: 10.1111/add.13819] [Medline: 28295758]
10. Riper H, Blankers M, Hadiwijaya H, Cunningham J, Clarke S, Wiers R, et al. Effectiveness of guided and unguided low-intensity internet interventions for adult alcohol misuse: a meta-analysis. PLoS One 2014;9(6):e99912 [FREE Full text] [doi: 10.1371/journal.pone.0099912] [Medline: 24937483]
11. Zachariae R, Lyby MS, Ritterband LM, O'Toole MS. Efficacy of internet-delivered cognitive-behavioral therapy for insomnia - a systematic review and meta-analysis of randomized controlled trials. Sleep Med Rev 2016;30:1-10. [doi: 10.1016/j.smrv.2015.10.004] [Medline: 26615572]
12. Brown M, Glendenning A, Hoon AE, John A. Effectiveness of web-delivered acceptance and commitment therapy in relation to mental health and well-being: a systematic review and meta-analysis. J Med Internet Res 2016;18(8):e221 [FREE Full text] [doi: 10.2196/jmir.6200] [Medline: 27558740]
13. Spijkerman MP, Pots WT, 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;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
14. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
15. Ruzek J, Hoffman J, Ciulla R, Prins A, Kuhn E, Gahm G. Bringing internet-based education and intervention into mental health practice: afterdeployment.org. Eur J Psychotraumatol 2011;2(1):7278 [FREE Full text] [doi: 10.3402/ejpt.v2i0.7278] [Medline: 22893824]
16. Moore BA, Fazzino T, Garnet B, Cutter CJ, Barry DT. Computer-based interventions for drug use disorders: a systematic review. J Subst Abuse Treat 2011;40(3):215-223 [FREE Full text] [doi: 10.1016/j.jsat.2010.11.002] [Medline: 21185683]
17. Rodriguez-Paras C, Tippey K, Brown E, Sasangohar F, Creech S, Kum HC, et al. Posttraumatic stress disorder and mobile health: app investigation and scoping literature review. JMIR Mhealth Uhealth 2017;5(10):e156 [FREE Full text] [doi: 10.2196/mhealth.7318] [Medline: 29074470]
18. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
19. Bedford LA, Dietch JR, Taylor DJ, Boals A, Zayfert C. Computer-guided problem-solving treatment for depression, PTSD, and insomnia symptoms in student veterans: a pilot randomized controlled trial. Behav Ther 2018;49(5):756-767. [doi: 10.1016/j.beth.2017.11.010] [Medline: 30146142]

XSL•FO
RenderX

20. Cooper DB, Bowles AO, Kennedy JE, Curtiss G, French LM, Tate DF, et al. Cognitive rehabilitation for military service members with mild traumatic brain injury: a randomized clinical trial. J Head Trauma Rehabil 2017;32(3):E1-15. [doi: 10.1097/HTR.0000000000000254] [Medline: 27603763]
21. Timmons PL, Oehlert ME, Sumerall SW, Timmons CW, Borgers SB. Stress inoculation training for maladaptive anger: comparison of group counseling versus computer guidance. Comput Hum Behav 1997;13(1):51-64. [doi: 10.1016/s0747-5632(96)00029-5]
22. Cucciare MA, Combs AS, Joshi G, Han X, Humphreys K. Computer-delivered brief alcohol intervention for patients with liver disease: a randomized controlled trial. Addiction 2021;116(5):1076-1087. [doi: 10.1111/add.15263] [Medline: 32924207]
23. Short NA, Fuller K, Norr AM, Schmidt NB. Acceptability of a brief computerized intervention targeting anxiety sensitivity. Cogn Behav Ther 2017;46(3):250-264. [doi: 10.1080/16506073.2016.1232748] [Medline: 27712458]
24. Brief DJ, Rubin A, Keane TM, Enggasser JL, Roy M, Helmuth E, et al. Web intervention for OEF/OIF veterans with problem drinking and PTSD symptoms: a randomized clinical trial. J Consult Clin Psychol 2013;81(5):890-900. [doi: 10.1037/a0033697] [Medline: 23875821]
25. Calhoun PS, Datta S, Olsen M, Smith VA, Moore SD, Hair LP, et al. Comparative effectiveness of an internet-based smoking cessation intervention versus clinic-based specialty care for veterans. J Subst Abuse Treat 2016;69:19-27 [FREE Full text] [doi: 10.1016/j.jsat.2016.06.004] [Medline: 27568506]
26. Hobfoll SE, Blais RK, Stevens NR, Walt L, Gengler R. Vets prevail online intervention reduces PTSD and depression in veterans with mild-to-moderate symptoms. J Consult Clin Psychol 2016;84(1):31-42. [doi: 10.1037/ccp0000041] [Medline: 26322788]
27. Kahn JR, Collinge W, Soltysik R. Post-9/11 veterans and their partners improve mental health outcomes with a self-directed mobile and web-based wellness training program: a randomized controlled trial. J Med Internet Res 2016;18(9):e255 [FREE Full text] [doi: 10.2196/jmir.5800] [Medline: 27678169]
28. Litz BT, Engel CC, Bryant RA, Papa A. A randomized, controlled proof-of-concept trial of an internet-based, therapist-assisted self-management treatment for posttraumatic stress disorder. Am J Psychiatry 2007;164(11):1676-1683. [doi: 10.1176/appi.ajp.2007.06122057] [Medline: 17974932]
29. Pemberton MR, Williams J, Herman-Stahl M, Calvin SL, Bradshaw MR, Bray RM, et al. Evaluation of two web-based alcohol interventions in the U.S. military. J Stud Alcohol Drugs 2011;72(3):480-489. [doi: 10.15288/jsad.2011.72.480] [Medline: 21513685]
30. Pfeiffer PN, Pope B, Houck M, Benn-Burton W, Zivin K, Ganoczy D, et al. Effectiveness of peer-supported computer-based CBT for depression among veterans in primary care. Psychiatr Serv 2020;71(3):256-262. [doi: 10.1176/appi.ps.201900283] [Medline: 31931686]
31. Taylor D, Peterson AL, Pruiksma KE, Young-McCaughan S, Nicholson K, Mintz J, STRONG STAR Consortium. Internet and in-person cognitive behavioral therapy for insomnia in military personnel: a randomized clinical trial. Sleep 2017;40(6):zsx075. [doi: 10.1093/sleep/zsx075] [Medline: 28472528]
32. McLean CP, Foa EB, Dondanville KA, Haddock CK, Miller ML, Rauch SA, et al. The effects of web-prolonged exposure among military personnel and veterans with posttraumatic stress disorder. Psychol Trauma 2021;13(6):621-631. [doi: 10.1037/tra0000978] [Medline: 33211517]
33. Pedersen ER, Parast L, Marshall GN, Schell TL, Neighbors C. A randomized controlled trial of a web-based, personalized normative feedback alcohol intervention for young-adult veterans. J Consult Clin Psychol 2017;85(5):459-470 [FREE Full text] [doi: 10.1037/ccp0000187] [Medline: 28287799]
34. Clausen AN, Thelen J, Francisco AJ, Bruce J, Martin L, McDowd J, et al. Computer-based executive function training for combat veterans with PTSD: a pilot clinical trial assessing feasibility and predictors of dropout. Front Psychiatry 2019;10:62 [FREE Full text] [doi: 10.3389/fpsyt.2019.00062] [Medline: 30881315]
35. Acosta MC, Possemato K, Maisto SA, Marsch LA, Barrie K, Lantinga L, et al. Web-delivered CBT reduces heavy drinking in OEF-OIF veterans in primary care with symptomatic substance use and PTSD. Behav Ther 2017;48(2):262-276 [FREE Full text] [doi: 10.1016/j.beth.2016.09.001] [Medline: 28270335]
36. Engel CC, Litz B, Magruder KM, Harper E, Gore K, Stein N, et al. Delivery of self training and education for stressful situations (DESTRESS-PC): a randomized trial of nurse assisted online self-management for PTSD in primary care. Gen Hosp Psychiatry 2015;37(4):323-328 [FREE Full text] [doi: 10.1016/j.genhosppsych.2015.04.007] [Medline: 25929985]
37. Larsen SE, Lotfi S, Bennett KP, Larson CL, Dean-Bernhoft C, Lee HJ. A pilot randomized trial of a dual n-back emotional working memory training program for veterans with elevated PTSD symptoms. Psychiatry Res 2019;275:261-268 [FREE Full text] [doi: 10.1016/j.psychres.2019.02.015] [Medline: 30939398]
38. Possemato K, Ouimette P, Knowlton P. A brief self-guided telehealth intervention for post-traumatic stress disorder in combat veterans: a pilot study. J Telemed Telecare 2011;17(5):245-250. [doi: 10.1258/jtt.2011.100909] [Medline: 21636687]
39. Sayer NA, Noorbaloochi S, Frazier PA, Pennebaker JW, Orazem RJ, Schnurr PP, et al. Randomized controlled trial of online expressive writing to address readjustment difficulties among U.S. Afghanistan and Iraq war veterans. J Trauma Stress 2015;28(5):381-390. [doi: 10.1002/jts.22047] [Medline: 26467326]

XSL•FO
RenderX

40. Krupnick JL, Green BL, Amdur R, Alaoui A, Belouali A, Roberge E, et al. An internet-based writing intervention for PTSD in veterans: a feasibility and pilot effectiveness trial. Psychol Trauma 2017;9(4):461-470. [doi: 10.1037/tra0000176] [Medline: 27607767]
41. Georgia Salivar E, Knopp K, Roddy MK, Morland LA, Doss BD. Effectiveness of online OurRelationship and ePREP programs for low-income military couples. J Consult Clin Psychol 2020;88(10):899-906. [doi: 10.1037/ccp0000606] [Medline: 33048570]
42. McLean CP, Miller ML, Gengler R, Henderson J, Sloan DM. The efficacy of written exposure therapy versus imaginal exposure delivered online for posttraumatic stress disorder: design of a randomized controlled trial in veterans. Contemp Clin Trials 2020;91:105990. [doi: 10.1016/j.cct.2020.105990] [Medline: 32184198]
43. Institute of Medicine. Treatment of PTSD: an assessment of the evidence. Washington, D.C: National Academies Press; 2008.
44. Berge EE, Hagen R, Øveraas Halvorsen J. PTSD relapse in veterans of Iraq and Afghanistan: a systematic review. Mil Psychol 2020;32(4):300-312. [doi: 10.1080/08995605.2020.1754123]
45. McDonald A, Eccles JA, Fallahkhair S, Critchley HD. Online psychotherapy: trailblazing digital healthcare. BJPsych Bull 2020;44(2):60-66 [FREE Full text] [doi: 10.1192/bjb.2019.66] [Medline: 31685068]
46. Goetter EM, Bui E, Ojserkis RA, Zakarian RJ, Brendel RW, Simon NM. A systematic review of dropout from psychotherapy for posttraumatic stress disorder among Iraq and Afghanistan combat veterans. J Trauma Stress 2015;28(5):401-409. [doi: 10.1002/jts.22038] [Medline: 26375387]
47. Andersson G. Internet-delivered psychological treatments. Annu Rev Clin Psychol 2016;12:157-179. [doi: 10.1146/annurev-clinpsy-021815-093006] [Medline: 26652054]
48. 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;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
49. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
50. Schleider JL, Weisz JR. Little treatments, promising effects? Meta-analysis of single-session interventions for youth psychiatric problems. J Am Acad Child Adolesc Psychiatry 2017;56(2):107-115. [doi: 10.1016/j.jaac.2016.11.007] [Medline: 28117056]
51. Tanner-Smith EE, Lipsey MW. Brief alcohol interventions for adolescents and young adults: a systematic review and meta-analysis. J Subst Abuse Treat 2015;51:1-18 [FREE Full text] [doi: 10.1016/j.jsat.2014.09.001] [Medline: 25300577]
52. 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;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
53. Manwaring JL, Bryson SW, Goldschmidt AB, Winzelberg AJ, Luce KH, Cunning D, et al. Do adherence variables predict outcome in an online program for the prevention of eating disorders? J Consult Clin Psychol 2008;76(2):341-346. [doi: 10.1037/0022-006X.76.2.341] [Medline: 18377129]
54. Creech SK, Pulverman CS, Shin ME, Roe KT, Tzilos Wernette G, Orchowski LM, et al. An open trial to test participant satisfaction with and feasibility of a computerized intervention for women veterans with sexual trauma histories seeking primary care treatment. Violence Against Women 2021;27(3-4):597-614. [doi: 10.1177/1077801219895102] [Medline: 31896315]
55. Schvey NA, Klein DA, Pearlman AT, Kraff RI, Riggs DS. Stigma, health, and psychosocial functioning among transgender active duty service members in the U.S. military. Stigma Health 2020;5(2):188-198 [FREE Full text] [doi: 10.1037/sah0000190]
56. Morris EE, Smith JC, Farooqui SY, Surís AM. Unseen battles: the recognition, assessment, and treatment issues of men with military sexual trauma (MST). Trauma Violence Abuse 2014;15(2):94-101. [doi: 10.1177/1524838013511540] [Medline: 24231941]
57. Amaral EF, Pollard MS, Mendelsohn J, Cefalu M. Current and future demographics of the veteran population, 2014–2024. Popul Rev 2018;57(1):28-60. [doi: 10.1353/prv.2018.0002]
58. Schueller SM, Hunter JF, Figueroa C, Aguilera A. Use of digital mental health for marginalized and underserved populations. Curr Treat Options Psych 2019;6(3):243-255. [doi: 10.1007/s40501-019-00181-z]
59. Pearson R, Pisner D, Meyer B, Shumake J, Beevers CG. A machine learning ensemble to predict treatment outcomes following an internet intervention for depression. Psychol Med 2019;49(14):2330-2341 [FREE Full text] [doi: 10.1017/S003329171800315X] [Medline: 30392475]
60. Frankfurt S, Frazier P, Litz BT, Schnurr PP, Orazem RJ, Gravely A, et al. Online expressive writing intervention for reintegration difficulties among veterans: who is most likely to benefit? Psychol Trauma 2019;11(8):861-868. [doi: 10.1037/tra0000462] [Medline: 30998059]
61. Spence J, Titov N, Johnston L, Jones MP, Dear BF, Solley K. Internet-based trauma-focused cognitive behavioural therapy for PTSD with and without exposure components: a randomised controlled trial. J Affect Disord 2014;162:73-80. [doi: 10.1016/j.jad.2014.03.009] [Medline: 24767009]

XSL•FO
RenderX

62. Buhrman M, Fredriksson A, Edström G, Shafiei D, Tärnqvist C, Ljótsson B, et al. Guided internet-delivered cognitive behavioural therapy for chronic pain patients who have residual symptoms after rehabilitation treatment: randomized controlled trial. Eur J Pain 2013;17(5):753-765. [doi: 10.1002/j.1532-2149.2012.00244.x] [Medline: 23139021]
63. Foroushani PS, Schneider J, Assareh N. Meta-review of the effectiveness of computerised CBT in treating depression. BMC Psychiatry 2011;11:131 [FREE Full text] [doi: 10.1186/1471-244X-11-131] [Medline: 21838902]
64. Sierra MA, Ruiz FJ, Flórez CL. A systematic review and meta-analysis of third-wave online interventions for depression. Rev Latinoam Psicol 2018;50(2):126-135. [doi: 10.14349/rlp.2018.v50.n2.6]

Abbreviations

CBT: cognitive behavioral therapy
EBT: evidence-based treatment
ITT: intention to treat
PTSD: posttraumatic stress disorder
RCT: randomized controlled trial
SUD: substance use disorder
VA: Veterans Affairs

Edited by A Mavragani; submitted 03.05.21; peer-reviewed by A Raines, F Beyer, MDG Pimentel; comments to author 30.07.21; revised version received 07.10.21; accepted 20.12.21; published 03.06.22.

Please cite as:
Pearson R, Carl E, Creech SK
Computerized Psychological Interventions in Veterans and Service Members: Systematic Review of Randomized Controlled Trials
J Med Internet Res 2022;24(6):e30065
URL: https://www.jmir.org/2022/6/e30065
doi:10.2196/30065
PMID:35657663

©Rahel Pearson, Emily Carl, Suzannah K Creech. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.06.2022. 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

Remote Delivery of Yoga Interventions Through Technology: Scoping Review

Aurora James-Palmer[1], PT, DPT; Ellen Zambo Anderson[1], PT, DPT; Jean-Francois Daneault[1], PhD

Department of Rehabilitation and Movement Sciences, Rutgers University, Newark, NJ, United States

Corresponding Author:
Jean-Francois Daneault, PhD
Department of Rehabilitation and Movement Sciences
Rutgers University
65 Bergen Street
Newark, NJ, 07107
United States
Phone: 1 973 972 8482
Email: jf.daneault@rutgers.edu

Abstract

Background: The popularity of yoga and the understanding of its potential health benefits have recently increased. Unfortunately, not everyone can easily engage in in-person yoga classes. Over the past decade, the use of remotely delivered yoga has increased in real-world applications. However, the state of the related scientific literature is unclear.

Objective: This scoping review aimed to identify gaps in the literature related to the remote delivery of yoga interventions, including gaps related to the populations studied, the yoga intervention characteristics (delivery methods and intervention components implemented), the safety and feasibility of the interventions, and the preliminary efficacy of the interventions.

Methods: This scoping review was conducted in accordance with the PRISMA-ScR (Preferred Reporting Item for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines. Scientific databases were searched throughout April 2021 for experimental studies involving yoga delivered through technology. Eligibility was assessed through abstract and title screening and a subsequent full-article review. The included articles were appraised for quality, and data were extracted from each article.

Results: A total of 12 studies of weak to moderate quality were included. Populations varied in physical and mental health status. Of the 12 studies, 10 (83%) implemented asynchronous delivery methods (via prerecorded material), 1 (8%) implemented synchronous delivery methods (through videoconferencing), and 1 (8%) did not clearly describe the delivery method. Yoga interventions were heterogeneous in style and prescribed dose but primarily included yoga intervention components of postures, breathing, and relaxation and meditation. Owing to the heterogeneous nature of the included studies, conclusive findings regarding the preliminary efficacy of the interventions could not be ascertained.

Conclusions: Several gaps in the literature were identified. Overall, this review showed that more attention needs to be paid to yoga intervention delivery methods while designing studies and developing interventions. Decisions regarding delivery methods should be justified and not made arbitrarily. Studies of high methodological rigor and robust reporting are needed.

(J Med Internet Res 2022;24(6):e29092) doi:10.2196/29092

KEYWORDS

complementary therapies; mind-body; remote delivery; telerehabilitation; eHealth; yoga; technology; mind-body

Introduction

Background

As of 2016, a total of 36 million Americans had engaged in some form of yoga practice [1], with other countries demonstrating similar patterns of yoga practice [2-4]. Although healthy individuals incorporate yoga into their fitness routines to improve physical and mental health [1], yoga has also been used to manage symptoms of disease [2-5]. Even before the COVID-19 pandemic in 2020, individuals sought access to yoga in their home environments. In fact, results from a 2016 survey revealed that yoga was most commonly practiced at home [1]. During the pandemic, public health policies and individuals' personal preference for social distancing and staying at home stopped or limited in-person practice. Under these

XSL•FO
RenderX

circumstances, the availability of resources to practice yoga remotely increased, which could be beneficial for individuals with functional limitations.

However, when examining the available scientific literature, yoga interventions are most often delivered in person [6]. Although this may be feasible for constrained research projects, it greatly limits real-world applicability. First, access to qualified yoga instructors inevitably varies based on geographical location [7]. Second, individuals' socioeconomic status and access to transportation limit their access [8]. Third, inherent physical limitations may hinder the patients' ability to attend in-person yoga classes. Therefore, in-person yoga instruction presents several barriers regardless of the public health climate, such that alternative delivery methods (eg, videos, videoconferencing, and mobile apps) have been developed and are being used extensively. However, there seems to be a gap in the literature regarding interventions conducted in research studies and real-world practices.

Little is known about the evidence regarding the practice and outcomes of yoga using remote delivery methods. A previous scoping review of the yoga literature [6] included studies in which yoga was conducted at sites other than yoga studios, such as *at home*, where yoga instructors may not be present. However, details regarding intervention delivery were not provided, although there are multiple ways of delivering yoga remotely [6]. For example, information about yoga can be provided through hard copy resources, such as pamphlets and card decks, or through more engaging technological means. Specifically, technologies such as videoconferencing, DVDs, websites, and mobile apps can be used to provide yoga interventions. Furthermore, content can be delivered either (1) synchronously (an instructor interacts with participants in real time, eg, through videoconference) or (2) asynchronously (material is prerecorded for participants to use without real-time interaction). Despite these options and their potential impact, few studies have explored their use. Thus, little is known about how yoga is delivered remotely, whether remote delivery is used for some specific populations more than others, and what types of yoga are being delivered remotely.

With this, it is important to know that yoga is a multilayered ancient philosophy and practice intended to facilitate well-being through the cultivation of awareness by integrating mind and body, with an emphasis on self-realization [9]. There are various forms, often referred to as branches of yoga, which facilitate one's ability to reach a greater state of being. For instance, *Karma yoga* is the branch that focuses on devotion to service, *Jnana yoga* focuses on the development of knowledge, and *Hatha yoga* involves the practice of physical postures. *Hatha yoga* is typically what comes to mind when one thinks of *yoga* in the Western world. The term *Hatha yoga* refers to the branch of the physical practice of yoga and is also usually used to refer to a broad style of yoga that incorporates postures and breathing. There are other styles of yoga that fit under the umbrella of the physical practice of yoga—*Hatha yoga*—but are often specified further. Examples of these styles include *Iyengar yoga* and *Vinyasa yoga*. For example, *Iyengar yoga* focuses on body alignment, sequencing, and timing. By contrast, *Vinyasa yoga* is generally energetic and involves flowing through sequences of postures with breath integration. In addition to the branches of yoga and specific styles of the physical practice of yoga, there are 8 limbs of yoga. These 8 limbs are described in the *Yoga Sutras of Patanjali*, a foundational text, and include (1) *yama* (abstinences), (2) *niyama* (observances), (3) *asanas* (postures), (4) *pranayama* (breath control), (5) *pratyahara* (withdrawal of the senses), (6) *dharana* (concentration), (7) *dhyana* (meditation), and (8) *samadhi* (bliss) [9]. These limbs help to inform the practice of yoga in general, and some of them are incorporated in *Hatha yoga*. Most commonly, *asana*, *pranayama*, and *dhyana* are incorporated into *Hatha yoga*; however, in some cases, different combinations of the limbs of yoga are used. A previous scoping review of the yoga literature showed that most yoga intervention studies did not specify the style of yoga used however, of those that did, the most commonly reported styles were *Hatha yoga* (129/456, 28.3% of studies) and *Iyengar yoga* (41/456, 8.9% of studies) [6]. It is unknown whether similar patterns occur when yoga is delivered remotely.

Objective

Yoga can be delivered both in person and remotely, and there has been an increase in the availability and use of remotely delivered interventions in real-world applications. However, it is not known whether the scientific literature reflects this or reflects what types of populations have been enrolled in studies that have investigated the remote delivery of yoga, what the characteristics of these interventions are (including how remote delivery occurs), and what components of yoga are incorporated. Furthermore, the general feasibility and safety of these interventions or their preliminary efficacy are yet to be clearly defined. Therefore, the purpose of this scoping review was to examine the existing literature regarding the practice of yoga through remote delivery methods and identify current gaps related to (1) the populations studied, (2) the intervention characteristics (delivery methods and intervention components implemented), (3) the safety and feasibility of the interventions, and (4) the preliminary efficacy of the interventions.

Methods

This review was conducted in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines [10].

Search Strategy

Scientific databases, including PubMed, Scopus, Web of Science, PEDro, CINAHL, PsycINFO, IEEE Xplore Digital Library, and Cochrane Library, were searched during April 2021. The search strategies were modified for each database, with the Medical Subject Headings terms used as applicable. The following terms were used in various combinations: *yoga*, *Iyengar*, *ashtanga*, *hatha*, *asana*, *telerehabilitation*, *tele-rehabilitation*, *telemedicine*, *videoconferencing*, *video*, *telenursing*, *DVD*, *remote delivery*, *eHealth*, *video games*, *online*, and *virtual reality* (see Multimedia Appendix 1 for the specific search terms and search methods used for each database).

XSL•FO
RenderX

Inclusion and Exclusion Criteria

Articles were included if they investigated (1) adults aged ≥18 years; (2) an experimental intervention study with pre- and posttesting for at least one group; (3) a yoga intervention using physical yoga postures (*asanas)* and at least one other of the [9] 8 limbs of yoga, as described in the *Yoga Sutras of Patanjali*, including *yama* (abstinences), *niyama* (observances), *pranayama* (breath control), *pratyahara* (withdrawal of the senses), *dharana* (concentration), *dhyana* (meditation), and *samadhi* (bliss); and (4) a method of remote yoga delivery through technology for at least one group (eg, video, mobile app, and videoconferencing platform) as the primary intervention.

Yoga interventions can be heterogeneous and are often poorly described in the literature [6]. In some cases, they do not reflect the integration of the mind and body as intended in traditional practice [9]. To address this, we stipulated that selected articles should explicitly include physical postures (body) and another of the [9] 8 limbs of yoga (mind) using either English or Sanskrit words. This was intended to limit poorly described interventions or interventions in which yoga was not the primary intervention. For example, although mindfulness-based stress reduction includes components of yoga, there are other significant intervention components that make these interventions different from typical yoga interventions, which, if included, introduce even more heterogeneity across studies. Furthermore, it also allowed us to exclude studies that used only physical postures that could be more akin to exercise interventions.

Article Screening and Data Extraction

All the data management processes discussed in this section were completed by the 3 authors (AJP, EZA, and JFD) in the following manner. The articles were divided into thirds. Then, 2 authors, screened, reviewed, and appraised, (two-thirds or 66.7% of the articles) with the third author available to resolve conflicts, such that each author (AJP, EZA, and JFD) was able to perform each task. The abovementioned criteria guided eligibility screening (completed via Rayyan QCRI [11]) and full-text article assessment. For articles that were included based on the full-text assessment, data were extracted simultaneously with the full-article review, and then a quality assessment was completed. The Quality Assessment Tool for Quantitative Studies Effective Public Health Practice Project was used to assess the level of bias and methodological quality of each study [12]. Data were extracted into a Microsoft Excel sheet designed by the scoping review authors (initial design by AJP, edited and approved by EZA and JFD) before the article review. The data extraction categories were determined by the authors (initial selection by AJP, edited and approved by EZA and JFD) based on the objective of the review (see Table 1 for descriptions and examples of data extracted from each study).

XSL•FO
RenderX

Table 1. Data extracted from selected articles (scoping review; information extracted from each selected article and examples).

Content area and extracted information	Examples
General study information	
Study type as defined by the study authors in the introduction or methods sections	Randomized controlled trial; single-group study
Country in which the study was conducted	United States
Comparison group used as described by the study authors (if applicable)	Regular activity control group; active control group such as a strengthening program
Study populations	
Number of participants (total [N] and per group [n])	N=50
Description of population, including defining characteristics such as the health condition, as described by the study authors	Women with depression
Mean age of the participants and SD (if provided) for the total sample and each group	Mean age of the total sample was 55.07 (SD 9.69) years
Sex distribution of participants in the total sample and in each group	Number of women in the study out of the total number of participants
Justification of delivery method in relation to the study population, as described by the study authors (ie, did the study authors describe why they delivered the intervention remotely, and if they did, what was the reason)	Yes—the study authors reported that individuals with cancer often have transportation and scheduling challenges that make it difficult to attend in-person appointments; no—the study authors did not describe why they chose remote delivery
Intervention characteristics: delivery methods	
Intervention setting	Home
Whether delivery was synchronous or asynchronous; delivery was considered synchronous if interventions were delivered in real time such that the instructor could interact with the participant or participants; delivery was considered asynchronous if intervention materials were prerecorded and could be accessed at any time	Synchronous (videoconferencing) and asynchronous (prerecorded video)
The technology used to deliver the intervention	Name of a specific videoconferencing platform; type of prerecorded video (ie, DVD)
Whether delivery was group or individual; it was considered group delivery if multiple people participated in the yoga intervention together at one time; it was considered individual delivery if a participant engaged in the intervention alone	If each participant received access to a prerecorded video and watched the video on their own (individual delivery)
Whether participants had additional interactions with the study team outside of assessment sessions and prescribed intervention sessions	Participants received an in-person introduction yoga class before starting the intervention period
Whether participants received supplementary materials	Participants received written instructions providing additional information on how to practice yoga
Intervention characteristics: yoga intervention components and other details	
Style of yoga implemented	Hatha yoga or Iyengar yoga
Specific limbs of yoga implemented	Breathing; postures; meditation; relaxation
Yoga instructor credentials, as reported by the study authors, including instructor training (ie, are they a yoga instructor, yoga therapist, or other health care professional) and their certification training hours	Yoga instructor (200 hours)
Yoga dose: frequency and duration reported in minutes per session, sessions per week, and total number of weeks	30 minutes per session with 2 sessions per week for 6 weeks
Whether the yoga sequences were designed, adapted, or selected for the specific study population, as described by the study authors, or whether this was not reported	Yes—the study authors reported that they designed the prerecorded videos specifically for the population enrolled in the study; no—the study authors did not report whether the yoga intervention was designed for the study population
Information about additional home practice (ie, did the study authors describe whether participants were encouraged to engage in additional practice outside of the prescribed intervention and how this was kept track of)	Yes—although the study authors required participants to watch the yoga video 1 time per week, the study authors encouraged participants to view the yoga video an additional 2 to 3 times per week if possible and asked them to log how often they did this
Intervention feasibility and safety	

XSL•FO
RenderX

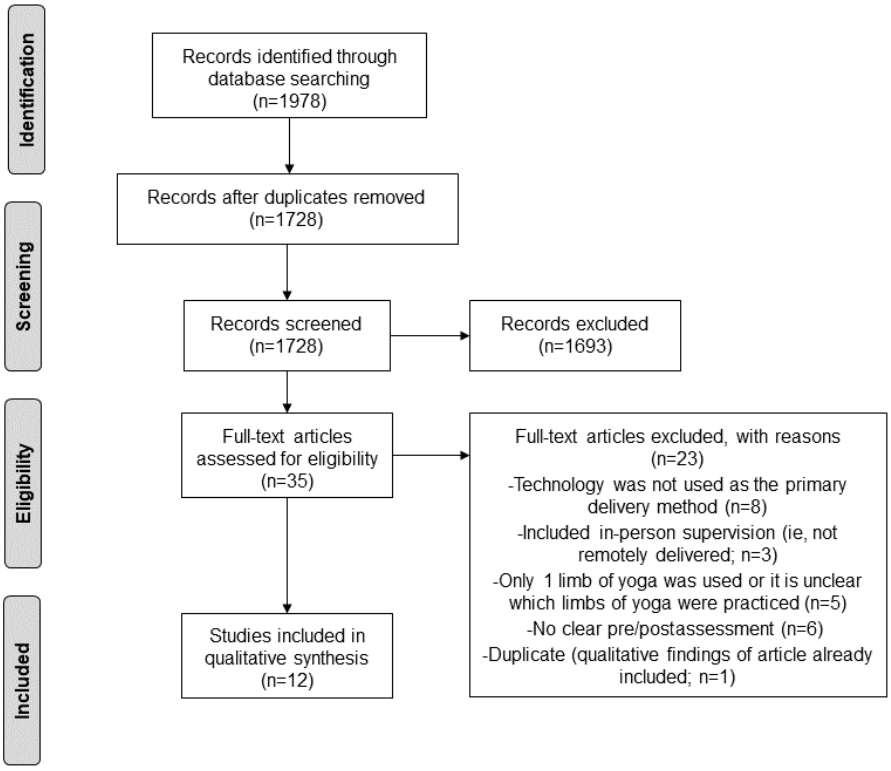
Content area and extracted information	Examples
Study adherence (ie, did participants complete the study overall, including the intervention period and assessment sessions) reported as how many people in each group completed the study	66% (44/67) of the yoga group completed the study
Intervention adherence (ie, did participants complete the intended yoga intervention dose); intervention adherence was reported as it was reported in each study; some studies reported it as the mean yoga practice, whereas others set a threshold or benchmark and reported intervention adherence as it related to the benchmark	Mean yoga practice was 44 min/week and the prescribed dose was 60 min/week; the benchmark for "good adherence" was participants who practiced yoga ≥6 times over 2 weeks, and 55% (37/67) of the yoga group met this benchmark
The presence or absence of technological challenges, as described by the study authors	Participants experienced technological challenges in 77% (24/31) of the sessions
The presence or absence of adverse events, as reported by the study authors for each study group, or whether the study authors did not report any information about adverse events	No adverse events occurred, the study did not report information about adverse events; 9 mild adverse events occurred in the yoga group and 4 mild adverse events occurred in the comparison group
Preliminary efficacy	
The outcome measures were categorized into patient-reported outcome measures, physical performance and function outcome measures, and physiological outcome measures based on what the measures assessed; subsequently, a summary of results for these outcomes was extracted (eg, were there significant improvements between groups and significant improvements within groups)	The patient-reported outcome measure—the Beck Depression Inventory—showed significant within-group improvements

Results

Overview

The search resulted in 1978 articles, with 1728 (87.36%) remaining after duplicates were removed. The title and abstract screening resulted in the inclusion of 2.03% (35/1728) of articles. Following a full-article review, of the 35 articles, 12 (34%) articles [13-24] published between 2003 and 2021 were included in the final qualitative analysis (see Figure 1 for the PRISMA [Preferred Reporting Items for Systematic Reviews and Meta-Analyses] flowchart).

Figure 1. Illustration showing the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the search, screening, and full-article review results.



XSL•FO
RenderX

General Study Information and Study Quality

Of the 12 studies, 8 (67%) were randomized controlled trials (RCTs) [13,14,18-23], 2 (17%) were quasi-experimental nonrandomized studies [16,17], 1 (8%) was an open-label single-group study [24], and 1 (8%) was a case series [15]. The comparison groups in the RCTs included regular activities [14,21], wait-list control [18], DVD program for strengthening [13], DVD program for walking [22], web-based stretching or toning program [23], and informational handouts about yoga [20]. One of the RCTs compared an in-person yoga intervention with a yoga DVD [19]. All active comparison groups were comparable in dose with the yoga interventions [13,19,22,23]. One of the quasi-experimental studies used health education [16], whereas the other [17] did not describe the comparison group. The methodological quality of the studies ranged from weak (7/12, 58% of studies [14-17,20,21,24]) to moderate (5/12, 42% of studies [13,18,19,22,23]; Table 2).

Table 2. General study information, including the first author, type, country, study groups, and results of the quality appraisal assessment.[a]

Author, study type, and country	Study groups	Selection bias	Study design	Confounders	Blinding	Data collection methods	Withdrawals or dropouts	Global rating[b]
Armstrong et al [14], RCT,[c] United States	Yoga video vs regular activity	Weak	Weak	Weak	Weak	Weak	Weak	Weak
Awdish et al [15], case series, United States	Yoga video; no comparison	Weak	Weak	Weak	Weak	Strong	Weak	Weak
Donesky et al [16], nonrandomized quasi-experimental, United States and United Kingdom	Yoga via videoconferencing vs health education phone call	Weak	Moderate	Weak	Weak	Strong	Moderate	Weak
Gunda et al [17], nonrandomized quasi-experimental, United States	Yoga DVD; control not clearly described	Moderate	Moderate	Strong	Weak	Weak	Moderate	Weak
Huberty et al [18], RCT, United States	Web-based yoga videos vs wait-list control	Moderate	Strong	Strong	Weak	Strong	Strong	Moderate
Huberty et al [23], RCT, United States	Web-based yoga videos (2 doses) vs stretch and tone control	Moderate	Strong	Moderate	Moderate	Strong	Weak	Moderate
Jasti et al [24], single group, India	Tele-yoga module	Moderate	Weak	Weak	Weak	Strong	Weak	Weak
Kyeongra et al [19], RCT, United States	Yoga DVD vs in-person yoga	Strong	Moderate	Strong	Weak	Strong	Moderate	Moderate
Mullur et al [20], RCT, United States	Yoga DVD vs handouts about yoga	Moderate	Weak	Strong	Weak	Moderate	Moderate	Weak
Sakuma et al [21], RCT, Japan	Yoga DVD vs regular activities	Moderate	Strong	Strong	Weak	Strong	Weak	Weak
Schuver et al [22], RCT, United States	Yoga DVD vs DVD on walking	Moderate	Strong	Strong	Moderate	Strong	Moderate	Moderate
Stan et al [13], RCT, United States	Yoga DVD vs DVD on strengthening	Moderate	Strong	Moderate	Weak	Strong	Moderate	Moderate

[a]The quality appraisal assessment was completed using the Quality Assessment Tool for Quantitative Studies with six domains contributing to the score: (1) selection bias, (2) study design, (3) confounders, (4) blinding, (5) data collection methods, and (6) withdrawals and dropouts.

[b]Global ratings were determined as follows: no weak ratings=strong, one weak rating=moderate, and ≥2 weak ratings=weak.

[c]RCT: randomized controlled trial.

Study Populations

The studies included individuals with physical and mental health conditions, as well as generally healthy adults. For example, 17% (2/12) of studies involved individuals with cancer, such as early-stage breast cancer [13] and myeloproliferative neoplasms (blood cancers) [18]. Approximately 25% (3/12) of studies involved individuals with cardiopulmonary conditions, including pulmonary hypertension [15], neurocardiogenic syncope [17], and a combination of chronic obstructive pulmonary disease (COPD) and heart failure (HF) [16]. Other studies involved sedentary adults who were overweight [19], veterans with diabetes [20], women who had experienced stillbirths [23], and women with depression [22]. Approximately 25% (3/12) of studies involved adults without any specified health conditions, including *adult women* [14], female childcare workers [21], and the general public in India during the COVID-19 pandemic [24]. The mean age of the participants ranged from 21 to 71 years. Across all studies, most participants were female, and 50% (6/12) of studies included only female participants [13-15,21-23]. For more details related to the study populations, refer to Table 3.

XSL•FO
RenderX

Table 3. Characteristics of the participants included in the reviewed studies.

Study	Population	Number of Participants			Age (years), mean (SD)			Sex, n (%)					
		Total	Yoga	Control	Total	Yoga	Control	Total		Yoga		Control	
								Female	Male	Female	Male	Female	Male
Armstrong et al [14]	Adult women	30	15	15	55.07 (9.69)	54 (10)	55 (9)	30 (100)	0 (0)	15 (100)	0 (0)	15 (100)	0 (0)
Awdish et al [15]	Women with pulmonary hypertension and additional chronic health conditions	3	3	N/A[a]	48, 32, and 24	48, 32, and 24	N/A	3 (100)	0 (0)	3 (100)	0 (0)	N/A	N/A
Donesky et al [16]	Adults with chronic obstructive pulmonary disease and heart failure	15	7	8	71 (8.5)	73 (14.3)	70.5 (2.7)	10 (66)	5 (33)	4 (57)	3 (43)	6 (75)	2 (25)
Gunda et al [17]	Adults with neurocardiogenic syncope	44	21	23	21 (3)	21 (3)	22 (3)	41 (93)	3 (7)	20 (95)	1 (5)	21 (91)	2 (9)
Huberty et al [18]	Adults with myeloproliferative neoplasm	48	27	21	56.9 (10.3)	58.3 (9.3)	55.0 (11.4)	45 (94)	38 (6)	25 (93)	2 (7)	20 (95)	1 (5)
Huberty et al [23]	Women who have experienced stillbirth	90	30 (YLD[b]) and 30 (YMD[c])	30	NS[d]	NS	NS	90 (100)	0 (0)	30 (100; YLD) and 30 (100; YMD)	0 (0)	30 (100)	0 (0)
Jasti et al [24]	General adult public	95	95	N/A	40.39 (13.33)	40.39 (13.33)	N/A	69 (73)	26 (27)	69 (73)	26 (27)	N/A	N/A
Kyeongra et al [19]	Sedentary adults who are overweight	14	7	7	58.6 (5.4)	58.7 (4.1)	58.4 (6.8)	12 (86)	2 (14)	6 (86)	1 (14)	6 (86)	1 (14)
Mullur et al [20]	Veterans with CKD[e] and diabetes	10	5	5	64.4 (NS)	60 (10.34)	68.8 (5.97)	1 (10)	9 (90)	1 (20)	4 (80)	0 (0)	5 (100)
Sakuma et al [21]	Female childcare workers	98	67	31	33.6 (NS)	32.6 (11.5)	35.8 (13.0)	98 (100)	0 (0)	67 (100)	0 (0)	31 (100)	0 (0)
Schuver et al [22]	Women with a history or diagnosis of depression	40	20	20	42.68 (4.95)	45.55 (12.30)	39.8 (11.23)	40 (100)	0 (0)	20 (100)	0 (0)	20 (100)	0 (0)
Stan et al [13]	Women with early-stage breast cancer and cancer-related fatigue	34	18	16	62.1 (8.1)	61.4 (7.0)	63.0 (9.3)	34 (100)	0 (0)	18 (100)	0 (0)	16 (100)	0 (0)

[a]N/A: not applicable.

[b]YLD: yoga low dose (60 min/week).

[c]YMD: yoga moderate dose (150 min/week).

[d]NS: not specified.

[e]CKD: chronic kidney disease.

In addition to extracting information about the enrolled populations, justification of the intervention delivery method based on the enrolled population was extracted when available. This was done to identify whether the study authors chose to implement remote delivery to address population-specific needs. Approximately 50% (6/12) of the studies reported some type of justification for the remote delivery method and related it to population-specific needs. Interestingly, one of these studies was designed during the COVID-19 pandemic and was conducted to provide easily accessible stress management strategies to the general public during social isolation [24]. Approximately 17% (2/12) of the studies [13,16], one for

XSL•FO
RenderX

individuals with COPD and HF [16] and the other for women with breast cancer [13], justified their choice of implementing a remote intervention by noting that transportation-related barriers impeded individuals' abilities to attend in-person interventions. Another study involving veterans with diabetes supported the choice of implementing a remote intervention by noting that physical impairments make it challenging for individuals to access and engage in physical activity [20]. Another study with older adults [14] reported low adherence to in-person exercise interventions in this population. Finally, a study with childcare workers [21] justified remote delivery using a yoga DVD as it was a convenient low-cost option. In contrast, 50% (6/12) of studies [15,17-19,22,23] did not provide any justification for the implementation of remote delivery.

Yoga Intervention Characteristics

Delivery Methods

As per the inclusion criteria, all reviewed studies [13-24] involved remote delivery of yoga. There were no inclusion criteria that stipulated the intervention setting; however, all the included studies stated that the interventions occurred in the participants' homes. Approximately 83% (10/12) of studies implemented asynchronous delivery through prerecorded yoga videos using DVDs [13,15,17,19-22], a mobile app [15], a website [18,23], and a *video* (type not specified) [14]. Among the studies that delivered yoga via DVDs, one of the studies compared in-person yoga with yoga delivered via DVD [19]. One of the studies implemented synchronous delivery [16] via the videoconferencing platform DocBox (MicroDesign) to provide real-time video interactions between the instructor and the participants. The study team installed the technology in the participants' homes, and a DocBox technician provided technological support for the entire study duration. One of the studies referred to using a *tele-yoga* module and mentioned *a minimum of supervised* sessions, implying that there were supervised and unsupervised sessions; however, further details of the delivery method were not clearly described [24].

By nature, 83% (10/12) of studies implementing asynchronous delivery were delivered individually [13-15,17-23]. For the study implementing synchronous delivery, the instructor saw the entire group to provide feedback; however, participants only saw the instructor and needed to be unmuted to ask questions [16]. Some studies that implemented asynchronous delivery included interactions with the study team at periodic intervals through phone calls [13,22], in-person visits [14], and introductory classes (one [14,20] or two [19]). Some studies that implemented asynchronous delivery used strategies to optimize safety, including written instructions [13,14,19,20], ability-based movement sequences [15], intensity ramp-ups across videos [18], and preparatory instructions within each video [23]. The study in which the intervention was delivered synchronously included safety measures in which a nurse observed and assessed participants before and after each tele-yoga class [16].

Approximately 50% (6/12) of studies implementing asynchronous delivery used only 1 yoga video or sequence for the study's duration [13,14,19-22]. In 25% (3/12) of studies, participants used a variety of videos [17,18,23]. One of these studies provided beginner and intermediate ratings to help participants choose videos [18], whereas the other provided 12 videos to be watched in a specific order to optimize safety [17]. In the case series, it was clear that each participant watched different videos; however, it was unclear whether a given participant watched the same yoga video or sequence for the study duration [15]. For the study comparing an in-person yoga intervention with a DVD, it was unclear whether the in-person sessions differed from week to week and whether the in-person and DVD groups performed the same sessions [19].

Yoga Intervention Components and Other Details

Approximately 50% (6/12) of studies did not specify the style of yoga used [13,14,17,20,21,24], whereas 50% (6/12) used specific styles, including gentle *Hatha* yoga [22,23], *Vinyasa* yoga [19], *Iyengar* yoga [16], a combination of *Hatha* yoga and *Iyengar* [15], and a combination of *Hatha* yoga and *Vinyasa* [18]. Approximately 50% (6/12) of studies [15,16,20,22-24] indicated that the yoga interventions were designed or adapted for their specific populations. One of these studies also provided additional preselected videos available on the web for participants to choose from after completing the videos designed for the intervention [23]. Approximately 17% (2/12) of studies [13,19] indicated that the yoga videos were selected with the population in mind. Approximately 17% (2/12) of studies [14,17] did not mention whether the programs were chosen for their populations. One of the studies used a yoga program that had previously shown benefits in people with low back pain; however, further rationale for its selection was not provided [21]. One of the studies selected specific yoga videos posted on the web but also created 6 videos specifically for the study [23].

Approximately 42% (5/12) of studies [14,17,20-22] did not specify the credentials of yoga instructors. One of the studies [13] specified that a certified yoga therapist provided the intervention, whereas another [16] specified that a certified yoga instructor and physical therapy assistant provided the intervention. The remaining studies specified that certified yoga instructors (500-hour certifications [18] or >200-hour certifications [19,23]) provided the interventions.

Approximately 33% (4/12) of studies specified using breathing exercises and physical postures [19-21,24]. One of the studies specified using postures and meditation [23]. Approximately 58% (7/12) of studies mentioned using breathing exercises, postures, and meditation and relaxation [13-16,18,22]. Of these 7 studies, 4 (57%) used meditation [15,17,18,22], 2 (29%) used relaxation [13,14], and 1 (14%) used both relaxation and meditation [16].

The yoga interventions ranged in duration from 2 to 12 weeks, with individual session lengths ranging from 10 to 90 minutes and frequency ranging from once a week to daily practice. Interestingly, one of the studies included 2 different prescribed yoga doses (low, 60 min/week, and moderate, 150 min/week) [23].

Most studies considered remotely delivered interventions as home practice and, therefore, did not assess or account for additional home practice. However, they accounted for all the

XSL•FO
RenderX

practices at home when reporting adherence (if reported). The study comparing in-person yoga with a yoga DVD specified instructions for the desired frequency of additional home practice [19]. See Table 4 for a summary of the yoga intervention characteristics.

Table 4. Information about intervention characteristics.[a]

Study	Asynchronous vs synchronous	Technology	Group vs individual	Yoga style	Yoga limbs	Duration (minutes per session)	Sessions per week	Number of weeks
Armstrong et al [14]	Asynchronous	Video (type unspecified)	Individual	NS[b]	Breathing, postures, and relaxation	30	4	10
Awdish et al [15]	Asynchronous	DVD and mobile app	Individual	Hatha and Iyengar	Breathing, postures, and meditation	NS	3 to 6	8
Donesky et al [16]	Synchronous	Videoconference	Group	Iyengar	Breathing, postures, meditation, and relaxation	60	2	8
Gunda et al [17]	Asynchronous	DVD	Individual	NS	Breathing, postures, and relaxation	60	3	12
Huberty et al [18]	Asynchronous	Web-based videos	Individual	Hatha and Vinyasa	Breathing, postures, and meditation	Requested 60 min/week	Requested 60 min/week	12
Huberty et al [23]	Asynchronous	Web-based videos	Individual	Gentle Hatha	Postures and meditation	60 min week (LD[c]); 150 min/week (HD[d])	60 min/week (LD); 150 min/week (HD)	12
Jasti et al [24]	Asynchronous and synchronous	Unspecified tele-yoga	NS	NS	Breathing and postures	40	≥1	4
Kyeongra et al [19]	Asynchronous	DVD	Individual	Vinyasa	Breathing and postures	Required one 90-minute session, encouraged 2 more for “home practice”	Required one 90-minute session, encouraged 2 more for “home practice”	8
Mullur et al [20]	Asynchronous	DVD	Individual	NS	Breathing and postures	10	As often as possible	12
Sakuma et al [21]	Asynchronous	DVD	Individual	NS	Breathing and postures	7.5	Daily	2
Schuver et al [22]	Asynchronous	DVD	Individual	Gentle Hatha	Breathing, postures, and meditation	60 to 75	2	12
Stan et al [13]	Asynchronous	DVD	Individual	NS	Breathing and postures	90	3 to 5	12

[a]Information about the intervention characteristics such as the delivery method, including whether the intervention was delivered synchronously or asynchronously; the type of technology used; whether the intervention was delivered to a group or individual; and yoga intervention components, including the yoga style, yoga limbs, and intervention dose (frequency and duration).

[b]NS: not specified.

[c]LD: low dose.

[d]HD: high dose.

Intervention Feasibility and Safety

Overview

The extracted data that were related to feasibility included information about adherence and occurrence or absence of technological challenges. The components of adherence were subcategorized into intervention adherence (ie, whether participants completed the intended yoga intervention dose) and study adherence (ie, whether participants completed the study overall). The following sections present the results related to adherence. We reported the results as specified in each study and used the term *compliance* instead of adherence for one of the studies that reported its results using that term.

Intervention Adherence

Adherence to the intervention was assessed or reported differently across studies. Approximately 33% (4/12) of studies

XSL•FO
RenderX

did not report on intervention adherence [14,15,17,20]. Approximately 25% (3/12) of studies [13,21,23] defined benchmarks or categories such as *compliance* or *good adherence* using specific thresholds. Approximately 42% (5/12) of studies reported intervention adherence through mean yoga practice using self-reported logs [18,19,22,23] or class attendance (synchronous [16] and in-person [19] interventions). In addition to self-reporting, 17% (2/12) of studies using web-based videos also used web analytics programs to monitor the time spent viewing these videos [18,23].

One of the studies set a benchmark for adherence (completion of 90% of the prescribed yoga dose in 9 out of 12 weeks). In this study, there was a low-dose group, in which 44% (8/18) of the participants met the benchmark, and a moderate-dose group, in which 6% (1/16) met the benchmark [23]. The study using the term *compliance* defined it as exercising >3 times per week for ≥7 weeks [13]. The authors showed that 39% (7/18) of the yoga group were compliant with the intervention as compared with 44% (7/16) in the strengthening group [13]. Finally, another study defined good adherence as those who practiced ≥6 times over 2 weeks and defined poor adherence as those who practiced 1 to 5 times over 2 weeks [21]. This study showed that 55% (37/67) of the yoga group had good adherence, 16% (11/67) had poor adherence, whereas 10% (7/67) did not report their adherence [21].

The mean yoga practice was also reported in several studies to indicate yoga intervention adherence. In the study comparing a yoga DVD with a walking DVD, a mean practice of 119.75 (SD 58.95) minutes for the yoga group was reported and a mean practice of 78.25 (SD 52.50) was reported for the walking group [22]. The target intervention dose for both groups was 120 minutes [22]. The study comparing in-person yoga with yoga delivered via a DVD reported that the in-person group practiced yoga for 75 minutes per week compared with 53.4 minutes per week in the DVD group [19]. The targeted dose was administered for 90 minutes per week. One of the studies using web-based videos showed mean yoga participation of 40.8 minutes per week by use of a software that counted how long they viewed web-based videos [18]. However, the self-reported mean practice of the same participants was 56 min/week. However, based on the self-reported practice measure, only 15% (4/27) of participants completed the required intervention dose of ≥60 minutes per week [18]. The other study using web-based videos showed, via self-report, that the low-dose group achieved a mean weekly yoga practice of 73% (44/60 minutes) of the target dose, and the moderate-dose group achieved a mean weekly yoga practice of 49% (77/150 minutes) of the prescribed dose [23]. However, web analytics revealed that these numbers overreported yoga practice [23]. Another study reported a mean yoga practice of 11.48 (SD 7.55) sessions but did not indicate a target dose [24]. Finally, the synchronous yoga study showed a mean attendance of 90% (14.5/16 required yoga classes) [16].

Study Adherence

In addition to intervention adherence, study adherence was reported as the number of individuals who completed the study compared with those who were enrolled. Approximately 33% (4/12) of studies did not report whether any individuals dropped out or whether all those enrolled completed the study [14,15,17,20]. In the study comparing yoga DVDs with a control group, 66% (44/67) of participants in the yoga group and 77% (24/31) of participants in the control group completed the study [21]. In the study comparing a yoga DVD with a strengthening DVD, 78% (14/18) of participants in the yoga group and 56% (9/16) of participants in the strengthening group completed the study [13]. The study comparing a yoga DVD with a walking program DVD reported that 90% (18/20) of participants in the yoga group and 80% (16/20) of participants in the walking group completed the study [22]. The study comparing yoga delivered via videoconferencing with an educational phone call control reported that 86% (6/7) of participants in the yoga group and 75% (6/8) of participants in the control group completed the study [16]. The study comparing in-person yoga with DVD yoga showed that 86% (6/7) of participants in the in-person group and 57% (4/7) of participants in the DVD group completed the study [19]. The study comparing web-based yoga videos with a wait-list control group reported that 79% (27/34) of participants in the yoga group and 75% (21/28) of participants in the control group completed the study [18]. The study comparing web-based yoga videos with a stretch and tone program reported that 57% (34/60) of participants in the yoga group and 47% (14/30) of participants in the control group completed the study [23]. Finally, the single-group study reported that 57% (54/95) of participants completed the study [24].

Technical Challenges and Satisfaction

The study implementing synchronous yoga reported technological challenges in 77% (24/31) of the yoga sessions and a mean enjoyment of 8.3 (SD 2.7) on a 10-point scale [16]. One of the studies using web-based videos reported that participants noted some technological challenges, attributing most to slow internet connections [23]. This study also noted high participant satisfaction in the web-based yoga video group and the web-based tone and stretch group [23]. The single-group study involving an unspecified tele-yoga program reported that 92.6% of the participants reported the intervention to be feasible (and safe) [24]. Finally, one of the studies assessed program satisfaction and found that the in-person group showed significantly greater satisfaction with the instruction method than the DVD group [19].

Adverse Events and Safety

Approximately 33% (4/12) of studies did not specify the occurrence or absence of adverse events [14,17,19,22]. Approximately 42% (5/12) of studies reported no adverse events [15,16,18,20,21]. One of the studies reported that 92.6% of the participants reported the intervention to be safe and feasible but did not provide further details [24]. One of the studies comparing a yoga DVD program with a strengthening DVD program for women with breast cancer reported 4 mild adverse events in the strengthening group and 9 mild adverse events in the yoga group [22]. However, the authors concluded that these mild adverse events could be attributed to recent reconstructive surgeries or medication side effects related to the participants' cancer treatment and management [22].

XSL•FO
RenderX

Preliminary Efficacy

Patient-Reported Outcome Measures

Patient-reported outcomes included various measures assessing anxiety, depression, sleep, fatigue, quality of life, general health, syncope functional status, multifactorial myeloproliferative neoplasm symptoms, sexual function, and pain. One of the studies did not analyze whether there were statistically significant differences between pre- and postmeasurements, despite having collected the data [18]. The single-group study showed significant improvements in perceived stress and yoga performance [24]. The study comparing a yoga DVD with a walking program DVD for women with depression showed within-group improvements but did not show between-group differences [22]. However, when controlling for baseline levels of rumination, the study showed significantly lower rumination in the yoga group [22]. The study comparing the use of a yoga DVD with a strengthening DVD showed significant within-group improvements in fatigue but did not show significant between-group differences [13]. The study that included childcare workers showed statistically significant improvements in low back pain, upper arm or neck pain, and menstrual pain in individuals who demonstrated good yoga intervention adherence [21]. The study comparing a web-based stretch or tone program with a web-based yoga program for women who had experienced stillbirth noted significant improvements in depression, perinatal grief, self-compassion, and self-rated health, favoring the yoga group [23]. Finally, the study that included individuals with neurocardiogenic syncope showed a statistically significant improvement on the Syncope Functional Status questionnaire following the yoga intervention [17].

Physical Performance and Function Outcome Measures

Physical performance and functional outcome measures comprised flexibility, strength, the 6-minute walk test, and balance. One of the studies assessing the effect of yoga on flexibility in older women showed statistically significant improvements in the sit and reach test in the yoga group, along with improvements in trunk extension, shoulder flexion, and left and right ankle flexibility [14]. In contrast, another study investigating the impact of a yoga DVD on childcare workers did not show any improvements in flexibility [21]. However, they did not specify the measure of flexibility used [21]. The studies assessing upper and lower extremity strength [16] and assessing grip strength [21] did not show any improvements. Approximately 17% (2/12) of studies [15,16] found no improvements in the 6-minute walk test; however, one of these studies [16] showed statistically significant improvements in shortness of breath and distress related to dyspnea in the yoga group following the 6-minute walk test. The study assessing balance used the functional reach test and did not report any improvements [21].

Physiological Outcome Measures

Physiological measures that showed significant improvements in at least one study included heart rate, blood pressure, oxygen saturation, presyncope and syncope events, and specific blood tests. One of the studies on veterans with diabetes showed statistically significant improvements in heart rate, diastolic blood pressure, and capillary blood glucose [20], whereas another study on individuals with neurocardiogenic syncope did not show any significant improvement in heart rate or blood pressure [17]. The study on individuals with neurocardiogenic syncope also showed statistically significant improvements in the number of presyncope and syncope events in participants who completed the yoga program [17]. See Table 5 for a summary of the results of each study for the patient-reported outcome measures, physical performance and performance outcome measures, and physiological measures.

XSL•FO
RenderX

Table 5. Outcome measures assessed in each study.

Study and type	Comparison	Outcomes
Armstrong et al [14], RCT[a]	Yoga video vs regular activity	• Sit and reach test[b] • Trunk extension[b] • Trunk flexion • Shoulder extension • Shoulder flexion[b] • Left ankle flexibility[b] • Right ankle flexibility[b]
Awdish et al [15], case series	Yoga video; no comparison	• Subjective changes via journaling[c] • Health Promoting Lifestyle Profile 2[c] • 6-minute walk test[c] • Oxygen saturation[c]
Donesky et al [16], quasi-experimental nonrandomized study	Yoga via videoconferencing vs health education phone call	• Safety: see the Adverse Events and Safety section • Acceptability: see the Intervention Adherence section • Technical issues: see the Patient-Reported Outcome Measures section • Upper and lower body muscle strength • 6-minute walk test • Symptoms following the 6-minute walk test[b] • Quality of life: St George's Respiratory Questionnaire and Kansas City Cardiomyopathy Questionnaire[d] • Depression Personal Health Questionnaire-8 • Overall dyspnea: Dyspnea-12 questionnaire • General Sleep Disturbance Scale
Gunda et al [17], quasi-experimental nonrandomized pilot study	Yoga DVD; control not clearly described	• Log of the number of presyncope and syncope events: in the intervention group, for those who finished the yoga regimen, there was a statistically significant improvement in the number of episodes of syncope and presyncope • SFSQ[e]: statistically significant decrease in the mean SFSQ score from the control phase to completion of the intervention phase • Head-up tilt table: resting heart rate • Blood pressure
Huberty et al [18], pilot RCT	Web-based yoga videos vs wait-list control	• Yoga participation: see the Intervention Adherence section • Adverse events: see the Adverse Events and Safety section • Blood draw feasibility and practicality • Inflammatory biomarkers[c] • Fatigue: single item from the multifactor MPN-SAF[c,f] • Multifactor MPN-SAF: total symptom score[c] • Quality of life: single item from the NIH[g] PROMIS[h] Global Health measure[c] • Sleep disturbance: Sleep Disturbance Scale • Short Form 8a[c] • Pain intensity: Pain Intensity Short Form 3a[c] • Anxiety distress: Emotional Anxiety Short Form 8a[c] • Depression emotional • Distress: Depression Short Form 8a[c] • Mental health: PROMIS[c] • Sexual function: 8-item for men; 10-item for women[c] • Physical health: PROMIS[c]

XSL•FO
RenderX

Study and type	Comparison	Outcomes
Huberty et al [23], RCT	Web-based yoga videos, including 2 different doses vs stretch and tone control	• Adherence: see the Intervention Adherence section • Acceptability: all groups achieved satisfaction benchmarks • Adverse events: see the Adverse Events and Safety section • Demand: no group met the demand benchmark • Impact of Event Scale • State-trait Anxiety Inventory • Patient Health Questionnaire-9[i] • Perinatal Grief Scale[i] • Self-Compassion Scale[i] • Self-rated health (Short Form-12)[i]: a significant decrease in low-dose and control groups • Emotion Regulation Questionnaire • Mindful Attention Awareness Scale • Pittsburg Sleep Quality Index
Jasti et al [24], single-group open-label trial	Tele-yoga module	• Adherence: see the Intervention Adherence section • Difficulty rating: mean score indicated that the yoga module was easy to practice • Feasibility: 92.6% of participants found the yoga to be safe and feasible • Yoga Performance Assessment scale[d] • Perceived Stress Scale[d]
Kyeongra et al [19], pilot RCT	Yoga DVD vs in-person yoga	• Adherence: see the Intervention Adherence section • Modifiable Activity Questionnaire • Program satisfaction: see the Patient-Reported Outcome Measures section
Mullur et al [20], pilot RCT	Yoga DVD vs handout about yoga	• Capillary blood glucose[b] • Heart rate[b] • Diastolic blood pressure[b] • Hemoglobin A1c • Systolic blood pressure • Weight • BMI
Sakuma et al [21], RCT	Yoga DVD vs regular activities	• Measure of body pain according to the Visual Analog Scale (range 0-100)[j] • Japanese version of the General Health Questionnaire • Body weight • BMI • Flexibility (measure not described reported in comparator) • Grip strength • Functional reach test
Schuver et al [22], pilot RCT	Yoga DVD vs walking DVD	• Beck Depression Inventory[d] • Ruminative Responses Scale[k]
Stan et al [13], pilot RCT	DVD vs strengthening DVD	• Feasibility: see the Intervention Adherence section • Safety: see the Adverse Events and Safety section • Fatigue: Multidimensional Fatigue Symptom Intervention Short Form[d] • Quality of life: Functional Assessment of Cancer Therapies–Breast[d]

[a]RCT: randomized controlled trial.

[b]Statistically significant between-group difference favoring the remote-delivered yoga group.

[c]Only effect sizes were calculated.

[d]Statistically significant within-group differences.

[e]SFSQ: Syncope Functional Status Questionnaire.

[f]MPN-SAF: Myeloproliferative Neoplasm Symptom Assessment Form.

XSL•FO
RenderX

[g]NIH: National Institutes of Health.

[h]PROMIS: Patient-Reported Outcomes Measurement Information System.

[i]Statistically significant improvements in the yoga group compared with the control group. Yoga comprised 2 different yoga intervention doses. For further details, refer to the study by Huberty et al [23].

[j]Statistically significant improvement for individuals who demonstrated good adherence to the yoga group (≥6 times per 2 weeks) for low back pain, upper arm or neck pain, and menstrual pain.

[k]Statistically significant between-group difference when controlling for baseline levels.

Discussion

Principal Findings and Comparison With Prior Work

The purpose of this scoping review was to examine the existing literature regarding the practice of yoga through remote delivery methods and identify current gaps related to (1) the populations studied, (2) the intervention characteristics (delivery methods and intervention components implemented), (3) the safety and feasibility of the interventions, and (4) the preliminary efficacy of the interventions. In summary, the studied populations included adults across their life spans, including individuals with physical and mental health conditions and some generally healthy adults. The review showed that, to date, most studies implementing remotely delivered yoga have implemented asynchronous delivery. In addition, the delivered interventions were primarily *Hatha yoga* interventions, including postures, breathing exercises, and meditation and relaxation exercises. The interventions were shown to be generally safe and feasible, with some feasibility challenges present in the study that implemented synchronous delivery. The heterogeneity of the included studies did not allow for an adequate evaluation of the preliminary efficacy of remotely delivered yoga interventions.

This scoping review showed that remotely delivered yoga has been successfully implemented in a heterogeneous sample of populations with and without chronic conditions. This is similar to the body of literature investigating the impact of in-person yoga [6]. In this review, studies included older adults; childcare workers; and individuals with chronic conditions such as COPD, HF, cancer, depression, diabetes, pulmonary hypertension, and cardiogenic syncope. Notably, most of these populations have also been implicated in in-person yoga studies. For instance, individuals with COPD [25] and HF [26], individuals with cancer [27,28], women with depression [29], older adults [30], adults with diabetes [31], and adults with cardiovascular conditions [32] have all been included in previous in-person yoga studies. Interestingly, no two studies reviewed here were conducted on the same population. Replication of studies for similar populations is seen in the in-person yoga literature [6] but is missing from the body of literature reviewed in this study. Therefore, replication using larger samples is needed. In addition, despite the variety of populations enrolled in the included studies, some populations that were commonly enrolled in in-person yoga studies were not included in the reviewed studies. For example, none of the included studies involved populations of individuals with neurological conditions or balance impairments. This differs from the body of literature investigating in-person yoga, which shows a large body of studies investigating yoga in these individuals [33,34].

It is unclear why some populations were enrolled in studies investigating remotely delivered yoga, whereas others were not. In an attempt to gain a better understanding of this, information related to the authors' motivation to implement remote delivery was extracted. However, as reported in the *Results* section, half of the studies (6/12, 50%) [15,17-19,22,23] did not provide any reason for choosing the remote delivery method. Previous guidelines for the development of yoga interventions have highlighted that intervention delivery must be considered in intervention design and are discussed in the second domain of *dose and delivery of yoga* [35]. There are multiple reasons why remote delivery may be beneficial, some of which have been mentioned in the reviewed studies. For example, previous studies have shown that multiple factors can impede engagement in exercise and physical activity, including transportation barriers, lack of time, decreased motivation, feelings of intimidation because of physical or environmental barriers, costliness, stigma, and a general lack of resources [36-38]. These barriers may also impede access to in-person yoga, and as such, a better understanding of facilitators and barriers associated with in-person and remote delivery methods is needed. In fact, previous studies support the notion that interventions delivered through technology show higher adherence in older adults [39].

When examining delivery method characteristics, most studies (10/12, 83%) implemented asynchronous delivery. This is similar to the results of another scoping review that investigated web-based mindfulness interventions for people with physical health conditions, which showed that 69% (11/16 studies) implemented asynchronous delivery [40]. Furthermore, another systematic review investigating remotely delivered therapy for mental health showed that 73% (8/11 studies) implemented asynchronous delivery [41]. This may be because asynchronous delivery requires fewer resources; however, this has not been formally established. Specifically, for the purpose of this review, no studies to date have compared synchronous and asynchronous remote delivery methods in yoga interventions. A total of 2 previous reviews also showed that there were no studies comparing synchronous and asynchronous remote delivery methods for their respective interventions: web-based mindfulness [40] and remotely delivered therapy [41]. One of the studies comparing synchronous and asynchronous delivery of tele-exercise for individuals with spinal cord injury showed significantly higher adherence and average weekly training load for the synchronous training group compared with the asynchronous group, suggesting that synchronous training may offer added benefits [42]. Thus, a current gap in the literature relates to the investigation of synchronous remote delivery of yoga, despite promising outcomes for other interventions [40-42].

Although more studies included in our review implemented asynchronous delivery than synchronous delivery, the one study investigating synchronous delivery reported high adherence and

XSL•FO
RenderX

enjoyment [16]. However, it also showed some feasibility challenges. It should be acknowledged that we cannot arrive at conclusions about the feasibility of all synchronously delivered, remotely delivered yoga interventions as there was only one study included in this review that implemented synchronous delivery. In addition, the other scoping reviews that investigated remotely delivered interventions did not speak to the feasibility of synchronously delivered interventions [40,41]. A previously published study indicated that certain videoconferencing platforms and practices may facilitate web-based engagement in remote research [43]. Thus, synchronous yoga interventions using widely available and commonly used videoconferencing platforms should be implemented and investigated to determine the approaches that limit technological challenges.

When examining the other intervention characteristics, such as the implemented style and limbs of yoga, our findings are similar to those found in the in-person yoga literature [44]. Although some of the studies reviewed here did not report the yoga style or limbs used, others reported a style of yoga that fell under the umbrella of *Hatha yoga* (the physical practice of yoga) without further specification. Limited available guidance on the optimal style may be is the reason behind a lack of specific reporting surrounding yoga style in general. In fact, a previously published systematic review [45] indicated that RCTs using different yoga styles did not differ significantly in their odds of achieving their desired outcomes. This demonstrated that there may not be an optimal style of yoga, and other factors such as preference and availability can help determine style [45]. However, future work should strive to report the style and limbs of yoga being implemented in interventions to help identify whether there are population-specific intervention characteristics that can optimize outcomes. For instance, a specific yoga style may be more beneficial for a given population. For example, a study comparing meditative yoga to power yoga for stress reduction in physically active, yoga-naive women showed significant improvements in salivary cortisol and state anxiety following a single meditative yoga session compared with power yoga [46], demonstrating that certain types of yoga may be more beneficial in specific situations. In addition, one yoga style may be delivered remotely more easily than another. For example, it may be more challenging for individuals with chronic health conditions and limited yoga experience to engage in a *Vinyasa* flow intervention involving rapid transitions from posture to posture while trying to follow a prerecorded video. However, this has yet to be investigated and could not be examined in this review or previously published reviews [6,45]. Thus, future studies should investigate the interactions among yoga style, study population, and other intervention characteristics, especially delivery method.

Limitations

This scoping review had some limitations. First, only studies with pre- and posttesting were included. This was stipulated in the eligibility criteria and was intended to facilitate the exploration of preliminary efficacy. However, this may have resulted in the exclusion of studies, such as feasibility studies or qualitative analyses, which could have provided additional insight into the current state of the literature. Specifically, we are aware of 4 studies [47-50] that were excluded because of a lack of explicit pre- and posttesting. Second, the heterogeneity of the included studies and assessed outcomes prevented the identification of conclusive preliminary efficacy findings. Third, the searches were completed in April 2021. With the continuation of the COVID-19 pandemic, publications involving telerehabilitation in general [51-53] and remotely delivered yoga interventions, in particular, may increase over the next few years. This could possibly require an update to this review. Regardless of these limitations, this scoping review identifies key gaps in the related literature and provides a strong foundation to optimize future research.

Strengths

This scoping review had several strengths. A broad range of databases was searched, which allowed a comprehensive search. This review provides a robust quality assessment of the included studies, provides a realistic picture of the literature and facilitates the interpretation of the findings. In addition, this review covers a broad range of content areas, including (1) the populations studied, (2) the intervention characteristics (delivery methods and intervention components implemented), (3) the safety and feasibility of the interventions, and (4) the preliminary efficacy of the interventions. This broad range of content allows readers to obtain a full picture of the state of the related literature and understand the current gaps. This is intended to help provide a path forward to optimize future research.

Future Directions

Multiple steps can be taken to address the gaps identified in this review and optimize future research. Future studies involving larger sample sizes should assess populations similar to those enrolled in the reviewed studies to determine whether the results can be replicated. In addition, populations not examined in the included studies, such as those with neurological conditions or other populations that have been shown to benefit from in-person yoga, should be enrolled in future studies that implement remotely delivered yoga. Future studies should justify the choice of delivery methods and relate this justification to population-specific needs. Moreover, future studies should consider and investigate the interactions among delivery methods, yoga intervention components, and other study characteristics. They should explore the implementation of synchronous delivery and compare different delivery methods. Specifically, synchronous yoga interventions using widely available and commonly used videoconferencing platforms should be investigated to determine whether this approach limits technological challenges and facilitates feasibility. Finally, future studies should report information regarding adverse events, adherence, and other safety and feasibility measures to provide robust information regarding the implementation of these interventions.

Conclusions

This review synthesized the literature regarding the remote delivery of yoga and provided information about gaps in the literature related to study populations, intervention characteristics, intervention safety and feasibility, and intervention efficacy. Overall, this review revealed a broad gap in the literature, showing that little attention has been paid to

XSL•FO
RenderX

yoga intervention delivery methods. Future studies and yoga intervention development guidelines should further consider the delivery methods when developing interventions. For instance, population-specific needs and barriers should be accounted for when determining delivery methods. In addition, more studies implementing synchronous delivery methods and studies comparing delivery methods should be conducted, and robust reporting of intervention characteristics is required.

Authors' Contributions

AJP approved the search strategy, completed the search, participated in the abstract and title screening, participated in the full-article review, extracted the results, synthesized the data, completed the first draft of the manuscript, and revised the manuscript.
EZA approved the search strategy, participated in the abstract and title screening, participated in the full-article review, extracted results, synthesized data, and revised the manuscript.
JFD approved the search strategy, participated in the abstract and title screening, participated in the full-article review, extracted the results, synthesized data, and revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The search strategy used for this scoping review, including the search terms used for each database, along with notes detailing the search methods.
[PDF File (Adobe PDF File), 103 KB - jmir_v24i6e29092_app1.pdf]

References

1. 2016 yoga in America study conducted by yoga journal and yoga alliance reveals growth and benefits of the practice. Yoga Alliance. URL: https://www.yogaalliance.org/Get_Involved/Media_Inquiries/2016_Yoga_in_America_Study_Conducted_by_Yoga_Journal_and_Yoga_Alliance_Reveals_Growth_and_Benefits_of_the_Practice [accessed 2021-04-01]
2. Penman S, Cohen M, Stevens P, Jackson S. Yoga in Australia: results of a national survey. Int J Yoga 2012 Jul;5(2):92-101 [FREE Full text] [doi: 10.4103/0973-6131.98217] [Medline: 22869991]
3. Cartwright T, Mason H, Porter A, Pilkington K. Yoga practice in the UK: a cross-sectional survey of motivation, health benefits and behaviours. BMJ Open 2020 Jan 12;10(1):e031848 [FREE Full text] [doi: 10.1136/bmjopen-2019-031848] [Medline: 31932388]
4. Park CL, Quinker D, Dobos G, Cramer H. Motivations for adopting and maintaining a yoga practice: a national cross-sectional survey. J Altern Complement Med 2019 Oct;25(10):1009-1014. [doi: 10.1089/acm.2019.0232] [Medline: 31460773]
5. Quilty MT, Saper RB, Goldstein R, Khalsa SB. Yoga in the real world: perceptions, motivators, barriers, and patterns of use. Glob Adv Health Med 2013 Jan;2(1):44-49 [FREE Full text] [doi: 10.7453/gahmj.2013.2.1.008] [Medline: 24381824]
6. Elwy AR, Johnston JM, Bormann JE, Hull A, Taylor SL. A systematic scoping review of complementary and alternative medicine mind and body practices to improve the health of veterans and military personnel. Med Care 2014 Dec;52(12 Suppl 5):S70-S82. [doi: 10.1097/MLR.0000000000000228] [Medline: 25397827]
7. Brems C, Justice L, Sulenes K, Girasa L, Ray J, Davis M, et al. Improving access to yoga: barriers to and motivators for practice among health professions students. Adv Mind Body Med 2015;29(3):6-13. [Medline: 26026151]
8. Becker WC, Dorflinger L, Edmond SN, Islam L, Heapy AA, Fraenkel L. Barriers and facilitators to use of non-pharmacological treatments in chronic pain. BMC Fam Pract 2017 Mar 20;18(1):41 [FREE Full text] [doi: 10.1186/s12875-017-0608-2] [Medline: 28320337]
9. Patañjali. The Yoga Sutras of Patanjali: The Book of the Spiritual Man. Bexar County, Texas: Bibliotech Press; 2020.
10. 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]
11. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
12. Thomas BH, Ciliska D, Dobbins M, Micucci S. A process for systematically reviewing the literature: providing the research evidence for public health nursing interventions. Worldviews Evid Based Nurs 2004;1(3):176-184. [doi: 10.1111/j.1524-475X.2004.04006.x] [Medline: 17163895]
13. Stan D, Croghan K, Croghan I, Jenkins S, Sutherland S, Cheville A, et al. Randomized pilot trial of yoga versus strengthening exercises in breast cancer survivors with cancer-related fatigue. Support Care Cancer 2016 Sep;24(9):4005-4015. [doi: 10.1007/s00520-016-3233-z] [Medline: 27129840]
14. Armstrong W, Smedley J. Effects of a home-based yoga exercise program on flexibility in older women. Clin Kinesiol 2003;57(1):6 [FREE Full text]

XSL•FO
RenderX

15. Awdish R, Small B, Cajigas H. Development of a modified yoga program for pulmonary hypertension: a case series. Altern Ther Health Med 2015;21(2):48-52. [Medline: 25830281]
16. Donesky D, Selman L, McDermott K, Citron T, Howie-Esquivel J. Evaluation of the feasibility of a home-based teleyoga intervention in participants with both chronic obstructive pulmonary disease and heart failure. J Altern Complement Med 2017 Sep;23(9):713-721. [doi: 10.1089/acm.2015.0279] [Medline: 28654302]
17. Gunda S, Kanmanthareddy A, Atkins D, Bommana S, Pimentel R, Drisko J, et al. Role of yoga as an adjunctive therapy in patients with neurocardiogenic syncope: a pilot study. J Interv Card Electrophysiol 2015 Aug;43(2):105-110. [doi: 10.1007/s10840-015-9996-1] [Medline: 25863799]
18. Huberty J, Eckert R, Dueck A, Kosiorek H, Larkey L, Gowin K, et al. Online yoga in myeloproliferative neoplasm patients: results of a randomized pilot trial to inform future research. BMC Complement Altern Med 2019 Jun 07;19(1):121 [FREE Full text] [doi: 10.1186/s12906-019-2530-8] [Medline: 31174535]
19. Yang K, James KA. Yoga, as a transitional platform to more active lifestyle: a 6-month pilot study in the USA. Health Promot Int 2016 Jun 18;31(2):423-429. [doi: 10.1093/heapro/dau108] [Medline: 25524471]
20. Mullur R, Ames D. Impact of a 10 minute seated yoga practice in the management of diabetes. J Yoga Phys Ther 2016 Jan 18;6(1):1000224 [FREE Full text] [doi: 10.4172/2157-7595.1000224] [Medline: 27774351]
21. Sakuma Y, Sasaki-Otomaru A, Ishida S, Kanoya Y, Arakawa C, Mochizuki Y, et al. Effect of a home-based simple yoga program in child-care workers: a randomized controlled trial. J Altern Complement Med 2012 Aug;18(8):769-776. [doi: 10.1089/acm.2011.0080] [Medline: 22808932]
22. Schuver KJ, Lewis BA. Mindfulness-based yoga intervention for women with depression. Complement Ther Med 2016 Jun;26:85-91. [doi: 10.1016/j.ctim.2016.03.003] [Medline: 27261987]
23. Huberty J, Sullivan M, Green J, Kurka J, Leiferman J, Gold K, et al. Online yoga to reduce post traumatic stress in women who have experienced stillbirth: a randomized control feasibility trial. BMC Complement Med Ther 2020 Jun 05;20(1):173 [FREE Full text] [doi: 10.1186/s12906-020-02926-3] [Medline: 32503517]
24. Jasti N, Bhargav H, George S, Varambally S, Gangadhar B. Tele-yoga for stress management: need of the hour during the COVID-19 pandemic and beyond? Asian J Psychiatr 2020 Dec;54:102334 [FREE Full text] [doi: 10.1016/j.ajp.2020.102334] [Medline: 32777755]
25. Donesky-Cuenco D, Nguyen HQ, Paul S, Carrieri-Kohlman V. Yoga therapy decreases dyspnea-related distress and improves functional performance in people with chronic obstructive pulmonary disease: a pilot study. J Altern Complement Med 2009 Mar;15(3):225-234 [FREE Full text] [doi: 10.1089/acm.2008.0389] [Medline: 19249998]
26. Howie-Esquivel J, Lee J, Collier G, Mehling W, Fleischmann K. Yoga in heart failure patients: a pilot study. J Card Fail 2010 Sep;16(9):742-749. [doi: 10.1016/j.cardfail.2010.04.011] [Medline: 20797598]
27. Greenlee H, DuPont-Reyes MJ, Balneaves LG, Carlson LE, Cohen MR, Deng G, et al. Clinical practice guidelines on the evidence-based use of integrative therapies during and after breast cancer treatment. CA Cancer J Clin 2017 May 06;67(3):194-232 [FREE Full text] [doi: 10.3322/caac.21397] [Medline: 28436999]
28. Buffart LM, van Uffelen JG, Riphagen II, Brug J, van Mechelen W, Brown WJ, et al. Physical and psychosocial benefits of yoga in cancer patients and survivors, a systematic review and meta-analysis of randomized controlled trials. BMC Cancer 2012 Nov 27;12:559 [FREE Full text] [doi: 10.1186/1471-2407-12-559] [Medline: 23181734]
29. Cramer H, Lauche R, Langhorst J, Dobos G. Yoga for depression: a systematic review and meta-analysis. Depress Anxiety 2013 Nov;30(11):1068-1083. [doi: 10.1002/da.22166] [Medline: 23922209]
30. Sivaramakrishnan D, Fitzsimons C, Kelly P, Ludwig K, Mutrie N, Saunders DH, et al. The effects of yoga compared to active and inactive controls on physical function and health related quality of life in older adults- systematic review and meta-analysis of randomised controlled trials. Int J Behav Nutr Phys Act 2019 Apr 05;16(1):33 [FREE Full text] [doi: 10.1186/s12966-019-0789-2] [Medline: 30953508]
31. Thind H, Lantini R, Balletto BL, Donahue ML, Salmoirago-Blotcher E, Bock BC, et al. The effects of yoga among adults with type 2 diabetes: a systematic review and meta-analysis. Prev Med 2017 Dec;105:116-126 [FREE Full text] [doi: 10.1016/j.ypmed.2017.08.017] [Medline: 28882745]
32. Cramer H, Lauche R, Haller H, Dobos G, Michalsen A. A systematic review of yoga for heart disease. Eur J Prev Cardiol 2015 Mar;22(3):284-295. [doi: 10.1177/2047487314523132] [Medline: 24491402]
33. Mooventhan A, Nivethitha L. Evidence based effects of yoga in neurological disorders. J Clin Neurosci 2017 Sep;43:61-67. [doi: 10.1016/j.jocn.2017.05.012] [Medline: 28599839]
34. Green E, Huynh A, Broussard L, Zunker B, Matthews J, Hilton CL, et al. Systematic review of yoga and balance: effect on adults with neuromuscular impairment. Am J Occup Ther 2019;73(1):7301205150p1-730120515011. [doi: 10.5014/ajot.2019.028944] [Medline: 30839270]
35. Sherman KJ. Guidelines for developing yoga interventions for randomized trials. Evid Based Complement Alternat Med 2012;2012:143271 [FREE Full text] [doi: 10.1155/2012/143271] [Medline: 23082079]
36. Rimmer JH, Riley B, Wang E, Rauworth A, Jurkowski J. Physical activity participation among persons with disabilities: barriers and facilitators. Am J Prev Med 2004 Jun;26(5):419-425. [doi: 10.1016/j.amepre.2004.02.002] [Medline: 15165658]
37. Syed ST, Gerber BS, Sharp LK. Traveling towards disease: transportation barriers to health care access. J Community Health 2013 Oct;38(5):976-993 [FREE Full text] [doi: 10.1007/s10900-013-9681-1] [Medline: 23543372]

XSL•FO
RenderX

38. Costello E, Kafchinski M, Vrazel J, Sullivan P. Motivators, barriers, and beliefs regarding physical activity in an older adult population. J Geriatr Phys Ther 2011;34(3):138-147. [doi: 10.1519/JPT.0b013e31820e0e71] [Medline: 21937904]
39. Valenzuela T, Okubo Y, Woodbury A, Lord SR, Delbaere K. Adherence to technology-based exercise programs in older adults: a systematic review. J Geriatr Phys Ther 2018;41(1):49-61. [doi: 10.1519/JPT.0000000000000095] [Medline: 27362526]
40. Toivonen KI, Zernicke K, Carlson LE. Web-based mindfulness interventions for people with physical health conditions: systematic review. J Med Internet Res 2017 Aug 31;19(8):e303 [FREE Full text] [doi: 10.2196/jmir.7487] [Medline: 28860106]
41. Sucala M, Schnur JB, Constantino MJ, Miller SJ, Brackman EH, Montgomery GH. The therapeutic relationship in e-therapy for mental health: a systematic review. J Med Internet Res 2012 Aug 02;14(4):e110 [FREE Full text] [doi: 10.2196/jmir.2084] [Medline: 22858538]
42. Costa RR, Dorneles JR, Veloso JH, Gonçalves CW, Neto FR. Synchronous and asynchronous tele-exercise during the coronavirus disease 2019 pandemic: comparisons of implementation and training load in individuals with spinal cord injury. J Telemed Telecare 2021 Jan 18:1357633X20982732 [FREE Full text] [doi: 10.1177/1357633X20982732] [Medline: 33461399]
43. Thayer EK, Pam M, Al Achkar M, Mentch L, Brown G, Kazmerski TM, et al. Best practices for virtual engagement of patient-centered outcomes research teams during and after the COVID-19 pandemic: qualitative study. J Particip Med 2021 Mar 11;13(1):e24966 [FREE Full text] [doi: 10.2196/24966] [Medline: 33646964]
44. Elwy AR, Groessl EJ, Eisen SV, Riley KE, Maiya M, Lee JP, et al. A systematic scoping review of yoga intervention components and study quality. Am J Prev Med 2014 Aug;47(2):220-232 [FREE Full text] [doi: 10.1016/j.amepre.2014.03.012] [Medline: 24996759]
45. Cramer H, Lauche R, Langhorst J, Dobos G. Is one yoga style better than another? A systematic review of associations of yoga style and conclusions in randomized yoga trials. Complement Ther Med 2016 Apr;25:178-187. [doi: 10.1016/j.ctim.2016.02.015] [Medline: 27062966]
46. Marshall M, McClanahan M, McArthur Warren S, Rogers R, Ballmann C. A comparison of the acute effects of different forms of yoga on physiological and psychological stress: a pilot study. Int J Environ Res Public Health 2020 Aug 21;17(17):6090 [FREE Full text] [doi: 10.3390/ijerph17176090] [Medline: 32825677]
47. Addington EL, Sohl SJ, Tooze JA, Danhauer SC. Convenient and Live Movement (CALM) for women undergoing breast cancer treatment: challenges and recommendations for internet-based yoga research. Complement Ther Med 2018 Apr;37:77-79 [FREE Full text] [doi: 10.1016/j.ctim.2018.02.001] [Medline: 29609942]
48. Guo SH, Lee C, Tsao C, Hsing H. A social media-based mindful yoga program for pregnant women in Taiwan. Stud Health Technol Inform 2016;225:621-622. [Medline: 27332280]
49. Johnson CC, Taylor AG, Anderson JG, Jones RA, Whaley DE. Feasibility and acceptability of an internet-based, African dance-modified yoga program for African-American women with or at risk for metabolic syndrome. J Yoga Phys Ther 2014;4:1000174 [FREE Full text] [doi: 10.4172/2157-7595.1000174] [Medline: 25593785]
50. Schulz-Heik RJ, Meyer H, Mahoney L, Stanton MV, Cho RH, Moore-Downing DP, et al. Results from a clinical yoga program for veterans: yoga via telehealth provides comparable satisfaction and health improvements to in-person yoga. BMC Complement Altern Med 2017 Apr 4;17(1):198 [FREE Full text] [doi: 10.1186/s12906-017-1705-4] [Medline: 28376861]
51. Ku B, Tse A, Pang B, Cheung N, Pang J, Chan J, et al. Tele-rehabilitation to combat rehabilitation service disruption during COVID-19 in Hong Kong: observational study. JMIR Rehabil Assist Technol 2021 Aug 19;8(3):e19946 [FREE Full text] [doi: 10.2196/19946] [Medline: 34254945]
52. Lai B, Chiu C, Pounds E, Tracy T, Mehta T, Young H, et al. COVID-19 modifications for remote teleassessment and teletraining of a complementary alternative medicine intervention for people with multiple sclerosis: protocol for a randomized controlled trial. JMIR Res Protoc 2020 Jul 03;9(7):e18415 [FREE Full text] [doi: 10.2196/18415] [Medline: 32540838]
53. Mukaino M, Tatemoto T, Kumazawa N, Tanabe S, Katoh M, Saitoh E, et al. An affordable, user-friendly telerehabilitation system assembled using existing technologies for individuals isolated with COVID-19: development and feasibility study. JMIR Rehabil Assist Technol 2020 Dec 10;7(2):e24960 [FREE Full text] [doi: 10.2196/24960] [Medline: 33279877]

Abbreviations

COPD: chronic obstructive pulmonary disease
HF: heart failure
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
RCT: randomized controlled trial

XSL•FO
RenderX

Edited by T Leung; submitted 25.03.21; peer-reviewed by J Huberty, J Johnson; comments to author 19.04.21; revised version received 10.05.21; accepted 26.04.22; published 06.06.22.

Please cite as:
James-Palmer A, Anderson EZ, Daneault JF
Remote Delivery of Yoga Interventions Through Technology: Scoping Review
J Med Internet Res 2022;24(6):e29092
URL: https://www.jmir.org/2022/6/e29092
doi:10.2196/29092
PMID:35666562

©Aurora James-Palmer, Ellen Zambo Anderson, Jean-Francois Daneault. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.06.2022. 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

Measurement of Adherence to mHealth Physical Activity Interventions and Exploration of the Factors That Affect the Adherence: Scoping Review and Proposed Framework

Yang Yang[1,2], MSc; Elisabeth Boulton[1,2], PhD; Chris Todd[1,2,3], PhD

[1]School of Health Sciences, Faculty of Biology, Medicine & Health, University of Manchester, Manchester, United Kingdom

[2]Manchester Academic Health Science Centre, Manchester, United Kingdom

[3]Manchester University NHS Foundation Trust, Manchester, United Kingdom

Corresponding Author:
Yang Yang, MSc
School of Health Sciences, Faculty of Biology, Medicine & Health
University of Manchester
Jean McFarlane Building
Manchester, M13 9PL
United Kingdom
Phone: 44 07422943521
Email: yang.yang-3@manchester.ac.uk

Abstract

Background: Mobile health (mHealth) is widely used as an innovative approach to delivering physical activity (PA) programs. Users' adherence to mHealth programs is important to ensure the effectiveness of mHealth-based programs.

Objective: Our primary aim was to review the literature on the methods used to assess adherence, factors that could affect users' adherence, and the investigation of the association between adherence and health outcomes. Our secondary aim was to develop a framework to understand the role of adherence in influencing the effectiveness of mHealth PA programs.

Methods: MEDLINE, PsycINFO, EMBASE, and CINAHL databases were searched to identify studies that evaluated the use of mHealth to promote PA in adults aged ≥18 years. We used critical interpretive synthesis methods to summarize the data collected.

Results: In total, 54 papers were included in this review. We identified 31 specific adherence measurement methods, which were summarized into 8 indicators; these indicators were mapped to 4 dimensions: length, breadth, depth, and interaction. Users' characteristics (5 factors), technology-related factors (12 factors), and contextual factors (1 factor) were reported to have impacts on adherence. The included studies reveal that adherence is significantly associated with intervention outcomes, including health behaviors, psychological indicators, and clinical indicators. A framework was developed based on these review findings.

Conclusions: This study developed an adherence framework linking together the adherence predictors, comprehensive adherence assessment, and clinical effectiveness. This framework could provide evidence for measuring adherence comprehensively and guide further studies on adherence to mHealth-based PA interventions. Future research should validate the utility of this proposed framework.

(J Med Internet Res 2022;24(6):e30817) doi:10.2196/30817

KEYWORDS

mobile health; mHealth; physical activity; adherence; framework; scoping review; mobile phone

Introduction

Background

There is strong evidence that physical activity (PA) is associated with improvements in physical health, mental health, and well-being [1]. However, many people are at risk of inactivity and there is poor uptake of, and adherence to, PA interventions [2]. Promoting PA and maintaining people's adherence are crucial public health issues.

Technological innovation (eg, mobile health [mHealth]) is developing rapidly and is being widely applied in the health care field [3]. mHealth mainly focuses on the delivery and

XSL•FO
RenderX

monitoring of health care services [4] and could also be an alternative approach to delivering PA interventions, overcoming the limitations of traditional PA approaches (eg, classes or workshops) [5]. Compared with traditional approaches, mHealth can use vivid video and pictures and may be more attractive and acceptable [6]. The use of mHealth can help deliver exercise programs to a wide audience at a low cost. In addition, mHealth technologies can provide timely feedback, reminders and support, continuous monitoring, and outcomes assessment [7,8].

There is a common issue with innovative health technologies, including mHealth, which is users' adherence to mHealth programs. For example, people who download an exercise app do not always use, or continue to use, the app. Research suggests that suboptimal exposure to the PA program lessens the effects of these interventions [9]. Measuring users' adherence to mHealth and exploring the factors that could influence users' adherence and the association between adherence and intervention outcomes are important to understand how PA and other outcomes can be improved.

Although there are systematic reviews summarizing evidence on the adherence to technology-based interventions, previous reviews are rather generic, do not differentiate mHealth from other technology-based interventions, do not focus on PA, and do not address the issue that adherence to a PA intervention may differ from adherence to other interventions such as those for medications or therapy. For example, Donkin et al [10] summarized measurement methods for adherence to any e-therapies and evaluated the association of adherence with intervention outcomes. Perski et al [11], focusing on digital behavior change interventions, developed a conceptual framework to explain the impacts of potential factors on people's engagement with digital behavior change interventions. By contrast, other reviews may focus on PA but restrict themselves to specific digital technologies. For example, Attig et al [12], focusing on wearable trackers for PA, summarized reported reasons for abandoning their use, such as usability issues and privacy concerns.

mHealth is a commonly used innovative solution to deliver health interventions. The characteristics of instant access, portability, and direct feedback make mHealth different from other technologies such as desktop computers [4]. We believe that these characteristics could affect the adherence of users of mHealth-based interventions in a way that is different from adherence to other technologies. Therefore, it would be better to consider mHealth-based interventions more specifically rather than grouping them together with other, generic technologies. Given this, we consider that the measurement of users' adherence to mHealth-based PA should not only reflect generic technologies' features, where applicable, but also incorporate mHealth-specific factors. For example, the automatically recorded number of days when mHealth devices are worn and the automatically recorded daily amount of PA, such as step counts, are commonly used indicators in measuring users' adherence to mHealth-based PA interventions [8]. However, such indicators have not been well considered in existing reviews that cover information technologies in general. Overall, there is a lack of evidence on determinants of the adherence to mHealth-based PA and the influence of adherence on health outcomes; in addition, there seems to be no agreed measurement method of adherence to mHealth devices that aim to improve PA engagement.

A scoping review of the literature was therefore carried out to explore how adherence to mHealth aiming at improving PA engagement is measured, to investigate which factors affect users' adherence, and the association between adherence and intervention outcomes. A framework is needed to identify the association among factors, adherence measurement, and health outcomes. The framework can be used in future to guide the measurement of adherence to mHealth-based PA programs and facilitate further research on the effectiveness of mHealth-based programs.

Objectives

The aims of this study were to synthesize evidence about (1) how adherence to mHealth PA interventions has been measured in the literature, (2) the factors that influence the adherence, and (3) the association between adherence to mHealth PA interventions and health outcomes. An additional aim was to propose an operational concept of adherence to mHealth-based PA programs and a framework for identifying the links among the determinants of adherence, adherence measurements, and intervention outcomes.

Methods

The PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines were followed for reporting this review [13]. Refer to Multimedia Appendix 1 for the PRISMA-ScR checklist. The review protocol was not registered on the web.

Literature Search and Search Strategy

To identify relevant studies, the initial electronic searches were run in September 2020 in 4 databases: MEDLINE, PsycINFO, EMBASE, and CINAHL [14]. Update searches were performed in February 2022. Refer to Multimedia Appendix 2 for the search strategies. There were no restrictions on publication year, but searches were limited to English language publications.

Eligibility Criteria

We considered a study eligible if it met the criteria presented in Table 1. Following the World Health Organization's definitions of mHealth [15], we defined mHealth-based PA programs as interventions that use mobile devices to deliver PA. The devices could be smartphones, smartwatches, PDAs such as wristbands, and other wireless technologies. In addition, there were no restrictions on how adherence to mHealth interventions was defined and assessed in this review as long as the authors described the measurement or definition of adherence. Furthermore, this review used the PA definition provided by the World Health Organization [16]: any bodily movement produced by skeletal muscles that requires energy expenditure.

XSL•FO
RenderX

Table 1. Inclusion and exclusion criteria.

Items	Inclusion criteria	Exclusion criteria
Types of study	• Any experimental and nonexperimental study design	• Unpublished studies • Papers that were not peer reviewed
Types of participants	• People aged ≥18 years (including older adults)	• Studies recruiting children (aged <18 years) or participants with cognitive impairment or psychiatric disorders
Types of interventions	• Studies that evaluated the use of mHealth[a] to promote PA[b] • mHealth devices could be used alone or in combination with other forms of interventions, such as physiotherapy. PA could be one part of the whole intervention, such as a behavior change program for weight	• Studies that delivered interventions using a desktop or laptop computer • Studies that used mHealth purely to monitor PA rather than deliver or guide PA
Types of outcomes	• Studies that measured any outcomes on the adherence to using mHealth to promote PA	• No exclusion criteria

[a]mHealth: mobile health.

[b]PA: physical activity.

Study Selection

We used EndNote (Clarivate Analytics) to manage records identified through the electronic searches. After removing duplicate records, we screened titles and abstracts at first to identify potentially eligible studies and then screened their full texts to include eligible studies. Given that this is a scoping review, only 1 reviewer (YY) was involved in this process. However, any problem was resolved by consulting another researcher (CT or EB).

Data Extraction

We used predefined data extraction forms and extracted the following details: (1) characteristics of the included studies: study design, study population, sample size, the description of the intervention, mHealth used, the goal of the intervention, control program, follow-up duration, and outcome measurement; and (2) factors that could influence adherence and the relationship between adherence and outcomes.

Data Synthesis

The critical interpretive synthesis approach was used to synthesize both qualitative and quantitative data [17]. Concepts identified in the full texts of included studies were labeled. The research questions were used as a top-down coding frame. We coded text fragments that were explicitly or implicitly related to any of the following three topics: (1) adherence measurement, (2) predictors of adherence, and (3) association between adherence and intervention outcomes. Synthetic constructs (ie, concepts that explain similar themes) were developed from the codes, and relationships between the synthetic constructs were specified.

Framework Development Methods

On the basis of the aforementioned synthetic constructs, we developed an integrative adherence framework by following the theoretical causal pathway to map the synthetic constructs of the scoping review. This framework can show the links among the determinants affecting adherence, multidimensional adherence measurements, and the association between adherence and health outcomes.

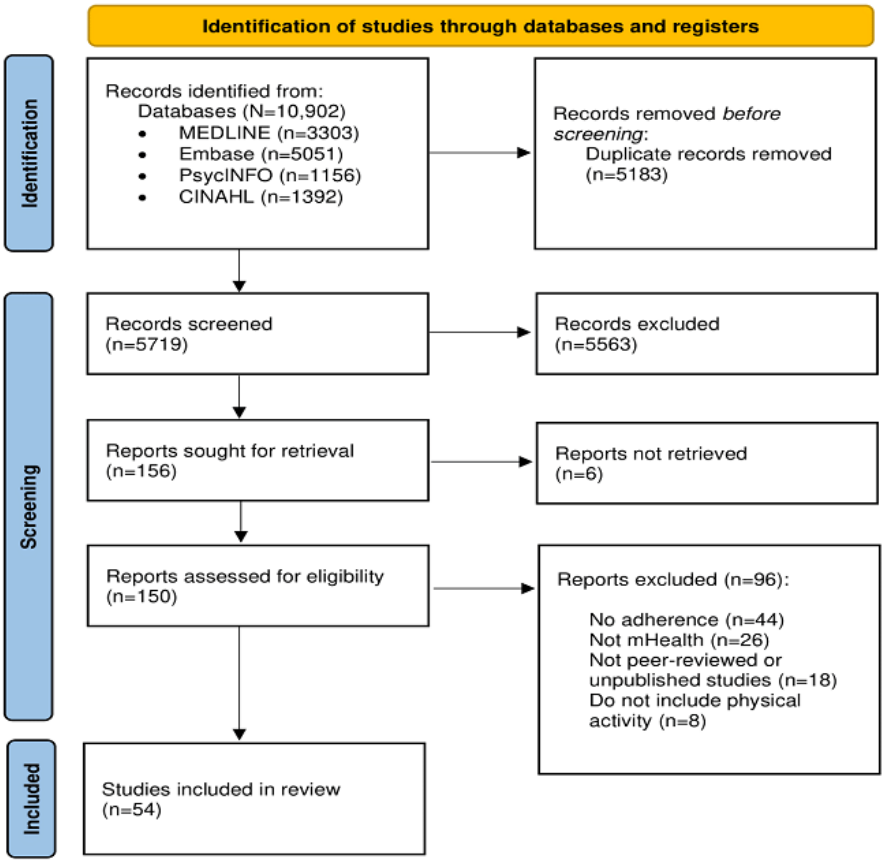
Results

Search Results

The electronic database searches retrieved a total of 10,902 records. Title and abstract screening of these 10,902 records resulted in 150 (1.38%) requiring full-text inspection. Of these 150 papers, 54 (36%) were included in this review (Figure 1).

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the study selection process. mHealth: mobile health.



Characteristics of the Included Studies

The characteristics of the included studies are summarized and reported in Textbox 1 (refer to Multimedia Appendix 3 [18-71] for full details). Most of the interventions were delivered through smartphone apps. The sample sizes of the included studies ranged from 10 to 16,948 (median 86). The average age of participants in the included studies ranged from 23.6 (SD 4.6) years to 73.2 (SD 7.3) years (median 51.7, SD 11.2).

XSL•FO
RenderX

Textbox 1. Summary characteristics of the included studies.

Types of studies

- Randomized controlled trial design (33 studies) [18,20-22,24,25,30,31,33,34,39-47,50-53,57,58,60,61,63,65-68,71], including 9 pilot studies with small sample sizes [20,21,33,40,42,47,50,51,68]
- Pre-post quasi-experimental studies with 1 arm (10 studies) [19,23,29,37,38,49,55,56,62,69]
- Observational study design (4 studies) [26,36,48,64]
- Subgroup analyses of the intervention groups of the randomized controlled trial (2 studies) [54,59]
- Nonrandomized controlled pilot trial (3 studies) [27,28,35]
- Nonrandomized 2-arm, matched case-control trial (1 study) [60]
- Cross-sectional web-based survey (1 study) [36]

Types of participants

- General population (such as university students and staff or healthy adults; 10 studies) [26,35,36,39,45,48,49,57,64,70]
- People with specific characteristics (17 studies)
  - Physically inactive community dwelling (4 studies) [41,53,59,65]
  - People who were overweight (4 studies) [24,68,69,71]
  - Older adults (4 studies) [22,23,31,37]
  - Pregnant or postpartum women who were overweight (2 studies) [28,60]
  - Shift workers (1 study) [47]
  - Mothers (1 study) [52]
  - Nurses (1 study) [55]
- People with specific diseases (27 studies)
  - Diabetes or at high risk of diabetes (9 studies) [19,32,38,42,43,46,50,51,63]
  - Cancer (5 studies) [20,44,56,61,62]
  - Cardiac event (4 studies) [18,30,54,67]
  - Musculoskeletal conditions (3 studies) [21,25,58]
  - Pulmonary disease (2 studies) [27,34]
  - Stroke (1 study) [33]
  - Parkinson disease (1 study) [40]
  - Excessive weight (in patients in primary care; 1 study) [66]
  - Patients awaiting surgery (1 study) [29]

Types of mobile health

- Delivering interventions through a smartphone app (52 studies) [18-55,57-70]
- Using a wrist-worn activity tracker (1 study) [56]
- Using a PDA (1 study) [71]

Functions of mobile health

- Helping users to self-monitor and document their behavior (such as physical activity, diet, and weight; 49 studies) [18-20,22-30,32-35,37-50,53-71]
- Providing feedback, reminders, and social support (38 studies) [18,22-24,28,30,32-35,38-43,45,47,49,50,53-69,71]
- Analytic and assessment features and setting activity targets or plan (38 studies) [18,19,22-25,28,32,33,35,38-45,47,51,53-57,59-61,63-68,70]
- Providing behavior change education and instruction (30 studies) [18,21,23,25,27,28,30-34,38,40-42,44-47,50,51,54,55,58,60,61,63,66,67,69]
- Game-based function (5 studies) [35,37,39,43,51]

Types of outcomes

- The feasibility of the mobile health interventions

XSL•FO
RenderX

- Adherence (27 studies) [18,21,23,25-27,29-31,33-35,40,42-44,50-54,56,58,67-69,71]
- Engagement (19 studies) [20,22,28,29,32,36-39,46-49,59-62,64,65]
- Retention (10 studies) [23,28,32,42,47,48,56,59,60,64]
- Acceptability (9 studies) [20,24,28,40,53,54,60,62,68]
- Usefulness or usability (7 studies) [23,47,49,50,54,57,70]
- Satisfaction (5 studies) [33,40,56,58,68]
- Recruitment (4 studies) [28,37,56,60]
- Uptake (4 studies) [30,42,44,57]
- Completion (3 studies) [44,59,67]
- Safety (2 studies) [40,56]
- Program use (2 studies) [57,70]
- Adoption (1 study) [55]
- Implementation (1 study) [55]
- Maintenance (1 study) [55]
- Fidelity (1 study) [60]

- Change in health behavior
  - Physical activity level (step count; 19 studies) [18,20,22,23,28,31,35,36,39,40,49,50,52,53,62,64-67]
  - Dieting (4 studies) [28,53,62,66]

- Clinical indicators
  - Change in weight or BMI (13 studies) [20,24,26,28,30,38,46,52,57,62,66,69,71]
  - Physical function and walking or exercise capacity (10 studies) [19,23,25,27,31,33,34,40,50,58]
  - Quality of life (9 studies) [19,31,33,34,39,40,50,65,67]
  - Glycated hemoglobin levels or fasting blood glucose (5 studies) [19,30,38,51,63]
  - Perceptions of treatment effectiveness (3 studies) [21,25,58]
  - Blood pressure (2 studies) [30,66]
  - Oxygen uptake peak (1 study) [18]

- Psychological indicators
  - Physical activity motivation (4 studies) [37,42,43,51]
  - Depression, anxiety stress, and mood (4 studies) [22,28,39,69]
  - Self-efficacy (5 studies) [18,22,28,29,66]
  - Education about heart-related health (1 study) [54]
  - Adverse events (2 studies) [23,27]
  - Cognitive performance (1 study) [22]
  - Disease knowledge (1 study) [18]

Follow-up

- Range: 3 weeks to 24 months (median 12 weeks)

Summary of Adherence Measurement Methods

We identified 31 specific adherence measurement methods used (Table 2). The top 3 most frequently used methods were manually entering self-monitored health behavior data into the device (17 studies), recording PA data (eg, step count) automatically recorded on the devices (10 studies), and recording the frequency of daily access to the app (8 studies).

These 31 measurement methods were related to 8 measurement indicators that generally reflect 4 measurement dimensions: length, breadth, depth, and interaction (Table 2). Among the 4 dimensions, the breadth dimension was the most frequently

XSL•FO
RenderX

measured (35 studies). Of the 54 included studies, 31 (57%) measured adherence in only 1 dimension, 15 (28%) measured adherence in 2 dimensions, and 6 (11%) measured adherence in 3 dimensions, whereas 2 (4%) included all 4 dimensions.

Table 2. Adherence measurement methods.

Dimensions and measurement indicators	Specific methods reported in the included studies
Length: the time users spend on the mHealth[a] devices (reported by 15 studies)	
Device use time and frequency	• Recorded the frequency of daily access to app (ie, app visit or log-in; 8 studies) [28,32,49,55,56,65,66,68] • Self-reported how frequently the app was used (2 studies) [36,51] • Recorded the duration of time spent on the device (2 studies) [32,60]
Duration of use until attrition	• Number of days devices were used (3 studies) [20,37,49] • Time to attrition (1 study) [48] • Trial retention (1 study) [68] • Duration of program use (1 study) [64]
Breadth: the proportion of mHealth functions and features used out of the total available (reported by 35 studies)	
Device functions used	• Having physical activity data (step count and exercise) automatically recorded on the devices (10 studies) [20,35,37,40,43,59,62,64,65,70] • Manually entering or uploading self-monitored health behavior data to the device (17 studies): physical activity [24,26,29,30,32,38,41,42,44,47,48,53,57,69,71], and meals, weight, and other behavioral targets [24,26,28,29,32,38,47,53,62,69] • Recorded actual use and each feature (6 studies) [22,37,39,42,46,49] • Recorded game played and the total duration of game played (1 study) [43]
Completion of modules	• The number of sessions attended, completed, or canceled by participants (6 studies) [23,27,48,52,54,69] • Self-reported adherence to treatment or the physical activity program assessed by means of standardized questionnaires (2 studies) [31,34] • Received counseling sessions (1 study) [41]
Depth: how well the program has been used (reported by 13 studies)	
Meeting tasks or challenges	• Self-reported adherence to the physical activity plan or target (3 studies) [31,50,58] • Points won when participants achieve their daily goals (2 studies) [51,59] • The duration of exercise performed vs prescribed (1 study) [54] • Attendance of the planned assessment (1 study) [67]
Behavior change (eg, physical activity level and diet habits)	• Devices monitored physical activity levels such as average daily step count or physical activity time (4 studies) [18,19,48,50] • Self-reported walking and sitting time (1 study) [33] and exercise time (1 study) [54] • Self-reported the number of behavior targets met (2 studies) [28,45]
Interaction: how users interact with the intervention programs (reported by 19 studies)	
Active interaction	• Writing, or responding to, a post (6 studies) [38,55,59-61,65] • Setting behavior change goals or challenges (6 studies) [22,28,45,60,63,64] • Receiving and responding to SMS text messages (4 studies) [21,22,38,62] • Sending digital gifts to teammates (2 studies) [59,65] • The number of notifications or prompts opened and responded to (1 study) [57] • Points earned when interacting with the program components (1 study) [63] • Join a Facebook group (1 study) [60]
Passive interaction	• Reading articles, texts, or watching video clips through app (6 studies) [25,28,38,41,48,60] • Completing telephone calls and the duration of calls (2 studies) [46,62] • The number of opened notifications (1 study) [32]

[a]mHealth: mobile health.

Factors That Affect Adherence to mHealth PA Programs

In the included studies, there are 3 factors affecting the adherence reported: user characteristics, technology-related factors, and contextual factors (Table 3). For the 5 specific user characteristics, the included studies showed inconsistent evidence on the influence of age, education status, and weight on adherence, whereas the studies by Ryan et al [59] and Guertler et al [64] consistently suggested that men had higher adherence. The study by Edney et al [39] reported that being overweight reduced adherence. The included studies consistently

XSL•FO
RenderX

showed that almost all 12 technology-related factors (9 mHealth functions and 3 specific factors related to the experience of using mHealth devices) increased users' adherence. Only 1 contextual factor was identified in this review: weekdays have higher adherence than weekends [59].

Table 3. Factors that affect adherence to mobile health (mHealth) physical activity interventions.

	Association with physical activity app adherence
User characteristics	

Age	• Inconsistent results
	  • Older age, more adherence (2 studies) [39,64]
	  • Unrelated to adherence (1 study) [51]

Sex	• Male with higher adherence (2 studies) [59,64]

Weight	• Inconsistent results
	  • Overweight reduced adherence (1 study) [39]
	  • Baseline weight or BMI was unrelated to adherence (1 study) [49]

Education	• Inconsistent results
	  • Middle education category with higher adherence (1 study) [59]
	  • Higher education status increased adherence (1 study) [46]

Baseline physical activity	• Baseline steps unrelated to adherence (1 study) [49]
Technology-related factors	
mHealth[a] functions	• Feedback on progress or motivation increased adherence (6 studies) [32,37,42,55,66,71] • Networking platforms or app-specific communities increased adherence (3 studies) [36,58,66] • Reminder feature increased adherence (3 studies) [57,59,66] • Access to historical physical activity data increased adherence (3 studies) [32,37,47] • Interpersonal contact function increased adherence (2 studies) [42,47] • Tailored interventions increased adherence (2 studies) [42,47] • Automation of data input increased adherence (1 study) [32] • Information update increased adherence (1 study) [32] • Multiple tasks decreased adherence (1 study) [55]
User experience	• Ease of use increased adherence (2 studies) [57,66] • Feeling challenged increased adherence (1 study) [37] • Fun-to-use intervention increased adherence (1 study) [66]
Contextual factors	Weekdays have higher adherence than weekends (1 study) [59]

[a]mHealth: mobile health.

Association Between Adherence and Intervention Outcomes

Associations between adherence and intervention outcomes were reported in 13 studies. Higher adherence was reported as having associations with higher PA in 5 studies [36,39,61,65,70] or with physical function in 1 study [27], with greater weight loss in 6 studies [26,38,39,46,59,71], and with greater reductions in glycated hemoglobin levels in 2 studies [38,51]. Höchsmann et al [43] reported that participants' adherence was positively associated with intrinsic PA motivation inventory change scores.

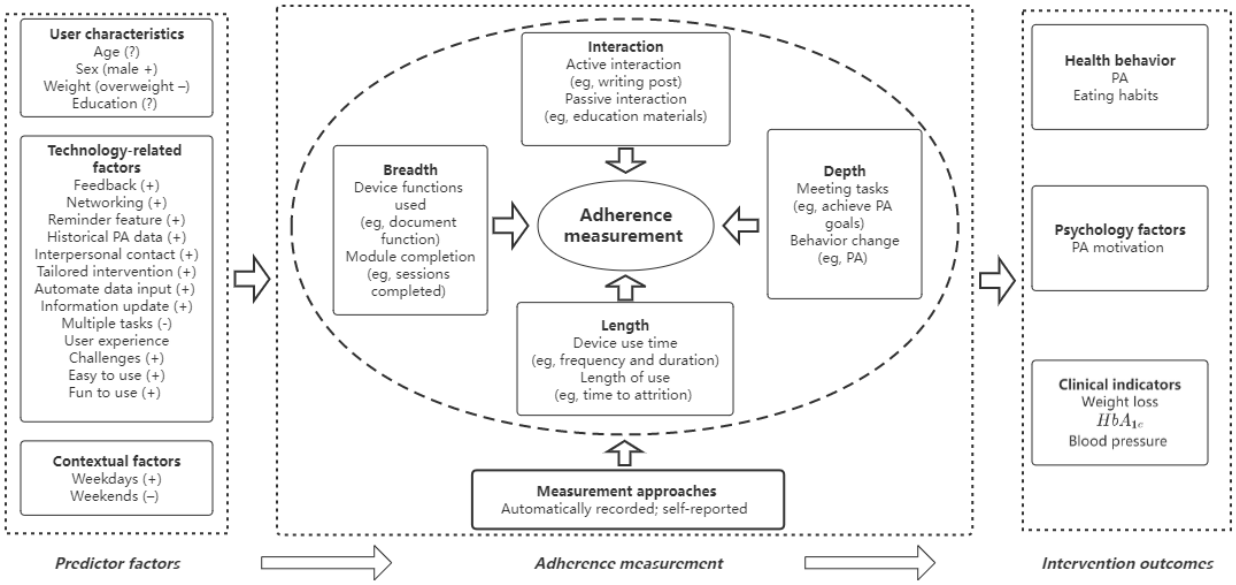
Proposed Framework of the Adherence to mHealth PA Interventions

The framework proposed by Perski et al [11] explicitly links potential influences on engagement and relationships between engagement and intervention effectiveness. We drew on the same structure for our adherence framework. We mapped our aforementioned review findings into a similar framework to illustrate interactions among the factors affecting adherence to mHealth PA interventions, the 4 dimensions of the adherence measurement, and intervention effectiveness (Figure 2).

Generally, most specific factors were related to aspects of user characteristics, the technology itself, and contexts that could increase users' adherence. The adherence to mHealth could be reflected by indicators associated with 4 dimensions (length, breadth, depth, and interaction) and could be measured largely by automatically recorded data and self-reported methods. Higher adherence could improve intervention effectiveness on health behaviors as well as psychological and clinical outcomes.

XSL•FO
RenderX

Figure 2. Framework of adherence to mHealth physical activity interventions. HbA_{1c}: glycated hemoglobin; PA: physical activity.



Discussion

Principal Findings

In this scoping review, we synthesized evidence from 54 studies. The findings are as follows:

1. Users' personal characteristics (eg, sex and education status), mHealth devices' features, and contextual factors were reported to increase users' adherence to mHealth PA interventions.
2. Adherence was reportedly reflected or measured in 4 dimensions and their corresponding indicators: breadth (device functions used and completion of modules), depth (meeting tasks or challenges and behavior change), length (device use time and length of use), and interaction (active or passive interaction with the program elements).
3. Higher adherence was associated with better outcomes in terms of health behaviors (eg, increasing PA), PA motivation, and clinical outcomes (eg, decrease in glycated hemoglobin levels, blood pressure, and weight).

These findings were summarized to propose an initial comprehensive adherence framework based on a theoretical causal pathway with three parts: (1) the factors that can affect users' adherence to mHealth PA interventions, (2) multidimensional adherence measurements, and (3) the association between adherence and intervention outcomes.

Adherence Measurement Issues

This scoping review identified 3 issues related to the measurement of adherence to mHealth PA programs in the literature. First, there is little consensus on how adherence should be defined and measured. In total, 31 specific adherence measurement methods were identified in the literature. The heterogeneity of the methods of assessing adherence makes comparison across studies difficult. Second, adherence measurement should be evidence based or theory informed; however, only the study by Adu et al [32] referred to a framework to measure adherence. Third, adherence to PA interventions is a complex definition rather than a single-dimension method [72], but many of the studies (31/54, 57%) included in this review measured adherence using only 1 dimension.

This review attempts to address the aforementioned issues through developing a new adherence conceptual framework based on the review findings. This framework, with 4 dimensions, is expected to inform the comprehensive measurement of users' adherence to mHealth PA interventions. The four dimensions are as follows:

1. The length dimension can reflect whether the users still use the mHealth devices and how much time they spend on the devices. This dimension should be considered essential in measuring adherence and can be reflected by the time to attrition, the frequency of access to the devices, and the duration of use.
2. The breadth dimension can reflect how many device functions are involved in PA and how many modules are completed by participants. This dimension can help understand the usability of each function of an mHealth device and the engagement and involvement of users. It was the most frequent consideration in measuring adherence in the included studies. This dimension is especially worth considering when the mHealth program has multiple functions or a number of intervention modules.
3. The depth dimension can assess how well users adhere to the task of an mHealth program. It can be measured by whether participants complete the program tasks or meet the PA target. For mHealth programs included in this review, tasks were usually related to behavior changes that could be considered to reflect the depth dimension. For example, the device recorded or self-reported PA time.
4. The interaction dimension reflects how users interact with mHealth programs. In the included studies, the most frequently used methods to assess interaction included recording and counting the number of posts written or users'

XSL•FO
RenderX

responses and assessing the users' access to educational materials.

Evidence on the topic of adherence or the related concept of engagement was included in 3 previous research studies [11,12,73]; however, all 3 differ from this review in terms of the focus of the interventions. Perski et al [11] developed a conceptual framework to highlight potential influences on engagement with a digital behavior change intervention and relationships between engagement and target behaviors. The authors' framework specified potential direct and indirect influences on engagement but did not aim to show how the engagement could be measured. The framework we have developed in this review focuses on the topic of adherence measurement and also explores the factors that influence users' adherence to mHealth. In addition, Perski et al [11] considered the engagement of users with a broad range of digital behavior change interventions rather than focusing on mHealth-based PA interventions. Attig et al [12] emphasized the exploration of factors that are related to the abandonment of an activity tracker rather than conceptualizing a framework for abandonment. Attig et al [12] identified less-intensive device use, less device interaction, and amount of PA achieved as important factors that affect abandonment of activity tracker use. We agree with the importance of these factors and have included them in our proposed adherence framework. Couper et al [73] analyzed data from a randomized controlled trial that aimed to evaluate the effectiveness of a web-based intervention in promoting dietary changes. They produced a 2D adherence measurement approach that consisted of breadth and depth. Breadth was defined as how widely users could access all available functions on the website, and depth was defined as how deeply users were engaged in the web-based material. Couper et al [73] considered that the *breadth-depth* engagement led to intervention outcomes: users' retention and behavior changes. However, we consider *retention* and *behavior change* as specific aspects of length and depth, respectively. This is because both *retention* and *behavior change* are only intermediate, moderating effects in terms of mHealth use patterns, rather than determining the influence of users' adherence on clinical outcomes resulting from PA, such as specific balance and strength outcomes. Given this, we consider our evidence-based adherence framework to be specifically relevant to mHealth-delivered PA interventions.

When choosing methods to measure adherence, researchers need to consider the following issues:

1. The characteristics of the mHealth program (eg, whether the mHealth program has multiple functions). The length and breadth dimensions are generally applicable to most mHealth programs, whereas the interaction and depth dimensions are usually applicable only to mHealth programs with relevant functions.
2. The purpose of the study. For example, if the study's purpose is to assess the effectiveness of the mHealth intervention, users' deep engagement with the mHealth intervention is needed, meaning that the interaction and depth dimensions should be considered.

The final issue regarding adherence measurement is how adherence data can be collected. Automatic recording and self-reporting adherence are 2 common approaches, but they both have strengths and limitations. Automatic recording can be objective but may not detect some specific types of activity (eg, stationary movement or upper or lower body movement) [74]. Self-reporting is easy to use and applicable to all types of PA, but the validity of this approach can be affected by social desirability bias [75]. The ideal way to measure adherence to mHealth PA interventions may be a combination of the objective measurement using mHealth and self-reporting.

Factors That Influence Users' Adherence

Some findings of this review are inconsistent with previous work: this review suggests that male sex, older age, and secondary education could be predictors of higher adherence. However, in a systematic review investigating the predictors of adherence to web-based psychological interventions, Beatty and Binnion [76] suggest that female sex and older age predict higher adherence. Another review finds that younger age is associated with higher adherence to internet interventions [77]. These disagreements may be related to the differences in the types of interventions and target populations considered.

This review found 12 technology-related factors that could affect users' adherence, including 9 related to mHealth functions and 3 related to users' experience. This is consistent with another systematic review in terms of technology-related factors (eg, PA tracking, PA goal setting, and customization of exercise) [78]. In the study by Attig et al [12], the reasons for abandonment included data inaccuracy, privacy concerns, discomfort, loss of motivation, and loss of tracking feasibility. This finding suggested that the mHealth devices themselves are the key to users' adherence, and in designing mHealth, the target users could be involved to improve device functions and thus adherence.

Adherence and Intervention Outcomes

As this review and the adherence framework demonstrates, adherence could affect intervention outcomes, including health behavior outcomes such as PA level, psychological indicators such as PA motivation, and clinical indicators such as weight loss and blood pressure. The findings suggest that further studies should explore the association between adherence and health outcomes from the aforementioned 3 perspectives.

Hawley-Hague et al [72] and Donkin et al [10] suggest that different types of clinical outcomes could be predicted using different types of adherence indicators (eg, log-ins were related to outcomes in physical health interventions, whereas module completion was most related to outcomes in psychological health interventions). Understanding how adherence influences the effectiveness of interventions could be crucial to understanding how adherence should be defined. However, this review identified only a limited number of studies that evaluate the association between adherence to mHealth and health outcomes. Further study is needed to explore in depth the association between each adherence dimension and each health outcome dimension.

XSL•FO
RenderX

Limitations of This Review

This review includes limitations. First, given that this is a scoping review, the searches for this review were limited to publications in English; hence, evidence published in other languages could have been missed. Second, the proposed adherence framework was based on the findings of the scoping review alone; therefore, it may not include all possible factors because of the limited evidence in the literature. Further research could improve this framework by considering evidence from other sources such as expert group meetings. Third, this framework explores the adherence predictors and the association between adherence and health outcomes. However, we did not further explore how these predictors affect each adherence dimension and the associations between each adherence dimension and intervention outcomes. This is because the included studies have limited evidence regarding these issues. Further studies are needed to explore the adherence to mHealth programs by each adherence dimension.

Conclusions

This review suggests that adherence can be measured using the dimensions of length, breadth, depth, and interaction; that users' characteristics, technology-related factors, and contextual factors can affect adherence; and that adherence is significantly associated with outcomes in terms of health behaviors, psychology, and clinical measures. These findings inform the development of a framework, linking together the adherence predictors, comprehensive adherence assessment, and clinical effectiveness. The framework could facilitate a comprehensive measurement of adherence as well as guide mHealth device development and further studies on adherence to mHealth PA interventions. Further research is needed to validate this framework; for example, by considering evidence from other sources such as expert group meetings or using Delphi approaches.

Acknowledgments

This study was funded by CT's National Institute for Health Research (NIHR) Senior Investigator Award (NIHR200299). CT is the Director and, at the time this paper was written, EB was a Research Fellow at the NIHR Older People and Frailty Policy Research Unit. CT leads the Healthy Ageing Research Group, and YY is employed by the NIHR Applied Research Collaboration Greater Manchester. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File , 85 KB - jmir_v24i6e30817_app1.docx]

Multimedia Appendix 2
Example of search strategies for the MEDLINE database.
[DOC File , 41 KB - jmir_v24i6e30817_app2.doc]

Multimedia Appendix 3
Characteristics of the included studies.
[DOC File , 198 KB - jmir_v24i6e30817_app3.doc]

References

1. Biddle SJ, Bennie JA, Bauman AE, Chau JY, Dunstan D, Owen N, et al. Too much sitting and all-cause mortality: is there a causal link? BMC Public Health 2016 Jul 26;16:635 [FREE Full text] [doi: 10.1186/s12889-016-3307-3] [Medline: 27456959]
2. Haseler T, Haseler C. Lack of physical activity is a global problem. BMJ 2022 Feb 23;376:o348 [FREE Full text] [doi: 10.1136/bmj.o348] [Medline: 35197324]
3. Lucivero F, Jongsma KR. A mobile revolution for healthcare? Setting the agenda for bioethics. J Med Ethics 2018 Oct;44(10):685-689 [FREE Full text] [doi: 10.1136/medethics-2017-104741] [Medline: 29907579]
4. Helbostad JL, Vereijken B, Becker C, Todd C, Taraldsen K, Pijnappels M, et al. Mobile health applications to promote active and healthy ageing. Sensors (Basel) 2017 Mar 18;17(3):622 [FREE Full text] [doi: 10.3390/s17030622] [Medline: 28335475]
5. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017 Apr;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]

XSL•FO
RenderX

6. Yerrakalva D, Yerrakalva D, Hajna S, Griffin S. Effects of mobile health app interventions on sedentary time, physical activity, and fitness in older adults: systematic review and meta-analysis. J Med Internet Res 2019 Nov 28;21(11):e14343 [FREE Full text] [doi: 10.2196/14343] [Medline: 31778121]
7. Valenzuela T, Okubo Y, Woodbury A, Lord SR, Delbaere K. Adherence to technology-based exercise programs in older adults: a systematic review. J Geriatr Phys Ther 2018;41(1):49-61. [doi: 10.1519/JPT.0000000000000095] [Medline: 27362526]
8. Jonkman NH, Del Panta V, Hoekstra T, Colpo M, van Schoor NM, Bandinelli S, et al. Predicting trajectories of functional decline in 60- to 70-year-old people. Gerontology 2018;64(3):212-221 [FREE Full text] [doi: 10.1159/000485135] [Medline: 29232671]
9. Gunnes M, Langhammer B, Aamot IL, Lydersen S, Ihle-Hansen H, Indredavik B, LAST Collaboration group. Adherence to a long-term physical activity and exercise program after stroke applied in a randomized controlled trial. Phys Ther 2019 Jan 01;99(1):74-85. [doi: 10.1093/ptj/pzy126] [Medline: 30329136]
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. 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]
12. Attig C, Franke T. Abandonment of personal quantification: a review and empirical study investigating reasons for wearable activity tracking attrition. Comput Human Behav 2020 Jan;102:223-237. [doi: 10.1016/j.chb.2019.08.025]
13. 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]
14. Betrán AP, Say L, Gülmezoglu AM, Allen T, Hampson L. Effectiveness of different databases in identifying studies for systematic reviews: experience from the WHO systematic review of maternal morbidity and mortality. BMC Med Res Methodol 2005 Jan 28;5(1):6 [FREE Full text] [doi: 10.1186/1471-2288-5-6] [Medline: 15679886]
15. World Health Organization. mHealth: new horizons for health through mobile technologies: second global survey on eHealth. World Health Organization. 2011. URL: https://apps.who.int/iris/bitstream/handle/10665/44607/9789241564250_eng.pdf?sequence=1&isAllowed=y [accessed 2020-09-28]
16. Global strategy on diet, physical activity and health: a framework to monitor and evaluate implementation. World Health Organization. 2006. URL: https://apps.who.int/iris/handle/10665/43524 [accessed 2020-09-28]
17. Dixon-Woods M, Sutton A, Shaw R, Miller T, Smith J, Young B, et al. Appraising qualitative research for inclusion in systematic reviews: a quantitative and qualitative comparison of three methods. J Health Serv Res Policy 2007 Jan;12(1):42-47. [doi: 10.1258/135581907779497486] [Medline: 17244397]
18. Cai C, Bao Z, Wu N, Wu F, Sun G, Yang G, et al. A novel model of home-based, patient-tailored and mobile application-guided cardiac telerehabilitation in patients with atrial fibrillation: a randomised controlled trial. Clin Rehabil 2022 Jan;36(1):40-50. [doi: 10.1177/02692155211032372] [Medline: 34266323]
19. Jiwani R, Wang J, Li C, Dennis B, Patel D, Gelfond J, et al. A behavioral lifestyle intervention to improve frailty in overweight or obese older adults with type 2 diabetes: a feasibility study. J Frailty Aging 2022;11(1):74-82 [FREE Full text] [doi: 10.14283/jfa.2021.17] [Medline: 35122094]
20. Allicock M, Kendzor D, Sedory A, Gabriel KP, Swartz MD, Thomas P, et al. A pilot and feasibility mobile health intervention to support healthy behaviors in African American breast cancer survivors. J Racial Ethn Health Disparities 2021 Feb;8(1):157-165. [doi: 10.1007/s40615-020-00767-x] [Medline: 32385847]
21. Anan T, Kajiki S, Oka H, Fujii T, Kawamata K, Mori K, et al. Effects of an artificial intelligence-assisted health program on workers with neck/shoulder pain/stiffness and low back pain: randomized controlled trial. JMIR Mhealth Uhealth 2021 Sep 24;9(9):e27535 [FREE Full text] [doi: 10.2196/27535] [Medline: 34559054]
22. Bisson AN, Sorrentino V, Lachman ME. Walking and daily affect among sedentary older adults measured using the StepMATE app: pilot randomized controlled trial. JMIR Mhealth Uhealth 2021 Dec 01;9(12):e27208 [FREE Full text] [doi: 10.2196/27208] [Medline: 34855609]
23. Daly RM, Gianoudis J, Hall T, Mundell NL, Maddison R. Feasibility, usability, and enjoyment of a home-based exercise program delivered via an exercise app for musculoskeletal health in community-dwelling older adults: short-term prospective pilot study. JMIR Mhealth Uhealth 2021 Jan 13;9(1):e21094 [FREE Full text] [doi: 10.2196/21094] [Medline: 33439147]
24. Eisenhauer CM, Brito F, Kupzyk K, Yoder A, Almeida F, Beller RJ, et al. Mobile health assisted self-monitoring is acceptable for supporting weight loss in rural men: a pragmatic randomized controlled feasibility trial. BMC Public Health 2021 Aug 18;21(1):1568 [FREE Full text] [doi: 10.1186/s12889-021-11618-7] [Medline: 34407782]
25. Gruner MP, Hogaboom N, Hasley I, Hoffman J, Gonzalez-Carta K, Cheville AL, et al. Prospective, single-blind, randomized controlled trial to evaluate the effectiveness of a digital exercise therapy application compared with conventional physical therapy for the treatment of nonoperative knee conditions. Arch Rehabil Res Clin Transl 2021 Aug 1;3(4):100151 [FREE Full text] [doi: 10.1016/j.arrct.2021.100151] [Medline: 34977534]

XSL•FO
RenderX

26. Han M, Rhee SY. Effect of adherence to smartphone app use on the long-term effectiveness of weight loss in developing and OECD countries: retrospective cohort study. JMIR Mhealth Uhealth 2021 Jul 12;9(7):e13496 [FREE Full text] [doi: 10.2196/13496] [Medline: 34255708]
27. Layton AM, Irwin AM, Mihalik EC, Fleisch E, Keating CL, DiMango EA, et al. Telerehabilitation using fitness application in patients with severe cystic fibrosis awaiting lung transplant: a pilot study. Int J Telemed Appl 2021 Feb 26;2021:6641853 [FREE Full text] [doi: 10.1155/2021/6641853] [Medline: 33727918]
28. Napolitano MA, Harrington CB, Patchen L, Ellis LP, Ma T, Chang K, et al. Feasibility of a digital intervention to promote healthy weight management among postpartum African American/Black women. Int J Environ Res Public Health 2021 Feb 23;18(4):2178 [FREE Full text] [doi: 10.3390/ijerph18042178] [Medline: 33672229]
29. Shelton E, Barreto NB, Bidwell S, Folk-Tolbert M, Shelton A, Trickey AW, et al. Engagement and adherence with a Web-based prehabilitation program for patients awaiting abdominal colorectal surgery. J Gastrointest Surg 2021 Dec;25(12):3198-3207. [doi: 10.1007/s11605-021-05171-2] [Medline: 34668165]
30. Yudi MB, Clark DJ, Tsang D, Jelinek M, Kalten K, Joshi SB, et al. SMARTphone-based, early cardiac REHABilitation in patients with acute coronary syndromes: a randomized controlled trial. Coron Artery Dis 2021 Aug 01;32(5):432-440. [doi: 10.1097/MCA.0000000000000938] [Medline: 32868661]
31. Taraldsen K, Mikolaizak AS, Maier AB, Mellone S, Boulton E, Aminian K, et al. Digital technology to deliver a lifestyle-integrated exercise intervention in young seniors-the PreventIT feasibility randomized controlled trial. Front Digit Health 2020 Jul 31;2:10 [FREE Full text] [doi: 10.3389/fdgth.2020.00010] [Medline: 34713023]
32. Adu MD, Malabu UH, Malau-Aduli AE, Drovandi A, Malau-Aduli BS. User retention and engagement with a mobile app intervention to support self-management in Australians with type 1 or type 2 diabetes (My Care Hub): mixed methods study. JMIR Mhealth Uhealth 2020 Jun 11;8(6):e17802 [FREE Full text] [doi: 10.2196/17802] [Medline: 32525491]
33. Grau-Pellicer M, Lalanza JF, Jovell-Fernández E, Capdevila L. Impact of mHealth technology on adherence to healthy PA after stroke: a randomized study. Top Stroke Rehabil 2020 Jul;27(5):354-368. [doi: 10.1080/10749357.2019.1691816] [Medline: 31790639]
34. Jiménez-Reguera B, Maroto López E, Fitch S, Juarros L, Sánchez Cortés M, Rodríguez Hermosa JL, et al. Development and preliminary evaluation of the effects of an mHealth Web-based platform (HappyAir) on adherence to a maintenance program after pulmonary rehabilitation in patients with Chronic Obstructive Pulmonary Disease: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jul 31;8(7):e18465 [FREE Full text] [doi: 10.2196/18465] [Medline: 32513646]
35. Liew SJ, Gorny AW, Tan CS, Müller-Riemenschneider F. A mobile health team challenge to promote stepping and stair climbing activities: exploratory feasibility study. JMIR Mhealth Uhealth 2020 Feb 04;8(2):e12665 [FREE Full text] [doi: 10.2196/12665] [Medline: 32014845]
36. Petersen JM, Kemps E, Lewis LK, Prichard I. Associations between commercial app use and physical activity: cross-sectional study. J Med Internet Res 2020 Jun 03;22(6):e17152 [FREE Full text] [doi: 10.2196/17152] [Medline: 32490836]
37. Tabak M, de Vette F, van Dijk H, Vollenbroek-Hutten M. A game-based, physical activity coaching application for older adults: design approach and user experience in daily life. Games Health J 2020 Jun;9(3):215-226. [doi: 10.1089/g4h.2018.0163] [Medline: 32053023]
38. Toro-Ramos T, Michaelides A, Anton M, Karim Z, Kang-Oh L, Argyrou C, et al. Mobile delivery of the diabetes prevention program in people with prediabetes: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jul 08;8(7):e17842 [FREE Full text] [doi: 10.2196/17842] [Medline: 32459631]
39. 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]
40. Ellis TD, Cavanaugh JT, DeAngelis T, Hendron K, Thomas CA, Saint-Hilaire M, et al. Comparative effectiveness of mHealth-supported exercise compared with exercise alone for people with Parkinson disease: randomized controlled pilot study. Phys Ther 2019 Feb 01;99(2):203-216. [doi: 10.1093/ptj/pzy131] [Medline: 30715489]
41. Fukuoka Y, Haskell W, Lin F, Vittinghoff E. Short- and long-term effects of a mobile phone app in conjunction with brief in-person counseling on physical activity among physically inactive women: the mPED randomized clinical trial. JAMA Netw Open 2019 May 03;2(5):e194281 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.4281] [Medline: 31125101]
42. Griauzde D, Kullgren JT, Liestenfeltz B, Ansari T, Johnson EH, Fedewa A, et al. A mobile phone-based program to promote healthy behaviors among adults with prediabetes who declined participation in free diabetes prevention programs: mixed-methods pilot randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 09;7(1):e11267 [FREE Full text] [doi: 10.2196/11267] [Medline: 30626566]
43. Höchsmann C, Infanger D, Klenk C, Königstein K, Walz SP, Schmidt-Trucksäss A. Effectiveness of a behavior change technique-based smartphone game to improve intrinsic motivation and physical activity adherence in patients with type 2 diabetes: randomized controlled trial. JMIR Serious Games 2019 Feb 13;7(1):e11444 [FREE Full text] [doi: 10.2196/11444] [Medline: 30758293]
44. Lee BJ, Park YH, Lee JY, Kim SJ, Jang Y, Lee JI. Smartphone application versus pedometer to promote physical activity in prostate cancer patients. Telemed J E Health 2019 Dec;25(12):1231-1236. [doi: 10.1089/tmj.2018.0233] [Medline: 30758247]

XSL•FO
RenderX

45. Ni Mhurchu C, Te Morenga L, Tupai-Firestone R, Grey J, Jiang Y, Jull A, et al. A co-designed mHealth programme to support healthy lifestyles in Māori and Pasifika peoples in New Zealand (OL@-OR@): a cluster-randomised controlled trial. Lancet Digit Health 2019 Oct;1(6):e298-e307 [FREE Full text] [doi: 10.1016/S2589-7500(19)30130-X] [Medline: 33323252]
46. Muralidharan S, Ranjani H, Mohan Anjana R, Jena S, Tandon N, Gupta Y, et al. Engagement and weight loss: results from the mobile health and diabetes trial. Diabetes Technol Ther 2019 Sep;21(9):507-513. [doi: 10.1089/dia.2019.0134] [Medline: 31184922]
47. 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]
48. Rayward AT, Vandelanotte C, Corry K, Van Itallie A, Duncan MJ. Impact of a social media campaign on reach, uptake, and engagement with a free Web- and app-based physical activity intervention: the 10,000 steps Australia program. Int J Environ Res Public Health 2019 Dec 12;16(24):5076 [FREE Full text] [doi: 10.3390/ijerph16245076] [Medline: 31842383]
49. Tong HL, Coiera E, Tong W, Wang Y, Quiroz JC, Martin P, et al. Efficacy of a mobile social networking intervention in promoting physical activity: quasi-experimental study. JMIR Mhealth Uhealth 2019 Mar 28;7(3):e12181 [FREE Full text] [doi: 10.2196/12181] [Medline: 30920379]
50. Valentiner LS, Thorsen IK, Kongstad MB, Brinkløv CF, Larsen RT, Karstoft K, et al. Effect of ecological momentary assessment, goal-setting and personalized phone-calls on adherence to interval walking training using the InterWalk application among patients with type 2 diabetes-a pilot randomized controlled trial. PLoS One 2019 Jan 10;14(1):e0208181 [FREE Full text] [doi: 10.1371/journal.pone.0208181] [Medline: 30629601]
51. 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]
52. Mascarenhas MN, Chan JM, Vittinghoff E, Van Blarigan EL, Hecht F. Increasing physical activity in mothers using video exercise groups and exercise mobile apps: randomized controlled trial. J Med Internet Res 2018 May 18;20(5):e179 [FREE Full text] [doi: 10.2196/jmir.9310] [Medline: 29776899]
53. Spring B, Pellegrini C, McFadden HG, Pfammatter AF, Stump TK, Siddique J, et al. Multicomponent mHealth intervention for large, sustained change in multiple diet and activity risk behaviors: the make better choices 2 randomized controlled trial. J Med Internet Res 2018 Jun 19;20(6):e10528 [FREE Full text] [doi: 10.2196/10528] [Medline: 29921561]
54. Salvi D, Ottaviano M, Muuraiskangas S, Martínez-Romero A, Vera-Muñoz C, Triantafyllidis A, et al. An m-Health system for education and motivation in cardiac rehabilitation: the experience of HeartCycle guided exercise. J Telemed Telecare 2018 May;24(4):303-316. [doi: 10.1177/1357633X17697501] [Medline: 28350282]
55. Torquati L, Kolbe-Alexander T, Pavey T, Leveritt M. Changing diet and physical activity in nurses: a pilot study and process evaluation highlighting challenges in workplace health promotion. J Nutr Educ Behav 2018;50(10):1015-1025. [doi: 10.1016/j.jneb.2017.12.001] [Medline: 29650395]
56. Trinh L, Arbour-Nicitopoulos KP, Sabiston CM, Berry SR, Loblaw A, Alibhai SM, et al. RiseTx: testing the feasibility of a Web application for reducing sedentary behavior among prostate cancer survivors receiving androgen deprivation therapy. Int J Behav Nutr Phys Act 2018 Jun 07;15(1):49 [FREE Full text] [doi: 10.1186/s12966-018-0686-0] [Medline: 29880049]
57. Levin ME, Pierce B, Schoendorff B. The acceptance and commitment therapy matrix mobile app: a pilot randomized trial on health behaviors. J Contextual Behav Sci 2017 Jul;6(3):268-275. [doi: 10.1016/j.jcbs.2017.05.003]
58. Lambert TE, Harvey LA, Avdalis C, Chen LW, Jeyalingam S, Pratt CA, et al. An app with remote support achieves better adherence to home exercise programs than paper handouts in people with musculoskeletal conditions: a randomised trial. J Physiother 2017 Jul;63(3):161-167 [FREE Full text] [doi: 10.1016/j.jphys.2017.05.015] [Medline: 28662834]
59. Ryan J, Edney S, Maher C. Engagement, compliance and retention with a gamified online social networking physical activity intervention. Transl Behav Med 2017 Dec;7(4):702-708 [FREE Full text] [doi: 10.1007/s13142-017-0499-8] [Medline: 28523603]
60. Willcox JC, Wilkinson SA, Lappas M, Ball K, Crawford D, McCarthy EA, et al. A mobile health intervention promoting healthy gestational weight gain for women entering pregnancy at a high body mass index: the txt4two pilot randomised controlled trial. BJOG 2017 Oct;124(11):1718-1728. [doi: 10.1111/1471-0528.14552] [Medline: 28220604]
61. Valle CG, Tate DF. Engagement of young adult cancer survivors within a Facebook-based physical activity intervention. Transl Behav Med 2017 Dec;7(4):667-679 [FREE Full text] [doi: 10.1007/s13142-017-0483-3] [Medline: 28374211]
62. Quintiliani LM, Mann DM, Puputti M, Quinn E, Bowen DJ. Pilot and feasibility test of a mobile health-supported behavioral counseling intervention for weight management among breast cancer survivors. JMIR Cancer 2016;2(1):e4 [FREE Full text] [doi: 10.2196/cancer.5305] [Medline: 27761518]
63. Block G, Azar KM, Romanelli RJ, Block TJ, Hopkins D, Carpenter HA, et al. Diabetes prevention and weight loss with a fully automated behavioral intervention by email, Web, and mobile phone: a randomized controlled trial among persons with prediabetes. J Med Internet Res 2015 Oct 23;17(10):e240 [FREE Full text] [doi: 10.2196/jmir.4897] [Medline: 26499966]

XSL•FO
RenderX

64. Guertler D, Vandelanotte C, Kirwan M, Duncan MJ. Engagement and nonusage attrition with a free physical activity promotion program: the case of 10,000 steps Australia. J Med Internet Res 2015 Jul 15;17(7):e176 [FREE Full text] [doi: 10.2196/jmir.4339] [Medline: 26180040]
65. Maher C, Ferguson M, Vandelanotte C, Plotnikoff R, De Bourdeaudhuij I, Thomas S, et al. A Web-based, social networking physical activity intervention for insufficiently active adults delivered via Facebook app: randomized controlled trial. J Med Internet Res 2015 Jul 13;17(7):e174 [FREE Full text] [doi: 10.2196/jmir.4086] [Medline: 26169067]
66. Laing BY, Mangione CM, Tseng CH, Leng M, Vaisberg E, Mahida M, et al. Effectiveness of a smartphone application for weight loss compared with usual care in overweight primary care patients: a randomized, controlled trial. Ann Intern Med 2014 Nov 18;161(10 Suppl):S5-12 [FREE Full text] [doi: 10.7326/M13-3005] [Medline: 25402403]
67. Varnfield M, Karunanithi M, Lee CK, Honeyman E, Arnold D, Ding H, et al. Smartphone-based home care model improved use of cardiac rehabilitation in postmyocardial infarction patients: results from a randomised controlled trial. Heart 2014 Nov;100(22):1770-1779 [FREE Full text] [doi: 10.1136/heartjnl-2014-305783] [Medline: 24973083]
68. 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]
69. Thomas JG, Wing RR. Health-e-call, a smartphone-assisted behavioral obesity treatment: pilot study. JMIR Mhealth Uhealth 2013 Apr 17;1(1):e3 [FREE Full text] [doi: 10.2196/mhealth.2164] [Medline: 25100672]
70. Kirwan M, Duncan MJ, Vandelanotte C, Mummery WK. Using smartphone technology to monitor physical activity in the 10,000 steps program: a matched case-control trial. J Med Internet Res 2012 Apr 20;14(2):e55 [FREE Full text] [doi: 10.2196/jmir.1950] [Medline: 22522112]
71. Burke LE, Conroy MB, Sereika SM, Elci OU, Styn MA, Acharya SD, et al. The effect of electronic self-monitoring on weight loss and dietary intake: a randomized behavioral weight loss trial. Obesity (Silver Spring) 2011 Feb;19(2):338-344 [FREE Full text] [doi: 10.1038/oby.2010.208] [Medline: 20847736]
72. Hawley-Hague H, Horne M, Skelton DA, Todd C. Review of how we should define (and measure) adherence in studies examining older adults' participation in exercise classes. BMJ Open 2016 Jun 23;6(6):e011560 [FREE Full text] [doi: 10.1136/bmjopen-2016-011560] [Medline: 27338884]
73. Couper MP, Alexander GL, Zhang N, Little RJ, Maddy N, Nowak MA, et al. Engagement and retention: measuring breadth and depth of participant use of an online intervention. J Med Internet Res 2010 Nov 18;12(4):e52 [FREE Full text] [doi: 10.2196/jmir.1430] [Medline: 21087922]
74. Sabia S, Cogranne P, van Hees VT, Bell JA, Elbaz A, Kivimaki M, et al. Physical activity and adiposity markers at older ages: accelerometer vs questionnaire data. J Am Med Dir Assoc 2015 May 01;16(5):438.e7-438.13 [FREE Full text] [doi: 10.1016/j.jamda.2015.01.086] [Medline: 25752539]
75. Mondal H, Mondal S. Social desirability bias: a confounding factor to consider in survey by self-administered questionnaire. Indian J Pharmacol 2018;50(3):143-144 [FREE Full text] [doi: 10.4103/ijp.IJP_15_17] [Medline: 30166752]
76. Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med 2016 Dec;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]
77. Christensen H, Griffiths KM, Farrer L. Adherence in Internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
78. Yang X, Ma L, Zhao X, Kankanhalli A. Factors influencing user's adherence to physical activity applications: a scoping literature review and future directions. Int J Med Inform 2020 Feb;134:104039. [doi: 10.1016/j.ijmedinf.2019.104039] [Medline: 31865054]

Abbreviations

mHealth: mobile health
PA: physical activity
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

XSL•FO
RenderX

Edited by A Mavragani; submitted 02.06.21; peer-reviewed by R Krukowski, L König; comments to author 04.08.21; revised version received 16.09.21; accepted 15.03.22; published 08.06.22.

Please cite as:
Yang Y, Boulton E, Todd C
Measurement of Adherence to mHealth Physical Activity Interventions and Exploration of the Factors That Affect the Adherence: Scoping Review and Proposed Framework
J Med Internet Res 2022;24(6):e30817
URL: https://www.jmir.org/2022/6/e30817
doi:10.2196/30817
PMID:35675111

©Yang Yang, Elisabeth Boulton, Chris Todd. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 08.06.2022. 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

Video Relay Interpretation and Overcoming Barriers in Health Care for Deaf Users: Scoping Review

Minerva Rivas Velarde[1], PhD; Caroline Jagoe[2], PhD; Jessica Cuculick[3], EdD

[1]Department of Radiology and Medical Informatics, Faculty of Medicine, University of Geneva, Geneve, Switzerland

[2]Department of Clinical Speech & Language Studies, Trinity College Dublin, University of Dublin, Dublin, Ireland

[3]Department of Liberal Studies, National Technical Institute for the Deaf, Rochester Institute of Technology, Rochester, NY, United States

Corresponding Author:
Minerva Rivas Velarde, PhD
Department of Radiology and Medical Informatics
Faculty of Medicine
University of Geneva
CMU 1 Rue Michel Servet
Geneve, 1211
Switzerland
Phone: 41 223794600
Fax: 41 22 379 4600
Email: minerva.rivas@unige.ch

Abstract

Background: Persons who are deaf are more likely to avoid health care providers than those who can hear, partially because of the lack of means of communication with these providers and the dearth of available interpreters. The use of video remote interpretation, namely the video camera on an electronic device, to connect deaf patients and health providers has rapidly expanded owing to its flexibility and advantageous cost compared with in-person sign language interpretation. Thus, we need to learn more about how this technology could effectively engage with and respond to the priorities of its users.

Objective: We aimed to identify existing evidence regarding the use of video remote interpretation (VRI) in health care settings and to assess whether VRI technology can enable deaf users to overcome barriers to interpretation and improve communication outcomes between them and health care personnel.

Methods: We conducted a search in 7 medical research databases (including MEDLINE, Web of Science, Embase, and Google Scholar) from 2006 including bibliographies and citations of relevant papers. The searches included articles in English, Spanish, and French. The eligibility criteria for study selection included original articles on the use of VRI for deaf or hard of hearing (DHH) sign language users for, or within, health care.

Results: From the original 176 articles identified, 120 were eliminated after reading the article title and abstract, and 41 articles were excluded after they were fully read. In total, 15 articles were included in this study: 4 studies were literature reviews, 4 were surveys, 3 were qualitative studies, and 1 was a mixed methods study that combined qualitative and quantitative data, 1 brief communication, 1 quality improvement report, and 1 secondary analysis. In this scoping review, we identified a knowledge gap regarding the quality of interpretation and training in sign language interpretation for health care. It also shows that this area is underresearched, and evidence is scant. All evidence came from high-income countries, which is particularly problematic given that most DHH persons live in low- and middle-income countries.

Conclusions: Furthering our understanding of the use of VRI technology is pertinent and relevant. The available literature shows that VRI may enable deaf users to overcome interpretation barriers and can potentially improve communication outcomes between them and health personnel within health care services. For VRI to be acceptable, sign language users require a VRI system supported by devices with large screens and a reliable internet connection, as well as qualified interpreters trained on medical interpretation.

(J Med Internet Res 2022;24(6):e32439) doi:10.2196/32439

XSL•FO
RenderX

KEYWORDS

deafness; disability; accessibility; communication; video; remote interpretation; health care; system; deaf users; sign language; interpreter; medical interpretation; mobile phone

Introduction

Background

Accessible information and communications technology (ICT), mobile phones, and tools such as video remote interpretation (VRI) aim to enable effective communication between persons who are D/deaf ("Deaf" refers to the linguistic minority while "deaf" refers to persons with hearing impairment) and hard of hearing and those who use sign language as their first language (hereafter, deaf or hard of hearing [DHH] sign language users) and health care personnel. VRI refers to a video camera on an electronic device, either a computer or tablet, that is used to connect patients and health providers with a sign language interpreter via video call. Its use has rapidly expanded owing to its flexibility and advantageous cost compared with in-person sign language interpretation [1]. The cost-efficiency of such technology is a serious concern given that 80% of the DHH population live in low- and middle-income countries (LMICs), where resource constraints tend to limit the availability of qualified sign language interpreters [2]. VRI aims to overcome communication barriers in health care. DHH persons are more likely to avoid health care providers than those who can hear, partially because of the lack of means of communication with these providers and the dearth of available interpreters [3,4]. Even if interpreters are available, the pool of sign language interpreters tends to be relatively narrow, even in high-income contexts [5]. Forthcoming research suggests that general sign language training does not cover skills to work effectively within the health care context; therefore, issues arise from the limited number of interpreters and their lack of skills [6-8]. Furthermore, health care personnel tend to lack awareness about working with sign language interpreters, alongside limited awareness of deaf communities in general. This results in poor communication, and ultimately, patients do not obtain the information they need to decide on their health or treatment [5].

DHH populations tend to be particularly disadvantaged compared with other persons with a disability. They tend to occupy poorer socioeconomic positions, hold lower health literacy, have insufficient knowledge of health-related vocabulary, and are often unaware of their family medical histories, all of which prevent them from outlining risk factors for their health [9]. DHH individuals have a greater prevalence of obesity, higher levels of hypertension, and higher levels of self-reported depression compared with hearing persons [6,9,10]. There is also a particular concern of underdiagnoses of raised blood pressure and undertreatment of hypertension, diabetes, hyperlipidemia, and cardiovascular disease, among others, due to the lack of effective means of communication between health personnel and deaf patients [6,9-11]. Recent studies claim that by improving communication between deaf persons and nondeaf persons hearing health personnel would have a positive impact on preventive care [12-14].

Objective

The rapid adoption of VRI technology in health care opens up opportunities to set up more accessible health care. Thus, we need to learn more about how this technology could effectively engage with and respond to the priorities of its users. Emerging literature shows that DHH users tend to prefer in-person to VRI interpretation [15-17]. Furthermore, satisfaction with VRI interpretation tends to be low [15]. We do not have evidence on whether users are comparing interpreters with the same level of skills one via VRI and one in-person, so they are comparing the sentiment of indeed like with like or not. Thus, we need more clarity on the elements of VRI systems that have been examined, such as procedures, available protocols, challenges, and successes. Having detailed data, all elements regarding in-person and VRI interpretation protocols would allow determining the technology that holds some constraints more clearly or the protocol could be improved and made more efficient. It is also necessary to identify the essential elements of VRI as a precondition to encourage rigorous studies and ensure fidelity when implemented. The scoping review approach chosen for this study will allow us to determine the state of available evidence, which is needed before rigorous empirical studies are conducted. Therefore, for the purpose of this study, we used the guidance for conducting systematic scoping reviews by Peter et al [18] to determine the following with respect to the use of VRI in the health care context: does the existing literature provide sufficient evidence on how VRI can enable deaf users to overcome interpretation barriers and improve communication outcomes between them and health care personnel within health care settings?

Methods

Overview

In this review, we identified relevant studies in English, Spanish, and French published between 2006 when the first relevant publication in the area was identified and March 2021 in PubMed, Web of Science, Embase, MEDLINE, and Google Scholar. The key search terms used were as follows: Sign language user*s, Deaf, Hard of Hearing, Deafblind and VRI, video remote sign language interpretation, video interpreting service, video conference interpreting and community health, health system, and health personnel. The search also covered all types of health-related activities that are often linked to community health. See search strategies in Multimedia Appendix 1.

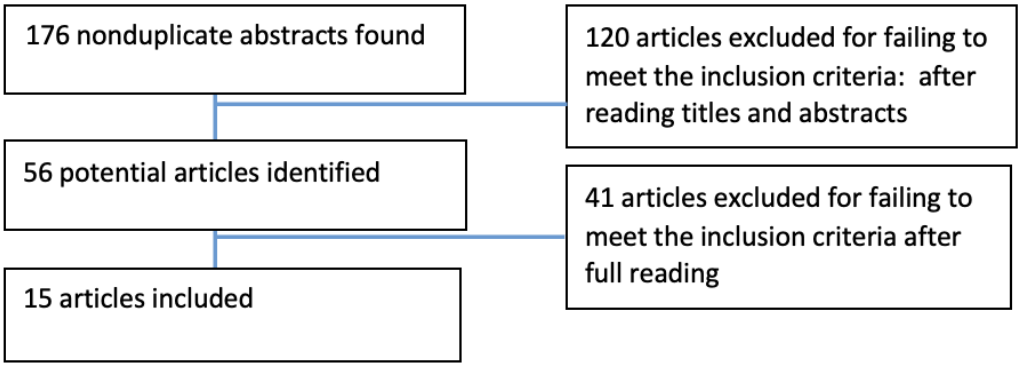
Study Selection

Articles were included for full-text reviews if they were about the use of VRI for DHH users for, or within, health care. Titles and abstracts were screened, and if an article was considered representative of the inclusion criteria, the full text was reviewed. Data extraction was conducted by 2 reviewers, independently, on 20% of the papers. The discrepancies were minimal.

XSL•FO
RenderX

If the paper was selected for full review, data related to the use of VRI for sign language users within the health care context were extracted. Data extracted from the articles that reported on the analysis, use, or implementation of VRI within the health care context were entered into an Excel (Microsoft Inc) form. Key findings were extracted in a summary format. Information on authorship, publication year, article type, methodology, population, lessons learned, and recommendations regarding the use of VRI were recorded in this form (Figure 1).

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



Analysis

We conducted an inductive content analysis of the selected records following the steps outlined by Elo [19]. The extracted findings from each study were subjected to open coding, and similar codes across articles were then identified as concepts coded inductively into the key concepts. Finally, in line with the aims of the study, the concepts were grouped into either the advantages of VRI or the challenges or limitations of VRI.

Patient and Public Involvement

This study was performed without the involvement of DHH patients. However, it does involve organizations for DHH individuals as well as persons with a disability. The National Deaf Federation of Colombia (FENASCOL) advised MRV on the pertinence of this research. JC, a DHH scientist, has coauthored this paper, contributing to its conceptualization, interpretation of the results, and attainment of clarity and accuracy of the writing.

Ethics Approval

The research protocol of this study was approved by the ethics committee of the University of Geneva (CUREG_2021-05-50).

Results

Overview

From the original 176 articles identified, 120 were eliminated after reading the article title and abstract, and 41 articles were excluded after they were fully read. In total, 15 articles were included in this study: 4 studies were literature reviews, 4 were surveys, 3 were qualitative studies, and 1 was a mixed methods study that combined qualitative and quantitative data, 1 brief communication, 1 quality improvement report, and 1 secondary analysis. Table 1 includes summaries of the articles that met our inclusion criteria.

There is limited research on the use and efficiency of VRI to improve communication between DHH individuals and health personnel within health settings. The current published scientific literature does not allow us to understand either the use of this technology or its impact on quality of care, patient satisfaction, or health outcomes. Nearly half (n=7, 46%) of the articles included empirical evidence on adult DHH VRI users, 1 (6%) on DHH children, 1 (6%) on sign language interpreters, and 1 (6%) on subject matter experts working with older DHH adults. Less than half (n=6, 40%) of the articles explicitly addressed the role of DHH persons as coauthors of the articles and steps followed to fulfill ethical and moral obligations of putting the voice of the DHH population at the center of their research, promoting well-being and the human rights of this population.

A limitation of the available literature is the lack of representation of the DHH population as a whole, given that all the articles are from high-income countries, namely 12 from the United States, 1 from Denmark, 1 from Norway, and 1 from Canada. This is a significant gap, given that 80% of persons with disabling hearing loss live in LMICs [19]. Currently, resource constraints and other social and political barriers in LMICs that could affect the availability, use, and efficiency of sign language interpretation via VRI within health care are not included in the published literature.

The current literature shows the key advantages of pursuing improvements in this technology as well as some recurring challenges and limitations (Textbox 1).

XSL•FO
RenderX

Table 1. Summaries of studies included in this review.

Study	Country	Aims	Study population and sample size	Design	Duration of the intervention	Main findings that related to the use of VRI[a] within health care
Berry and Stewart, 2006 [20]	United States	To outline challenges that D/deaf[b] people face within health care. It outlines recommendations to ensure a successful medical visit.	D/deaf	Literature review	No information	• Suggest capacity building for medical staff regarding communication needs of D/deaf patients. It provides a protocol to identify interpreters, as well as a list of tips for working with an interpreter, such as speaking to patients when using an interpreter.
Steinberg et al, 2006 [21]	United States	To better understand the health care experiences of deaf people who communicate in ASL[c]	Participants were deaf, communication preference for ASL, and willingness to share health care experiences	Qualitative studies (semistructured focus group meetings)	No information	• It points out that fear, mistrust, and frustration were prominent in participants' descriptions of health care encounters, as well as a list of inadequate common practices such as writing notes and using family members as interpreters.
Masland et al, 2010 [1]	United States	This study reviews published literature and unpublished data, documenting the use of telephonic and video interpretation methodologies to improve health care communication.	Published and unpublished literature on the interpretation in health care	Brief communication	No information	• This study looks at the cost-effectiveness of VRI for all language translation including sign language. • VRI advantages outlined in the study are flexibility, convenience, quality of interpretation, and cost. Some arguments are made that the savings in hiring an ASL interpreter can pay for the installation of video interpretation networks in some hospitals. • The results linked the use of VRI to fewer tests, less visits to the hospital, and better treatment adherence. However, evidence represented is in spoken leagues not sign language.
Hommes et al, 2018 [22]	United States	This research aimed to identify ASL interpreters' perceptions of barriers to effective communication between deaf and HOH[d] patients and health care providers.	ASL interpreters	A cross-sectional survey	June 15	• The results indicated that VRI technology in the absence of an ASL interpreter is considered a better option by many deaf and HOH patients than note-writing or lip-reading; however, the occasional technology malfunctions limit it as a consistently reliable tool.

XSL•FO
RenderX

Study	Country	Aims	Study population and sample size	Design	Duration of the intervention	Main findings that related to the use of VRI[a] within health care
Dammeyer et al, 2017 [23]	Denmark	This study examined the prevalence of technology use and interpreting services use among people with hearing loss as they relate to demographic characteristics of this population.	269 children (0-15 years of age) and 839 adults (16-65 years of age)	National surveys of children and adults with hearing loss	2014	• This study found that sign language users, both children and adults, prefer VRI over other communication technology. Adults with a bachelor's degree or higher reported more frequent use of mobile video interpretation and texting devices. • This study underlines the need for a user-centered approach and user involvements to address environmental and personal factors affecting assistive technology use. It recommends that deaf people may benefit from accessing well-trained personnel who understand the individual's needs and facilitate technology-person match.
Myers et al, 2021 [16]	United States	To examine the extent to which communication aids and services used by ASL users and their health care providers aligns with preferences, satisfaction, and unmet needs and to elicit from stakeholders' strategies to address disparities	ASL users in North Carolina	Web-based survey (cross-sectional study)	May 2018 until March 2019	• The study found that accessible communication was associated with 81% lower odds of dissatisfaction with communication. • Better communication was linked to better relationships with the health providers. The study claims that improving communication would have a positive impact on preventive care. • The study identifies several issues with the use of VRI. One of the most common barriers to accessible communication via VRI were technical problems, as well as quality of sign language interpreting services. • Communication via VRI was considered not user-friendly, creating frustrations for both deaf individuals and their professional health care providers. Health providers attempted to adapt to VRI issues by lipreading or speech or writing notes back-and-forth, both methods were inadequate and did not lead to improved communication. • The study made specific technical recommendations on when and how to use VRI in clinical settings.
Kushalnagar et al, 2019 [15]	United States		Persons that use ASL as a primary language, age of 18 years or above, and presence of bilateral hearing loss	Secondary Analysis of National health trends Survey in ASL	Between 2016 and 2018	

XSL•FO
RenderX

Study	Country	Aims	Study population and sample size	Design	Duration of the intervention	Main findings that related to the use of VRI[a] within health care
		This study aimed to investigate the national trends of deaf patients' satisfaction with the quality of VRI in health settings and recommend actions to improve VR quality and deaf patients' satisfaction with VRI in health care settings.				• The study shows that almost half of the people reached by the survey did not have access to VRI over the last 12 months. • It also shows that those who have access were largely dissatisfied with the quality of the service. About 41% (n=228) of the deaf patient sample rated the quality of VRI as satisfactory. The rest (n=327, 59%) rated their VRI experience as unsatisfactory. • VRI tends to be cost-effective and its flexibility is of great advantage to service providers, users, and interpreters. • The study claims that if D/deaf ASL health care users are provided with a fully functioning VRI system with qualified interpreters, this system can potentially reduce the number of emergency visits and unnecessary diagnostic tests, all of which are associated with cost burden.
Yabe, 2020 [17]	United States	This study identifies health care providers' and DHH[e] patients' interpreting preferences for VRI and in-person interpretation during critical care and noncritical care	1. Health care providers who had used VRI in clinical settings in the past 10 years were 18 years or older and spoke English. 2. DHH patients who had used VRI in clinical settings in the past 10 years were 18 years or older and used ASL	Mixed methods design incorporating both an online survey and qualitative interviews	No information	• This study provides the views of both health workers and sign language users—the findings pointed out that VRI is the preferred way of communication of patients and health providers for noncritical care. • VRI offers preparedness unattainable with in-person interpretation. Furthermore, in-person interpretation is limited in its availability and represents at times economic loss. • It outlines technical limitations regarding VRI and recommendation for its use. • It points out that patient's acceptance of VRI was linked to time constraints and type of care. Thus, acceptance was limited as it was described as waste of money as it did not prove effective for communication. • For providers, its convenience and flexibility were very important.
Kushalnagar et al, 2017 [24]	United States		Deaf adults (ages 18-90 years and above) who use ASL	Qualitative studies (cognitive interviews)	N/A[g]	• This article outlines the protocol of cultural adaptation national survey items exploring VRI. • Linguistic adaptation of items related to time, explanation of illness and use of diagrams, captions and videos is very useful for validation studies using sign language.

XSL•FO
RenderX

Study	Country	Aims	Study population and sample size	Design	Duration of the intervention	Main findings that related to the use of VRI[a] within health care
		The objectives of this study are (1) to culturally adapt and linguistically translate the HINTS[f] items to ASL (HINTS-ASL) and (2) to gather information about deaf people's health information–seeking behaviors across technology-mediated platforms.				
Singleton et al, 2019 [25]	United States	This study explored technology use among older deaf adults with regard to attitudes, adoption style, and frequency of use for a wide range of technologies, including ATs[h] for persons with hearing loss and general everyday technologies.	Participants had to be 50 years of age or older and self-identify as DHH	Online or paper copy questionnaire	__[i]	• Older adults are moving away from TTYj and TDDk to embrace VPsl and VRSm; 51% of respondents use VRI. They noticed that consumer service and support such as free delivery and personnel to set technology up had a very positive impact on the consumer experience. • Participants reported difficulty keeping up with software updates and other technology maintenance activities that require a higher level of computer literacy. Thus, many older adults in the deaf community appear to be comfortable with daily technologies and ATs and especially video-based internet technologies that support communication accessibility such as VP and VRS.
Kasales et al, 2020 [26]	United States	The goal of this review is to help members of the breast center team better understand (1) the mandates of the ADA[n] and the challenges faced by patients with select communication disabilities.	Descriptive review	Literature review (descriptive review)	N/A	• This article reviews some relevant literature and points out recommendations to use VRI. However, it does not include any empirical evidence. • They recommend using VRI when an in-person interpreter is not available and only in agreement with the patient. • It lays out the recommendation of the National Association of the Deaf Seniors of America for the use of VRI for ASL communication.
Meulder and Haualand, 2019 [27]	Norway	To critically assess the impact and role of SLIS[o] in those countries where SLIS have been institutionalized	VRI deaf users	Literature review (conceptual analysis)	N/A	

XSL•FO
RenderX

Study	Country	Aims	Study population and sample size	Design	Duration of the intervention	Main findings that related to the use of VRI[a] within health care
						• This article presents an analysis of the role that sign language interpretation has in social services including health care. • The paper makes a strong argument for the importance of language-concordant services. • It does refer broadly to sign language interpretation including VRI. It highlights that access and communication in the health care setting are mainly conceptualized and arranged with a hearing person's perspective. Little has been done to allow health settings or personnel to be bilingual and therefore more accommodating to the sign language users, cultural gaps, discriminatory set up, and other issues might not be apparent to the interpreter and shall be considered.

XSL•FO
RenderX

Study	Country	Aims	Study population and sample size	Design	Duration of the intervention	Main findings that related to the use of VRI[a] within health care
Preusse et al, 2016 [28]	United States	The goal of this study was to identify the range of challenges in everyday activities that might be experienced by older adults aging with preexisting impairments in vision, hearing, or mobility.	Interviews with subject matter experts working with older deaf adults	Qualitative study (interviews)	—	• Findings of the study revealed challenges faced by deaf persons as they age. These challenges include access to social services, adequate housing, and technology. • The findings state that access to interpreters is an issue in most health settings. Experts interviewed pointed out that this shortage of qualified sign language interpreters can be overcome by using VRI. Thus, they also pointed out that VRI may be inappropriate when people are dealing with high levels of stress such as a medical emergency. In these cases, in-person interpretation may be more appropriate, if available. • The findings show that device maintenance and software updates are difficult for this population. The study recommends one-to-one training for uptake of new technologies, as well as mixed available technologies such as haptic devices as medication reminders.
McKee et al, 2015 [29]	United States	The aim of this paper is to summarize evidence and good practices on how to enable better communication between DHH and health personnel, particularly physicians.	—	Literature review	N/A	• This paper offers an overview of good practices and questions regarding health service provision for DHH patients. • It lays out that DHH patients are more likely to experience poverty and less likely to access ICTp including smartphones. • VRI is mentioned as a tool to overcome communication barriers and improve satisfaction, quality of care, and health outcomes. However, it also mentioned that evidence on the impact of interpretation and VRI is lacking. These recommendations assume that interpretation availability either via VRI or in person is an efficient way forward.
Kwok et al, 2021 [30]	Canada	This report documents the experience in using web-based technology in an emergency department to meet communication needs of our patients who have LEP[q] including deaf sign language users during the COVID-19 pandemic.	—	Quality improvement report	March 30 and May 31, 2020	

XSL•FO
RenderX

Study	Country	Aims	Study population and sample size	Design	Duration of the intervention	Main findings that related to the use of VRI[a] within health care
						• This study focuses on the use of VRI more generally for patients of linguistic minorities including sign language. • It reports on the cost-efficiency of the intervention, laying out prices of VRI inclusive of sign language and claiming that such a cost is not problematic to absorb by the hospital. • VRI technical issues were easily overcome and personnel became acquainted to its use relatively easily. Furthermore, the study claims that the use of VRI also complies with security protocols in place during the COVID-19 pandemic and allows the protection of interpreters and others from exposure. • The authors of the paper judged that this intervention was successful for both hearing patients and DHH patients. Thus, there is no evidence that it was the case.

[a]VRI: video remote interpretation.

[b]D/deaf: “Deaf” refers to the linguistic minority while “deaf” refers to persons with hearing impairment.

[c]ASL: American Sign Language.

[d]HOH: hard of hearing.

[e]DHH: deaf or hard of hearing.

[f]HINTS: Health Information National Trends Survey.

[g]N/A: not applicable.

[h]AT: assistive technologies.

[i]Data not available.

[j]TTY: (teletypewriter) is a communication device used by people who are deaf, hard-of-hearing, or have severe speech impairment.

[k]TDD: test-driven development.

[l]VPs: videophones.

[m]VRS: video relay service.

[n]ADA: Americans with Disabilities Act.

[o]SLIS: Nottinghamshire Sign Language Interpreting Service.

[p]ICT: information and communications technology.

[q]LEP: limited English proficiency.

XSL•FO
RenderX

Textbox 1. Summary of advantages and disadvantages.

Advantages

- Convince
- Preparedness unattainable with in-person interpretation
- Access to qualified interpreters
- Possibility to work remotely for interpreters
- Safety, limiting social contact in health care environment
- Cost
- Flexibility

Disadvantages

- Technology malfunctions
- Inaccessible to deaf patients in certain physical positions and those with vision impairment
- Requires higher level of computer literacy
- Not user-friendly
- For some it might limit patient-provider relationship
- Relays on the availability of reliable internet access and adequate devices

Advantages of Using VRI Interpretation

Early literature [1,20,21] described sign language interpretation using VRI in health care settings as equally efficient as in-person interpretation. Advantages attributed to the technology, such as flexibility and affordability, encourage the idea that this technology could help overcome the shortage of qualified sign language interpretation in health care settings. It also pointed out that the use of VRI could help to override the use of inadequate techniques such as lipreading and note-reading, which are often used in health consultations with DHH patients. DHH sign language users prefer to use VRI over these techniques primarily because it allows them to communicate in their preferred language, sign language [22,26]. Lipreading and note-reading often assume that sign language users are proficient in reading and writing in a spoken language, which is often not the case. The literacy rates of DHH communities are at a sixth grade reading level or lower [29,31,32].

Articles exploring technology preferences highlight that sign language users (both children and adults) prefer VRI to other communication technology over texting devices (sign language, text, and speech interpretation linked by a call center or voice recognition technology) [16,23]. As the proliferation of VRI technology increases, consumer choices increase. With this technology, deaf patients have the possibility to choose communication tools and assistance that they deem more appropriate for their medical consultation [15-17,27]. For some noncritical medical services, VRI is preferred over in-person interpretation [16,17,24].

Sign language interpreters saw a significant advantage to this technology as it allowed them to eliminate time for transportation, given that most of their time assisting in a medical consultation is consumed by traveling to the location [22]. Saving in traveling time often translates to saving in the total cost of the interpretation. This is a key advantage often mentioned in the literature and an underlying motivation to continue expanding the use of VRI in health care settings [1,15,17,20-22,25,28,30]. VRI has also proven advantageous during the COVID-19 pandemic, allowing qualified interpreters to be available at emergency services while protecting both parties from risking potential exposure at the emergency room and complying with access restrictions [30].

The current literature suggests that the use of VRI to use qualified sign language interpreters, despite where they are located, has the potential to help overcome the scarcity of sign language interpretation and enable better communication between deaf patients and health care personnel. The advantages offered by VRI are likely to be enhanced as technology devices such as tablets, laptop computers, and smartphones become more affordable and reliable internet bandwidth becomes more available [15,17,25].

Challenges and Limitations of the Technology

As evidence grows, we are learning more about VRI technology because of its shortcomings, particularly with regard to the specificities of health care settings. A national survey conducted in the United States showed that only almost half of the representative sample did not have access to VRI during health care appointments over the last 12 months [15]. It is not clear whether the technology was needed but not available, suggesting that even in a high-income context, the availability of this technology remains limited or if participants chose not to use VRI because they had access to in-person interpretation or preferred to use other communication techniques.

Several articles in the hospital context in the United States showed that VRI was not user-friendly and led to frustration for both DHH individuals and their professional health care providers. The most common barriers noted were technical problems and poor quality of sign language interpreting services [16]. Although VRI is preferred for noncritical care, it is

XSL•FO
RenderX

considered inappropriate for critical care or stressful situations [16,25]. A second article reiterated the issues found in the first study and laid out other technical issues, such as limited placement and positioning of devices, negatively impacting the experience of using this technology [17]. They also found that VRI was seen as inconducive to enriching patient-health provider relationships and that providing VRI without previously notifying, seeking, and obtaining the agreement of the patient first was bothersome [17].

A national survey from the United States also looked at preferences of the DHH population between VRI or in-person interpretation within health care settings and found that 59% of their respondents rated their VRI experience as unsatisfactory and preferred in-person interpretation. Sign language interpreters have also reported concerns regarding technology. According to interpreters' views, in-person interpretation is more efficient at identifying when users do not understand a diagnosis, medical instructions, or other information compared with VRI. Interpreters also pointed out that the extra time before and after the appointment is useful for reviewing information available in the lobby and preparing for consultation, which enables them to provide better interpretation services [22]. The VRI does not allow interpreters to prepare or debrief DHH patients before and after consultations [31]. In turn, VRI could be more prone to incomplete communication between DHH and hearing health personnel. Capacity building among health personnel was noted as a significant communication barrier for DHH patients but also as a hindrance to technology development [20].

The efficiency of this technology is partially determined by the appropriateness of the video device used. The recommended screen of a minimum of 49.5 cm (19.5 in) is often not available [26]. Keeping up with software updates among other technologies, maintenance was considered burdensome among older DHH adults [25]. Other reported limitations of the technology included constraints due to the physical position of the patient. VRI is not accessible for patients undergoing clinical examination that requires them to be face down; VRI is also not accessible for DHH persons who are blind or have low vision [16]. The use of electronic means of communication for health information also raises security and privacy concerns. We found no information on whether the video feeds were encrypted.

The literature also shows methodological shortcomings of using health research instruments, such as surveys that explore VRI on DHH individuals, which have been developed and tested only with hearing participants. Given the cultural and linguistic differences between DHH and hearing populations, some concepts, questions, and wording may be inappropriate or incomprehensive for DHH individuals [24].

Adding to the technical and methodological issues, a more troubling challenge was assuming that an efficient VRI technology would be sufficient to overcome barriers to health care for DHH individuals (or communities). Research has shown that the use of VRI services alone is not fully accessible to DHH communities. Little research has been conducted to promote bilingualism or language-concordant practices across health settings or personnel and accessibility in broader health-related communication practices [27]. Furthermore, there is a risk that the VRI could be conceptualized and put in place from a hearing person's perspective. This limited, 1-sided view ignores issues related to cultural differences and discrepancies, discriminatory practices, intrinsic bias, and intersectionality issues related to hearing status, ethnicity, race, or multiple disabilities.

Discussion

Principal Findings

This scoping review provides an overview of the current evidence on the efficiency of the use of VRI with deaf users within health care settings. It shows that this area is under research, and the evidence is scant. It is particularly concerning that all articles found were from high-income countries, given that most DHH people live in LMICs. There is a dearth of evidence on the use of VRI and its efficiency and potential across LMICs. This reflects the long-lasting absence of voices of persons with disabilities from non-Western nations on both disability scholarship and technology innovation [33-35]. The lack of knowledge regarding the needs and realities of DHH individuals in LMICs extends beyond VRI technology. Technological progress has often overlooked the experience of disability and the everyday needs and constraints of DHH persons from the Global South. Nearly all research on assistive technology and ICT accessibility for DHH individuals and for persons with disabilities, whether from the legal, technical, or development fields, has focused on high-income countries and very little to no attention has been paid to LMICs [36]. Technological progress has often overlooked the experience of disability and the everyday needs and constraints of persons with disabilities from the Global South, among other reasons, because it is perceived as nonprofitable [34]. Failing to address this gap will cause persons with disabilities in LMICs to continue to be left behind in relation to universal health coverage.

At present, 164 countries are signatories to the Convention on the Rights of Persons with Disabilities (CRPD). CRPD Article 25 on health and Article 9 on accessibility provided the legal basis for ensuring the right to the highest attainable standard. Thus, the implementation of the CRPD remains limited, particularly in LMICs. The dominant presence of the literature from the United States may be linked to the Americans with Disabilities Act of 1990 [37], which lays the legal grounds for accessibility and nondiscrimination, as well as for the adoption of reasonable accommodation. However, similar legal frameworks have been adopted in other high-income countries with sufficient infrastructure to provide VRI services, such as the Disability Discrimination Act 2005 [38] in the United Kingdom, and we did not observe the same level of engagement on behalf of either public health or disability scholars. Nevertheless, the implementation of such CRPD rights to health and accessibility in health care settings will require robust evidence regarding the priorities, needs, and constraints of persons with disabilities in LMICs.

A major strength of this review is the use of a comprehensive search in 3 languages in a rapidly expanding technology and a focus on highlighting available evidence and gaps. A key issue highlighted by the available literature is that the availability of

XSL•FO
RenderX

VRI technology has the *potential* to address communication barriers within the health care setting, in addition to other available services and tools aside from, inter alia, in-person interpretation, telephone typewriters, and telecommunications relay services. The views, needs, and rights of the DHH community should be at the core of the development of these technologies. However, the VRI is not a quick fix to overcome accessibility issues [15,27,39]. It is important that its expansion and convenience do not undermine the possibility for DHH communities to choose whichever means of communication they prefer or which is more appropriate for the type of care they seek.

This review also pointed out a knowledge gap regarding the quality of interpretation and training in sign language interpretation for health care. It is not clear if poor-quality interpretation is a recurring issue when using in-person interpretation or if it is only an issue when using VRI [15-17]. There are no data on whether in-person interpretation, as requested in advance, the assigned interpreter is likely to use time before the consultation to undergo a prescreening for interpretation competencies, allowing better preparation for their job. Perhaps interpreters are better matched at the time of assigning the task; thus, we do not know whether this could improve the quality of interpretation. Nor do we know if such prescreening for qualification takes place for VRI interpreters or if such practice would lead to better outcomes and positive experiences across DHH users. There is a gap in the evidence on this issue, although most articles mentioned the pertinence of training for sign language interpreters on health interpretation for better communication outcomes.

The challenges documented in the literature highlight recurring technical issues regarding internet reliability, availability, and adequacy of devices in hospital settings. Although the internet is growing globally [40], it is clear that internet reliability has imposed utmost complex infrastructural challenges that could hamper VRI development in LMICs. The literature is not clear on whether, when VRI is used, users use their own devices or if they have to personally purchase internet data (and devices). This raises questions and concerns, as persons with disabilities are more likely to experience poverty in both high- and low-income countries. The financial challenges of DHH communities will have an impact on access to devices and the internet, and in turn, these challenges will impose further barriers to communication and health care. This is perhaps more acute in the Global South.

For future research, there is a need to raise awareness and build capabilities across health systems to improve accessibility for DHH individuals. The literature suggests that having more bilingual health workers, language-concordant services, better technologies, and raising awareness will contribute to better communication between DHH communities and health personnel [41-47]. New developments include technologies such as intelligent personal assistants such as Alexa, which can be used with sign language to improve communication [48]. Thus, we need to learn more about how to make health systems more accessible to DHH individuals. Accessible communication in health settings has been linked to fewer hospital visits, better treatment adherence, more cancer screening, and better oral health [10,12,14,41,42,49,50].

Comparison With Previous Literature

There have been no similar publications in this area. This study provides a well-needed analysis regarding knowledge gaps and the need for future research on the efficiency of VRI technology for sign language users in the health care context.

Limitations

Our study has a few limitations. We looked at articles examining VRI in health care settings, including hospitals, preventive care, and community health. Few rigorous articles have studied VRI for sign language users in the health care context. The protocols used and examined regarding the use of VRI for sign language are not generalizable at a national level or international level. We attempted to map and assess the available scientific literature.

Conclusions

The available literature shows that VRI may enable deaf users to overcome interpretation barriers and can potentially improve communication outcomes between them and health personnel within health care services. Communication between DHH health care users and personnel shall improve if sign language users are provided with a VRI system supported by devices with large screens and a reliable internet connection, as well as qualified interpreters trained on medical interpretation. Perhaps issues regarding lack of preparation for interpreters could be overcome by providing VRI interpreters with a brief summary of the purpose of the visit, as well as the background of the consultation before the discussion. Such preparation may allow interpreters and users to develop a rapport during health visits, and research is needed in this area.

Furthermore, our understanding of the use of VRI technology is pertinent and relevant. All articles mentioned that sign language interpretation is a scarce resource within health care systems, even in high-income countries. Thus, learning more about the possibilities and limitations of VRI is even more urgent in LMICs, because the dearth of data and in-person interpretation are largely unavailable and perhaps unfeasible in the near future in resource-constrained contexts.

Acknowledgments

The first author's (MRV) manuscript was supported by the Swiss National Science Foundation under award Ambizione (186035). The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. The authors thank Prof Antoine Geissbuhler and Prof Samia Hurst for their comments on the manuscript. They also thank Muriel Leclerc for her help in setting up the search strategies.

XSL•FO
RenderX

Authors' Contributions

MRV developed the review question. MRV was the first reviewer, and CJ was the second reviewer. MRV and CJ conducted the study and the analysis. MRV drafted the manuscript. MRV, CJ, and JC reviewed and edited the manuscript. MRV is the senior author and acts as guarantor. All authors have reviewed and revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 28 KB - jmir_v24i6e32439_app1.docx]

References

1. Masland MC, Lou C, Snowden L. Use of communication technologies to cost-effectively increase the availability of interpretation services in healthcare settings. Telemed J E Health 2010;16(6):739-745 [FREE Full text] [doi: 10.1089/tmj.2009.0186] [Medline: 20626299]
2. Zulu T, Heap M, Sinanovic E. The cost and utilisation patterns of a pilot sign language interpreter service for primary health care services in South Africa. PLoS One 2017 Dec 22;12(12):e0189983 [FREE Full text] [doi: 10.1371/journal.pone.0189983] [Medline: 29272272]
3. Kritzinger J, Schneider M, Swartz L, Braathen SH. "I just answer 'yes' to everything they say": access to health care for deaf people in Worcester, South Africa and the politics of exclusion. Patient Educ Couns 2014 Mar;94(3):379-383. [doi: 10.1016/j.pec.2013.12.006] [Medline: 24388666]
4. Bachman SS, Vedrani M, Drainoni ML, Tobias C, Andrew J. Variations in provider capacity to offer accessible health care for people with disabilities. J Soc Work Disabil Rehabil 2007;6(3):47-63. [doi: 10.1300/J198v06n03_03] [Medline: 17989034]
5. Leeson L, Saeed JI. Irish Sign Language: A Cognitive Linguistic Approach. Edinburgh, UK: Edinburgh University Press; 2012.
6. Kuenburg A, Fellinger P, Fellinger J. Health care access among deaf people. J Deaf Stud Deaf Educ 2016 Jan;21(1):1-10. [doi: 10.1093/deafed/env042] [Medline: 26405210]
7. Napier J, Skinner R, Turner G. “It’s good for them but not so for me”: inside the sign language interpreting call centre. Transl Interpreting 2017 Jul 24;9(2):1-23. [doi: 10.12807/ti.109202.2017.a01]
8. Martín MC, Phelan M. Interpreters and cultural mediators – different but complementary roles. Translocations 2010;6(1):1-19 [FREE Full text]
9. Barnett S, McKee M, Smith SR, Pearson TA. Deaf sign language users, health inequities, and public health: opportunity for social justice. Prev Chronic Dis 2011 Mar;8(2):A45 [FREE Full text] [Medline: 21324259]
10. Berman BA, Jo A, Cumberland WG, Booth H, Britt J, Stern C, et al. Breast cancer knowledge and practices among D/deaf women. Disabil Health J 2013 Oct;6(4):303-316 [FREE Full text] [doi: 10.1016/j.dhjo.2013.05.001] [Medline: 24060253]
11. Emond A, Ridd M, Sutherland H, Allsop L, Alexander A, Kyle J. The current health of the signing deaf community in the UK compared with the general population: a cross-sectional study. BMJ Open 2015 Jan 25;5(1):e006668 [FREE Full text] [doi: 10.1136/bmjopen-2014-006668] [Medline: 25619200]
12. Kushalnagar P, Engelman A, Sadler G. Deaf patient-provider communication and lung cancer screening: Health Information National Trends survey in American Sign Language (HINTS-ASL). Patient Educ Couns 2018 Jul;101(7):1232-1239 [FREE Full text] [doi: 10.1016/j.pec.2018.03.003] [Medline: 29548598]
13. Kushalnagar P, Hill C, Carrizales S, Sadler GR. Prostate-Specimen Antigen (PSA) screening and shared decision making among deaf and hearing male patients. J Cancer Educ 2020 Feb;35(1):28-35 [FREE Full text] [doi: 10.1007/s13187-018-1436-3] [Medline: 30353474]
14. McKee MM, Barnett SL, Block RC, Pearson TA. Impact of communication on preventive services among deaf American Sign Language users. Am J Prev Med 2011 Jul;41(1):75-79 [FREE Full text] [doi: 10.1016/j.amepre.2011.03.004] [Medline: 21665066]
15. Kushalnagar P, Paludneviciene R, Kushalnagar R. Video remote interpreting technology in health care: cross-sectional study of deaf patients' experiences. JMIR Rehabil Assist Technol 2019 Mar 11;6(1):e13233 [FREE Full text] [doi: 10.2196/13233] [Medline: 30855233]
16. Myers MJ, Annis IE, Withers J, Williamson L, Thomas KC. Access to effective communication aids and services among American sign language users across North Carolina: disparities and strategies to address them. Health Commun 2022 Jul;37(8):962-971. [doi: 10.1080/10410236.2021.1878594] [Medline: 33541113]
17. Yabe M. Healthcare providers' and deaf patients' interpreting preferences for critical care and non-critical care: video remote interpreting. Disabil Health J 2020 Apr;13(2):100870. [doi: 10.1016/j.dhjo.2019.100870] [Medline: 31791822]

XSL•FO
RenderX

18. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
19. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
20. Berry JA, Stewart AJ. Communicating with the deaf during the health examination visit. J Nurs Pract 2006 Sep 1;2(8):509-515. [doi: 10.1016/j.nurpra.2006.06.013]
21. Steinberg AG, Barnett S, Meador HE, Wiggins EA, Zazove P. Health care system accessibility. Experiences and perceptions of deaf people. J Gen Intern Med 2006 Mar;21(3):260-266 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00340.x] [Medline: 16499543]
22. Hommes RE, Borash AI, Hartwig K, DeGracia D. American sign language interpreters perceptions of barriers to healthcare communication in deaf and hard of hearing patients. J Community Health 2018 Oct;43(5):956-961. [doi: 10.1007/s10900-018-0511-3] [Medline: 29696596]
23. Dammeyer J, Lehane C, Marschark M. Use of technological aids and interpretation services among children and adults with hearing loss. Int J Audiol 2017 Oct;56(10):740-748. [doi: 10.1080/14992027.2017.1325970] [Medline: 28509597]
24. Kushalnagar P, Harris R, Paludneviciene R, Hoglind T. Health Information National Trends Survey in American Sign Language (HINTS-ASL): protocol for the cultural adaptation and linguistic validation of a national survey. JMIR Res Protoc 2017 Sep 13;6(9):e172 [FREE Full text] [doi: 10.2196/resprot.8067] [Medline: 28903891]
25. Singleton JL, Remillard ET, Mitzner TL, Rogers WA. Everyday technology use among older deaf adults. Disabil Rehabil Assist Technol 2019 May;14(4):325-332. [doi: 10.1080/17483107.2018.1447609] [Medline: 29522377]
26. Kasales CJ, Alkebsi ZA, Tong NT, Stephens AW. Caring for the deaf, hard-of-hearing, blind, and low-vision patients in the breast center. J Breast Imaging 2020;2(6):598-602. [doi: 10.1093/jbi/wbaa069]
27. De Meulder M, Haualand H. Sign language interpreting services: a quick fix for inclusion? Transl Interpreting Stud 2019 Sep 6;16(1):19-40. [doi: 10.1075/tis.18008.dem]
28. Preusse KC, Gonzalez ET, Singleton JL, Mitzner TL, Rogers WA. Understanding the needs of individuals ageing with impairment. Int J Human Factors Ergon 2016;4(2):144-168.
29. McKee MM, Moreland C, Atcherson SR, Zazove P. Hearing loss: communicating with the patient who is deaf or hard of hearing. FP Essent 2015 Jul;434:24-28. [Medline: 26161525]
30. Kwok MM, Chan RK, Hansen C, Thibault K, Wong HY. BMJ Open Qual 2021 Feb;10(1):e001062 [FREE Full text] [doi: 10.1136/bmjoq-2020-001062] [Medline: 33547156]
31. Napier J, Kidd MR. English literacy as a barrier to health care information for deaf people who use Auslan. Aust Fam Physician 2013 Dec;42(12):896-899 [FREE Full text] [Medline: 24324995]
32. Henning MA, Krägeloh CU, Sameshima S, Shepherd D, Shepherd G, Billington R. Access to New Zealand Sign Language interpreters and quality of life for the deaf: a pilot study. Disabil Rehabil 2011;33(25-26):2559-2566. [doi: 10.3109/09638288.2011.579225] [Medline: 21591984]
33. Rivas Velarde M. Indigenous perspectives of disability. Disabil Stud Q 2018 Dec 21;38(4):6114. [doi: 10.18061/dsq.v38i4.6114]
34. Chavarria MA, Schönenberger K, Mugeere A, Hurst S, Rivas Velarde M. Design approaches for creating person-centered, context sensitive, and sustainable assistive technology with the global south. In: Stein MA, Lazar J, editors. Accessible Technology and the Developing World. Oxford, UK: Oxford University Press; 2021.
35. Rivas Velarde MC. Addressing double layers of discrimination as barriers to health care: indigenous peoples with disabilities. AbOrig 2017 Oct 1;1(2):269-278. [doi: 10.5325/aboriginal.1.2.0269]
36. Stein AM, Lazar J. Accessible Technology and the Developing World. Oxford, UK: Oxford University Press; 2021.
37. Americans With Disabilities Act of 1990. Public Law 101-336 – 108th Congress. 1990. URL: https://www.govinfo.gov/content/pkg/STATUTE-104/pdf/STATUTE-104-Pg327.pdf [accessed 2022-03-03]
38. Disability Discrimination Act 2005. Act of the Parliament of the United Kingdom. 2005. URL: https://www.legislation.gov.uk/ukpga/2005/13/pdfs/ukpga_20050013_en.pdf [accessed 2005-12-31]
39. Kushalnagar RS, Tart JA. A survey on video relay service application interface preferences. J Technol Persons Disabil 2016;4:223-233.
40. Gould M, Montenegro V. 2016 CRPD ICT Accessibility Progress Report — A global analysis of the progress made by states parties to the convention on the rights of persons with disabilities to implement its dispositions on the accessibility of information and communication technologies and assistive technologies. G3ict. 2017. URL: https://g3ict.org/publication/2016-crpd-ict-accessibility-progress-report [accessed 2021-02-18]
41. The Lancet. The health of deaf people: communication breakdown. Lancet 2012 Mar 17;379(9820):977. [doi: 10.1016/S0140-6736(12)60411-5] [Medline: 22423869]
42. Laplante-Lévesque A, Hickson L, Worrall L. What makes adults with hearing impairment take up hearing AIDS or communication programs and achieve successful outcomes? Ear Hear 2012;33(1):79-93. [doi: 10.1097/AUD.0b013e31822c26dc] [Medline: 21841487]

XSL•FO
RenderX

43. McKee M, Schlehofer D, Cuculick J, Starr M, Smith S, Chin NP. Perceptions of cardiovascular health in an underserved community of deaf adults using American Sign Language. Disabil Health J 2011 Jul;4(3):192-197 [FREE Full text] [doi: 10.1016/j.dhjo.2011.04.001] [Medline: 21723526]
44. Thew D, Smith SR, Chang C, Starr M. The deaf strong hospital program: a model of diversity and inclusion training for first-year medical students. Acad Med 2012 Nov;87(11):1496-1500 [FREE Full text] [doi: 10.1097/ACM.0b013e31826d322d] [Medline: 23018327]
45. Hatakeyama T, Watanabe T, Takahashi K, Doi K, Fukuda A. Development of communication assistive technology for persons with deaf-blindness and physical limitation. Stud Health Technol Inform 2015;217:974-979. [Medline: 26294595]
46. Laubreton J, Morvan R, Roblot P. [Hospital consultations for deaf people]. Presse Med 2013 Nov;42(11):1427-1429. [doi: 10.1016/j.lpm.2013.05.002] [Medline: 23809425]
47. Barnett DD, Koul R, Coppola NM. Satisfaction with health care among people with hearing impairment: a survey of Medicare beneficiaries. Disabil Rehabil 2014;36(1):39-48. [doi: 10.3109/09638288.2013.777803] [Medline: 23594058]
48. Shahin N, Watfa M. Deaf and hard of hearing in the United Arab Emirates interacting with Alexa, an intelligent personal assistant. Technol Disabil 2020 Nov 20;32(4):255-269. [doi: 10.3233/tad-200286]
49. Cumberbatch K, Jones T. Use of Jamaican Sign language in the provision of dental health care. Community Dent Health 2017 Jun;34(2):72-76. [doi: 10.1922/CDH_3913Cumberbatch05] [Medline: 28573834]
50. Bown S, Aldersson R, Dekesel K. Supporting patients who are deaf who use a signed language in general practice. Br J Gen Pract 2020 Jan;70(690):10-11 [FREE Full text] [doi: 10.3399/bjgp20X707285] [Medline: 31879290]

Abbreviations

CRPD: Convention on the Rights of Persons with Disabilities
DHH: deaf or hard of hearing
FENASCOL: National Deaf Federation of Colombia
LMIC: low- and middle-income country
VRI: video remote interpretation

Edited by R Kukafka; submitted 28.07.21; peer-reviewed by N Hu, R Paludneviciene; comments to author 16.02.22; revised version received 21.03.22; accepted 20.04.22; published 09.06.22.

Please cite as:
Rivas Velarde M, Jagoe C, Cuculick J
Video Relay Interpretation and Overcoming Barriers in Health Care for Deaf Users: Scoping Review
J Med Internet Res 2022;24(6):e32439
URL: https://www.jmir.org/2022/6/e32439
doi:10.2196/32439
PMID:35679099

©Minerva Rivas Velarde, Caroline Jagoe, Jessica Cuculick. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.06.2022. 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

Internet-Delivered Cognitive Behavioral Therapy in Patients With Irritable Bowel Syndrome: Systematic Review and Meta-Analysis

Hyunjung Kim[1], PhD; Younjae Oh[1], PhD; Sun Ju Chang[2], PhD

[1]School of Nursing, Research Institute of Nursing Science, Hallym University, Chuncheon, Republic of Korea

[2]College of Nursing, Research Institute of Nursing Science, Seoul National University, Seoul, Republic of Korea

Corresponding Author:
Sun Ju Chang, PhD
College of Nursing
Research Institute of Nursing Science
Seoul National University
103 Daehak-ro
Jongno-gu
Seoul, 03080
Republic of Korea
Phone: 82 10 5226 8262
Email: changsj@snu.ac.kr

Abstract

Background: Irritable bowel syndrome is a common functional gastrointestinal disorder that negatively affects all aspects of life. With the widespread use of the internet, internet-delivered cognitive behavioral therapy has been developed and applied to control symptoms and improve the quality of life of those with irritable bowel syndrome. However, few studies have systematically reviewed the effectiveness of internet-delivered cognitive behavioral therapy on irritable bowel syndrome.

Objective: This study aimed to systematically review studies that examined the use of internet-delivered cognitive behavioral therapy in patients with irritable bowel syndrome and to evaluate the effects of internet-delivered cognitive behavioral therapy on the improvement of symptom severity, quality of life, psychological status, and cost-effectiveness.

Methods: This meta-analysis involved the search of 6 databases for relevant publications. From the 1224 publications identified through database searches, 9 randomized controlled trials were finally included in the analysis.

Results: The internet-delivered cognitive behavioral therapies including exposure-based cognitive behavioral therapy, cognitive behavioral therapy for self-management, and cognitive behavioral therapy for stress management were provided in 5 to 13 sessions for 5 to 10 weeks. Internet-delivered cognitive behavioral therapy had medium-to-large effects on symptom severity (standardized mean difference [SMD] –0.633; 95% CI –0.861 to –0.4304), quality of life (SMD 0.582; 95% CI 0.396-0.769), and cost-effectiveness (–0.372; 95% CI –0.704 to –0.039) at postintervention. The effects on symptom severity remained over time even after the intervention, short-term follow-up (SMD –0.391; 95% CI –0.560 to –0.221), and long-term follow-up (SMD –0.357; 95% CI –0.541 to –0.172). There was no significant difference in psychological status, including anxiety and depression, in those with irritable bowel syndrome compared to the controls during the postintervention period.

Conclusions: This review demonstrates that internet-delivered cognitive behavioral therapy could be a cost-effective intervention for improving symptoms and the quality of life in patients with irritable bowel syndrome. However, studies are still insufficient regarding the use of internet-delivered cognitive behavioral therapy in these patients; therefore, more high-quality studies are required in the future.

(J Med Internet Res 2022;24(6):e35260) doi:10.2196/35260

KEYWORDS

cognitive behavioral therapy; irritable bowel syndrome; internet; symptom; quality of life

XSL•FO
RenderX

Introduction

Irritable bowel syndrome (IBS), a common chronic gastrointestinal disorder, has a high prevalence of 5% to 20% worldwide [1]. Most patients with IBS experience intestinal symptoms, such as bloating, cramps, diarrhea, and constipation, in addition to abdominal pain and discomfort, for an average of 8.1 days per month [2]. Psychological symptoms include depression, anxiety, sensitivity, anger, and somatization. The symptoms can be so severe that up to 38% of patients consider suicide [3]. IBS is a social problem that causes absence, anxiety about unemployment, decreased work productivity, and increased medical costs, while also being a health problem that causes stress and negatively affects the quality of life (QOL) [4]. Therefore, symptom management and health promotion are essential in patients with IBS.

Although the mechanism has not been identified exactly, IBS can be explained with a biopsychosocial model in which somatization symptoms occur as psychosocial factors influencing the physiological functions of the brain-gut axis [5]. IBS treatment includes providing psychological comfort to the patient and assessing and correcting factors that stimulate bowel movement and sensation. The patients' quality of life (QOL) can be enhanced by improving their symptoms through lifestyle modification, the use of appropriate medication, and psychiatric treatment [6].

Based on a cognitive-behavioral model in which situation, thoughts, emotions, behaviors, and physiological responses interact with each other, cognitive behavioral therapy (CBT) has been considered as a treatment choice for IBS. CBT is a broad intervention that can include the following features: educational therapy for IBS; cognitive therapy to understand the relationship between thought, emotions, and IBS symptoms; and behavioral therapy, such as stress management, self-management, and self-help treatment [7]. CBT-based exposure therapy, including exposure training to symptom control by exposure to feared and avoided stimuli, has also been used for patients with IBS [8]. CBT is effective in improving the physical and psychological symptoms of IBS and the QOL [9,10]. In a meta-analysis of 18 randomized controlled trials, CBT was found to be more effective in patients with IBS than in control groups consisting of, for instance, those on waiting lists or receiving basic support [7]. With the implementation of CBT, it is expected that patients with IBS will gradually become healthier, more active, and more confident [5]. However, it is difficult for most patients to access CBT due to a shortage of trained therapists, especially in rural areas [11].

As the internet becomes popular worldwide, internet-delivered CBT (ICBT) can compensate for the treatment limitations of CBT. Whereas computerized CBT provides therapy via a computer system but without a therapist's input, ICBT adds that advantage while keeping the therapist's contact to a minimum [12]. ICBT consists of online psychoeducational material provided via the internet and therapist guidance, which can include providing feedback or encouragement via SMS text message, email, or chat rooms [13]. It has the advantages of reduced time for the therapist compared to conventional CBT and the ability for patients to access the treatment at any time and place [11]. Accordingly, ICBT has been applied to many psychiatric disorders, and a systematic review showed efficacy in 25 clinical applications, including psychiatric (eg, depression and anxiety), functional (eg, chronic pain and IBS), and eating disorders. Substantial evidence for the positive effects of ICBT on depression, panic disorder, and social phobia can be found [12]. Some randomized controlled trials (RCTs) have recently proven the effects of ICBT on patients with IBS. However, these studies have limitations due to the small sample size and heterogeneity [8,14,15]. To date, only a few papers have systematically reviewed the intervention methods and effectiveness of ICBT in this population. Therefore, this study attempts to comprehensively review and analyze the contents and effects of ICBT programs currently being tested in patients with IBS.

The objectives of this study are to systematically review the studies that examined the application of ICBT in patients with IBS and to evaluate the effects of ICBT on the improvement in symptom severity of IBS, QOL, anxiety, depression, and cost-effectiveness. This will provide comprehensive evidence regarding this topic.

Methods

Study Design

This study is a meta-analysis conducted to measure the effect size of ICBT in patients with IBS.

Literature Search

This study was conducted in accordance with the systematic literature review guidelines suggested by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) group [16]. A literature search was conducted using the popular search databases, PubMed, Cochrane Library, and PsycINFO, as well as Korean databases, Korean Studies Information Service System (KISS), Korean Medical Database (KMBASE), and Research Information Sharing Service (RISS) up to June 2020. Related studies in the reference list were searched for to find additional studies. Search key terms merged "irritable bowel syndrome" with "cognitive behavior therapy" or "cognitive therapy" or "cognitive psychotherapy." The complete search strategy is provided in Multimedia Appendix 1. The parameters set for the search were RCTs, journal articles, English or Korean language, and adults. The year of publication was not limited, so we could obtain a comprehensive overview of how ICBT was provided to patients with IBS. To prevent missing relevant publications, the general key term "CBT" (and not "ICBT") was selected as the key term, and abstracts of studies were screened for eligibility.

Study Selection

The data inclusion criteria were based on the PICO framework (Participant, Intervention, Comparator, Outcome), where the participant was defined as an adult patient with underlying IBS, intervention consisted of at least one of the elements of CBT and was delivered over the internet, the comparator was a group that did not receive ICBT, and the outcome was the measurable effects of ICBT.

XSL•FO
RenderX

The inclusion criteria were following: an RCT research design, adult participants with IBS; ICBT intervention (exposure-based ICBT, ICBT for self-management, ICBT for stress management), controls (patients on a waiting list who receive intervention after the treatment group, consisting of standard care, psychological treatment, or usual medical treatment), and measurable outcomes (IBS symptom severity, QOL, anxiety, depression, cost-effectiveness, visceral sensitivity, cognitive function, disability, stress, relief).

The exclusion criteria were the following: a non-RCT or secondary data analysis, studies in which ICBT was provided to both the experimental and control groups, studies with an objective other than assessing the effects of ICBT, and studies that presented insufficient data to measure the effect size.

First, duplicates were removed from the list of publications found via database searches. The titles and abstracts of publications were screened, and then the full-text studies were reviewed for eligibility. If the full text was not available, it was requested from the author. If the abstract was insufficient to determine whether the paper met the inclusion criteria, the full text was also searched for and screened. According to the inclusion and exclusion criteria, 2 researchers (HK and YO) reviewed and selected the studies separately. In the case of disagreement between them, a third researcher's (SC) opinion was to be consulted; however, the study selection results were consistent among the researchers.

Data Collection and Quality Assessment

Two researchers (HK and YO) independently collected the data from the selected papers using a data extraction form. The form was used to obtain data on the author, year, country, sample characteristics (sample size, mean age), intervention (type, duration, length of follow-up), control category (waiting list, standard care, other psychological therapy), primary and secondary outcome variables, intention to treat (ITT), and results. The primary outcome was the effect of ICBT on IBS symptom severity, which was evaluated using the following: the IBS-Symptom Severity Scale (IBS-SSS) [17], the Gastrointestinal Symptom Rating Scale (GSRS)-IBS [18], and the Bowel Symptom Severity Scale (BSSS) [19]. The secondary outcomes included QOL measured with the IBS-QOL [20], mood status measured with the Hospital Anxiety and Depression Scale (HADS) [21], the State-Trait Anxiety Inventory (STAI-S) [22], the Montgomery Asberg Depression Rating Scale (MADRS) [23], and the Center for Epidemiological Studies Depression scale (CES-D) [24]. Cost-effectiveness was measured using the Trimbos/Institute of Medical Technology Assessment Cost Questionnaire for Psychiatry (TIC-P) [25].

The methodological quality of the selected studies was assessed using the 7 criteria of the Cochrane's Risk of Bias of the Cochrane Collaboration [26]. Two researchers (HK and YO) independently evaluated the risk of bias in individual papers, and if the results were inconsistent, a consensus was reached through discussion.

Data Analysis

Comprehensive Meta-Analysis version 3.3 (Biostat) was used to assess heterogeneity and publication bias and to calculate the effect size. Heterogeneity was analyzed using the Q test and I^2 test. When the significance level of the Q statistic was less than 0.05, the results were considered heterogeneous. The I^2 value means that the closer the value is to 100%, the higher the heterogeneity: 25% (small), 50% (medium), and 75% (large) [27]. In this study, a fixed effects model was used when the studies were homogenous, and a random effects model was used when the studies were heterogeneous. For analyzing the effect size in subgroups, the recommended fixed effects model was used [28].

To verify the effect size, standardized mean difference (SMD) values with 95% CIs were calculated because the outcome variables were measured with several tools. The effect size for each outcome was analyzed postintervention. The primary outcome, IBS symptom severity, was further evaluated for the effects of short-term (4 to 6 months from intervention) and long-term (12 to 24 months) follow-up. Additionally, IBS symptom severity was analyzed in subgroups to evaluate the effect size according to the type of intervention, such as self-management and exposure therapy. Cohen's d guidelines were used to interpret the effect size, where a value of 0.2 indicated a small, 0.5 a medium, and 0.8 a large effect size [29]. For the QOL, a positive effect size indicated improvement, while a negative effect size of IBS symptom severity, psychological status, and cost indicated improvement.

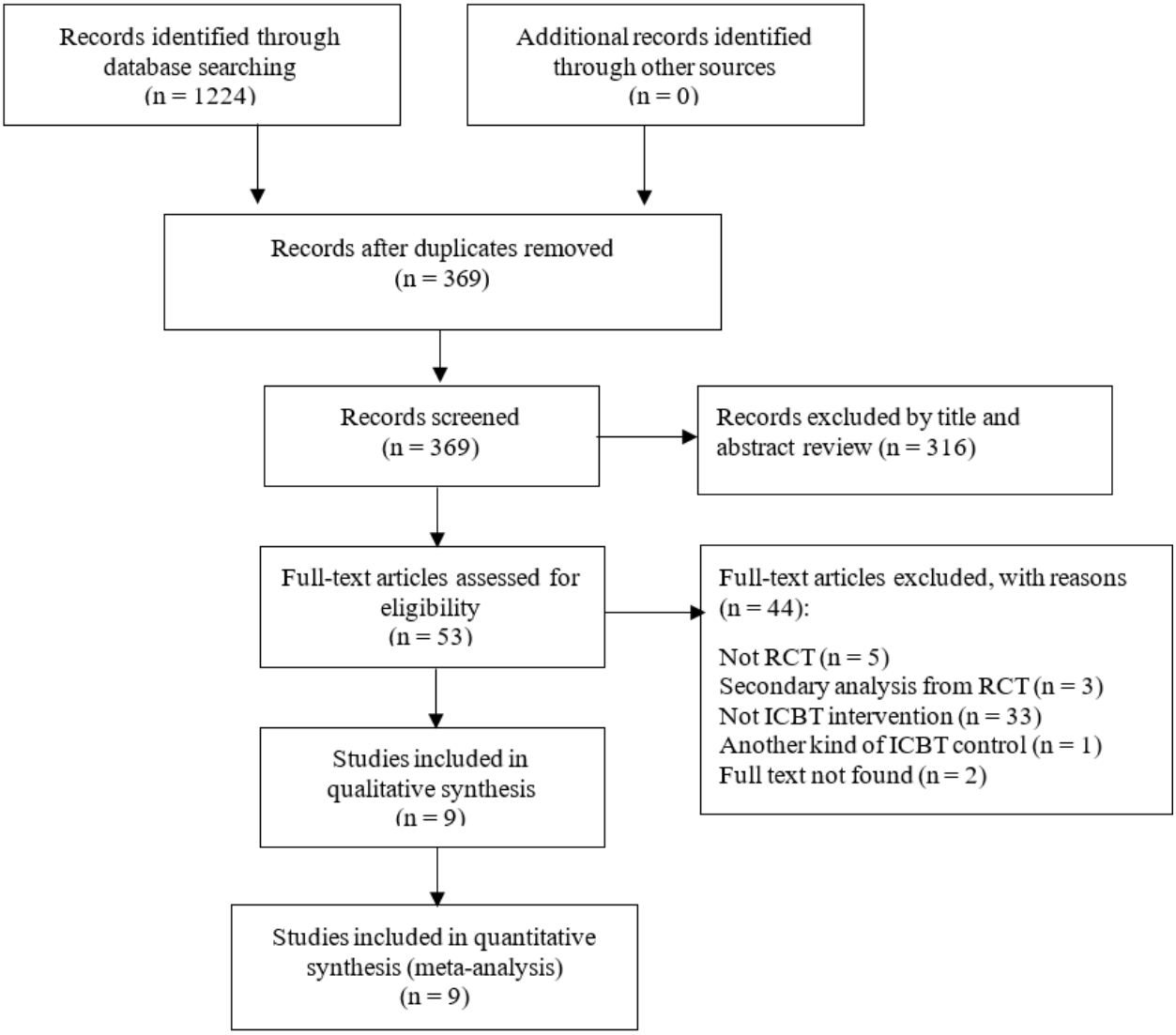
The publication bias was assessed using Egger's regression intercept: if Egger's regression intercept was not significant, publication bias was considered present. If publication bias was present, the effect size would be corrected using Duval and Tweedie's trim and fill [30].

Results

The titles and abstracts for 369 publications were screened after 855 duplicates were excluded among the 1224 publications initially identified from the search of 6 databases. Full-text screening of 53 studies was performed for eligibility, but 2 studies without the full text were eventually unable to be accessed due to no response being received from the authors. Finally, 9 studies were selected for the analysis (Figure 1).

XSL•FO
RenderX

Figure 1. Flow diagram for study inclusion. ICBT: internet-delivered cognitive behavioral therapy; RCT: randomized controlled trial.



Study Characteristics

Ultimately, 9 RCT studies were included in the analysis (Table 1). A summary of the data extraction results are presented in Multimedia Appendix 2. The studies were published between 2009 and 2019, and 6 out of the 9 studies were conducted by 2 different teams, one led by Everitt [15,31,32] and the other by Ljotsson [8,33,34]. One study was only conducted among women [14], while the rest of the studies included between 73.8% to 84.7% females, with an average age ranging from 18.5 to 44.4 years. All studies excluded participants with medical conditions that could affect the results, such as other gastrointestinal disorders (inflammatory bowel disease, celiac disease, rectal bleeding, and colorectal carcinoma) or psychiatric disorders (severe depressive symptoms, suicide ideation, psychosis, manic episodes, anorexia, and substance dependence). With the exception of 1 study [11] in which participants were included based on the self-report of a diagnosis with IBS by a medical professional, 7 studies included those diagnosed by the Rome III criteria. One study included both patients who self-reported a diagnosis by a medical professional and those who met the Rome III criteria [35]. The mean score of baseline IBS symptom severity ranged from 241 to 265 (out of 500) in 3 studies using IBS-SSS [15,31,32], 42.2 to 53.6 in 4 studies using GSRS [8,11,33,34], and 27.9 in a study using BSSS [14].

XSL•FO
RenderX

Table 1. Characteristics of included studies.

Authors	Female (%)	Age (years)	Intervention	Duration	Length of f/u[a]	Controls	ITT[b]
Andersson et al [35]	84.7	34.6	Exposure therapy (n=42)	10 w[c]/5 s[d]	10 w, 3 m[e], 12 m	Waiting list (n=43)	Yes
Everitt et al [32]	77.8	44.4	Self-management (n=45)	6 w/8 s	6 w, 12 w	Standard care (n=45)	Yes
Everitt et al [15]	76.3	42.9	Self-management (n=185)	9 w/8 s	3 m, 6 m, 12 m	Standard care (n=187)	Yes
Everitt et al [31]	N/A[f]	42.9	Self-management (n=99)	9 w/8 s	24 m	Standard care (n=105)	Yes
Hunt et al [11]	81.5	38.5	Exposure therapy (n=28)	5 w/5 s	5 w, 3 m	Waiting list (n=26)	Yes
Lee et al [14]	100	18.5	Stress management (n=48)	6 w/13 s	2 w, 6 w, 18 w	Waiting list (n=70)	N/A
Ljótsson et al [33]	84.7	34.6	Exposure therapy (n=42)	10 w/5 s	10 w, 3 m	Waiting list (n=43)	Yes
Ljótsson et al [8]	73.8	34.9	Exposure therapy (n=30)	10 w/5 s	10 w, 12 m	Waiting list (n=31)	Yes
Ljótsson et al [34]	79	38.9	Exposure therapy (n=98)	10 w/5 s	10 w, 6 m	Internet-delivered stress management (n=97)	Yes

[a]f/u: follow-up.

[b]ITT: intention to treat.

[c]w: weeks.

[d]s: sessions.

[e]m: month.

[f]N/A: not available

ICBT Program Characteristics

Among the types of CBT provided through the internet, exposure-based CBT was provided in 5 studies [8,11,33-35], CBT for self-management in 3 studies [15,31,32], and CBT for stress management in 1 study [14]. ICBT was provided as 5 to 13 sessions during a period of 5 to 10 weeks. For the control group, a waiting list was applied in 5 studies [8,11,14,33,35], standard care in 3 studies [15,31,32], and stress management techniques that did not involve CBT were applied in 1 study [34]. In all studies, a therapist contacted the patients in the ICBT group via email, telephone, or internet platform; the main contact method was email in 6 of the 9 studies (67%). The average time of therapist contact was reported in 6 studies (67%) and varied from 73 to 165 minutes in total.

After the intervention, postintervention assessments were performed, and follow-up assessments were performed at 3, 4, 6, 12, and 24 months. However, the follow-up assessments for studies with patients on a waiting list as a control group were only performed in the experimental group. In 1 study [14], ICBT was also administered to the control group (waiting list) after all follow-up assessments were completed.

For the primary outcome, 1 study [35] evaluated cost-effectiveness, while all other studies evaluated the symptom severity of IBS. In addition, QOL, anxiety, depression, visual sensitivity index, adequate relief, and cognitive function were evaluated as outcome variables. ITT data were reported in all except 1 study [14].

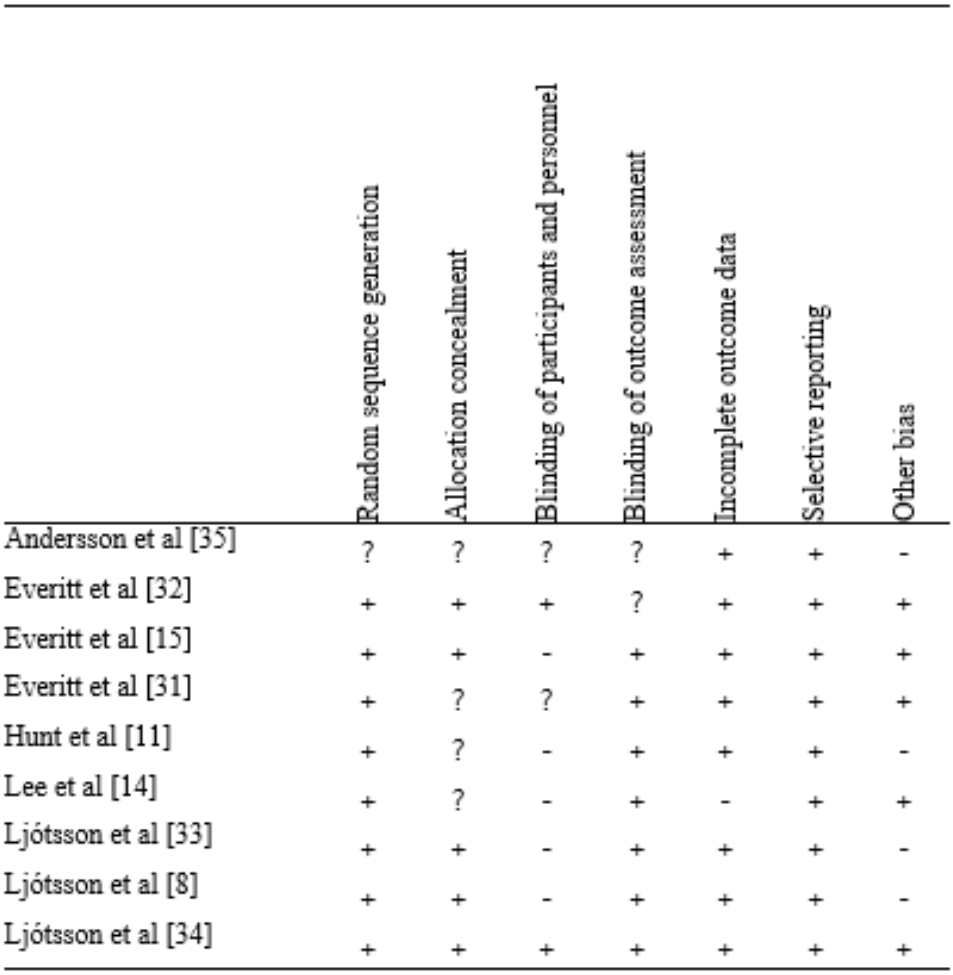
Quality Assessment

The methodological quality of the 9 included studies varied (Figure 2): 8 studies (89%) met at least 4 of the quality criteria, including 1 study [34] that met all 7 criteria. Only 1 study (11%) met 2 of the criteria [35]. All studies had a random sequence generation except for 1 study [35], 5 studies provided adequate information on allocation concealment, and only 2 studies [32,34] described the blinding of participants and personnel clearly. All studies involved the blinding of outcome assessments except for 2 studies [32,35], which did not provide sufficient information.

Regarding incomplete outcome data, all studies reported outcome data analysis completely except for 1 [14]. All studies reported all expected outcomes, including those that were prespecified to minimize bias due to selective outcome reporting. Finally, 5 studies appeared to be free of biased sources [14,15,31,32,34], whereas the other 4 studies did not report the outcomes from the waiting list control group in the follow-up stage.

XSL•FO
RenderX

Figure 2. Quality assessment of selected studies.



	Random sequence generation	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Other bias
Andersson et al [35]	?	?	?	?	+	+	-
Everitt et al [32]	+	+	+	?	+	+	+
Everitt et al [15]	+	+	-	+	+	+	+
Everitt et al [31]	+	?	?	+	+	+	+
Hunt et al [11]	+	?	-	+	+	+	-
Lee et al [14]	+	?	-	+	-	+	+
Ljótsson et al [33]	+	+	-	+	+	+	-
Ljótsson et al [8]	+	+	-	+	+	+	-
Ljótsson et al [34]	+	+	+	+	+	+	+

Effects of ICBT on Patients with IBS

Symptom Severity of IBS

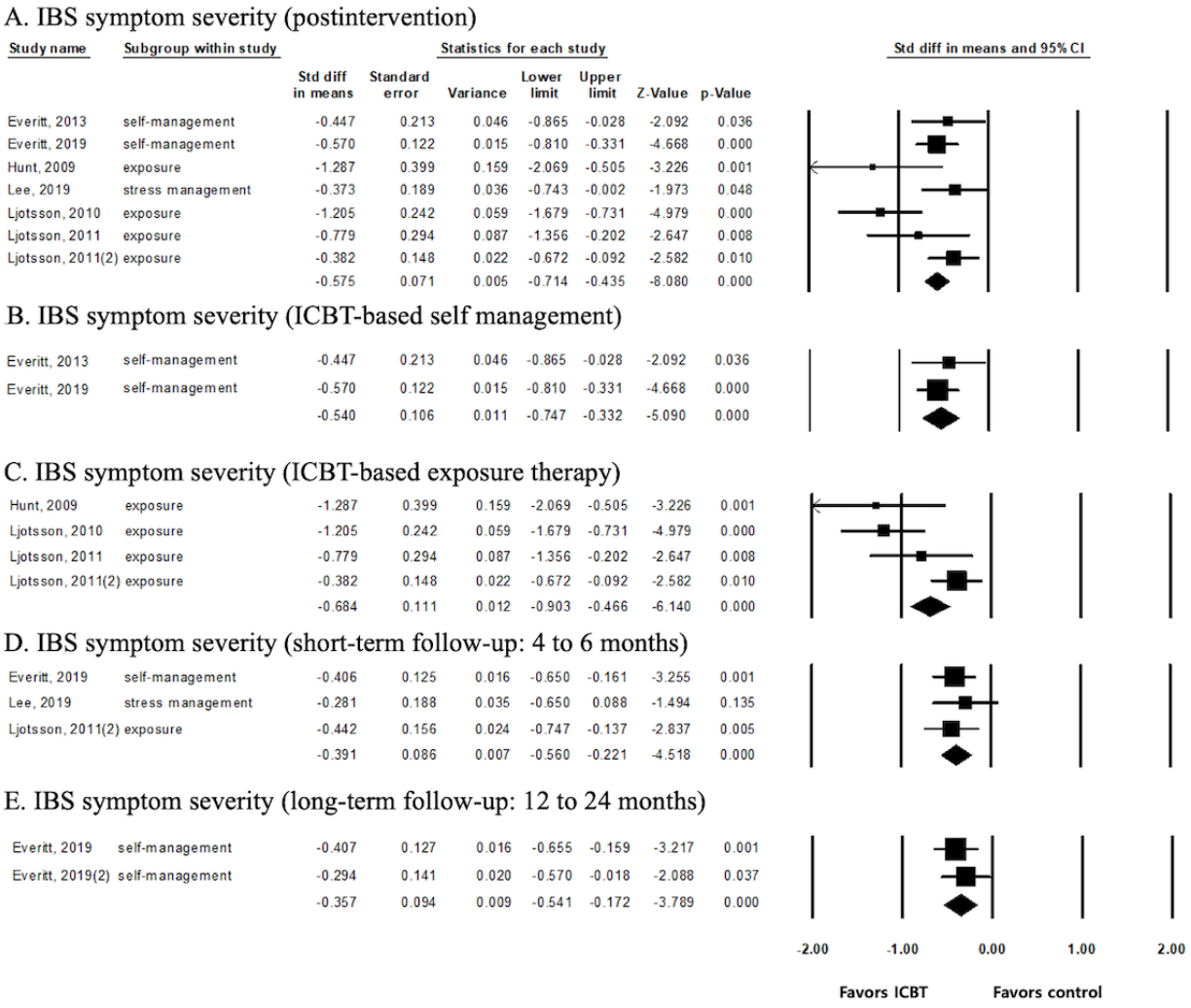
IBS symptom severity was the most reported variable as a primary outcome (7 out of 9 studies) [8,11,14,15,32-34]. Since 7 studies showed significant heterogeneity (I^2=56.01; P=.03), the overall effect on symptom severity was analyzed using a random model in postintervention. The subgroup analysis was performed using a fixed model.

Postintervention, the ICBT group had a significant reduction in IBS symptom severity compared with the control group (SMD –0.575; 95% CI –0.714 to –0.435), indicating a medium-to-large overall effect size (Figure 3A). In the subgroup analysis, we evaluated whether the effect differed according to the type of intervention. The group receiving ICBT-based self-management intervention reported significantly reduced symptom severity compared with the control group (SMD –0.540; 95% CI –0.747 to –0.332). Additionally, the group that received exposure therapy was compared with the control group, and there was a significant effect on symptom severity (SMD –0.684; 95% CI –0.903 to –0.466; Figure 3B and 3C). ICBT-based stress management was evaluated in 1 study [14], so a subgroup analysis could not be conducted.

Three short-term follow-up studies [14,15,34] had small-to-medium effect sizes in the ICBT group (SMD –0.391; 95% CI –0.560 to –0.221), and the effects remained even in the 2 long-term follow-up studies (SMD –0.357; 95% CI –0.541 to –0.172; Figure 3D and 3E) [15,31].

XSL•FO
RenderX

Figure 3. Effects size of ICBT on IBS symptom severity. IBS: irritable bowel syndrome; ICBT: internet-delivered cognitive behavioral therapy; STD: standard.

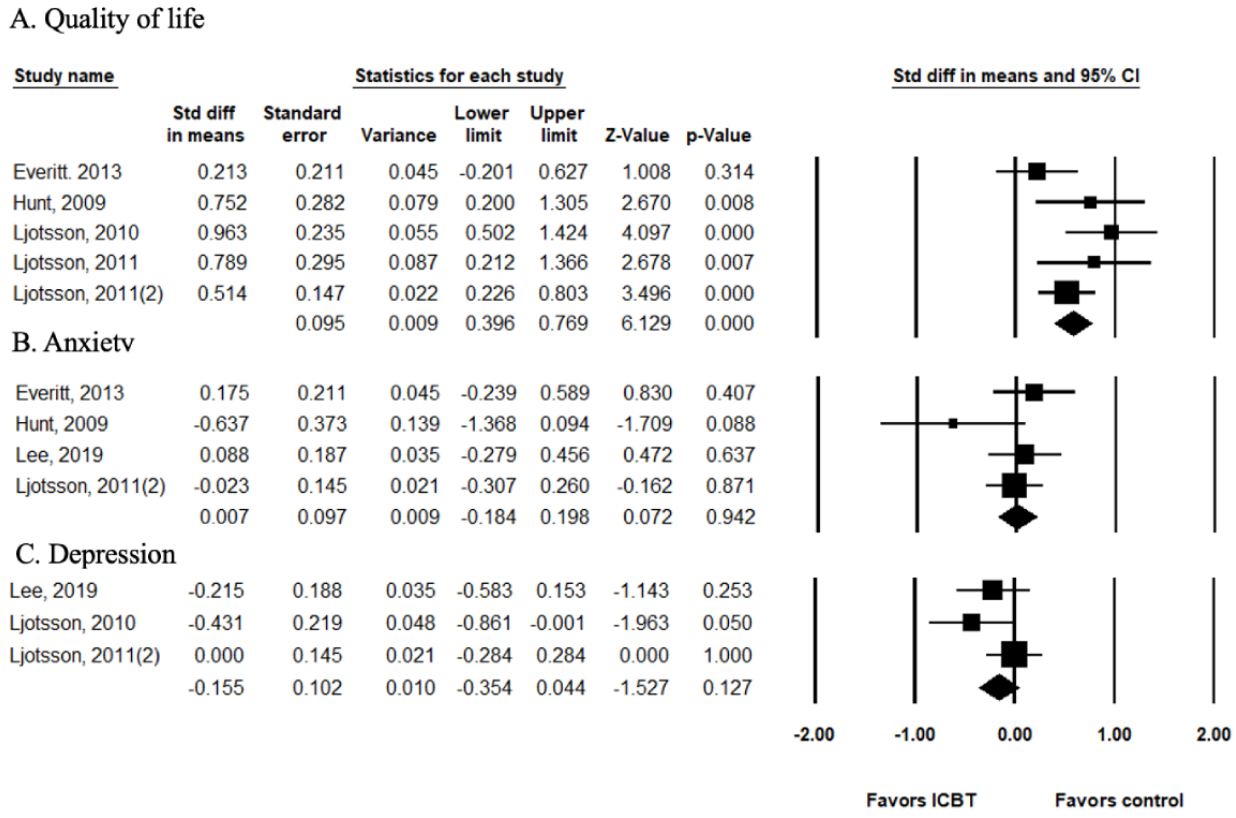


Quality of Life

Among the outcome variables, 5 studies [8,11,32-34] evaluated QOL using the same tool, the IBS-QOL developed by Patrick et al [20]. Therefore, the studies were not significantly heterogeneous (I^2=40.71; P=.15), and the effect size was analyzed using a fixed model. The effect size of ICBT on the QOL of patients with IBS was significant at 0.582 (95% CI 0.396-0.769) compared with the control group (Figure 4A).

XSL•FO
RenderX

Figure 4. Effect size of ICBT on the quality of life and psychological status. ICBT: internet-delivered cognitive behavioral therapy.



Psychological Status

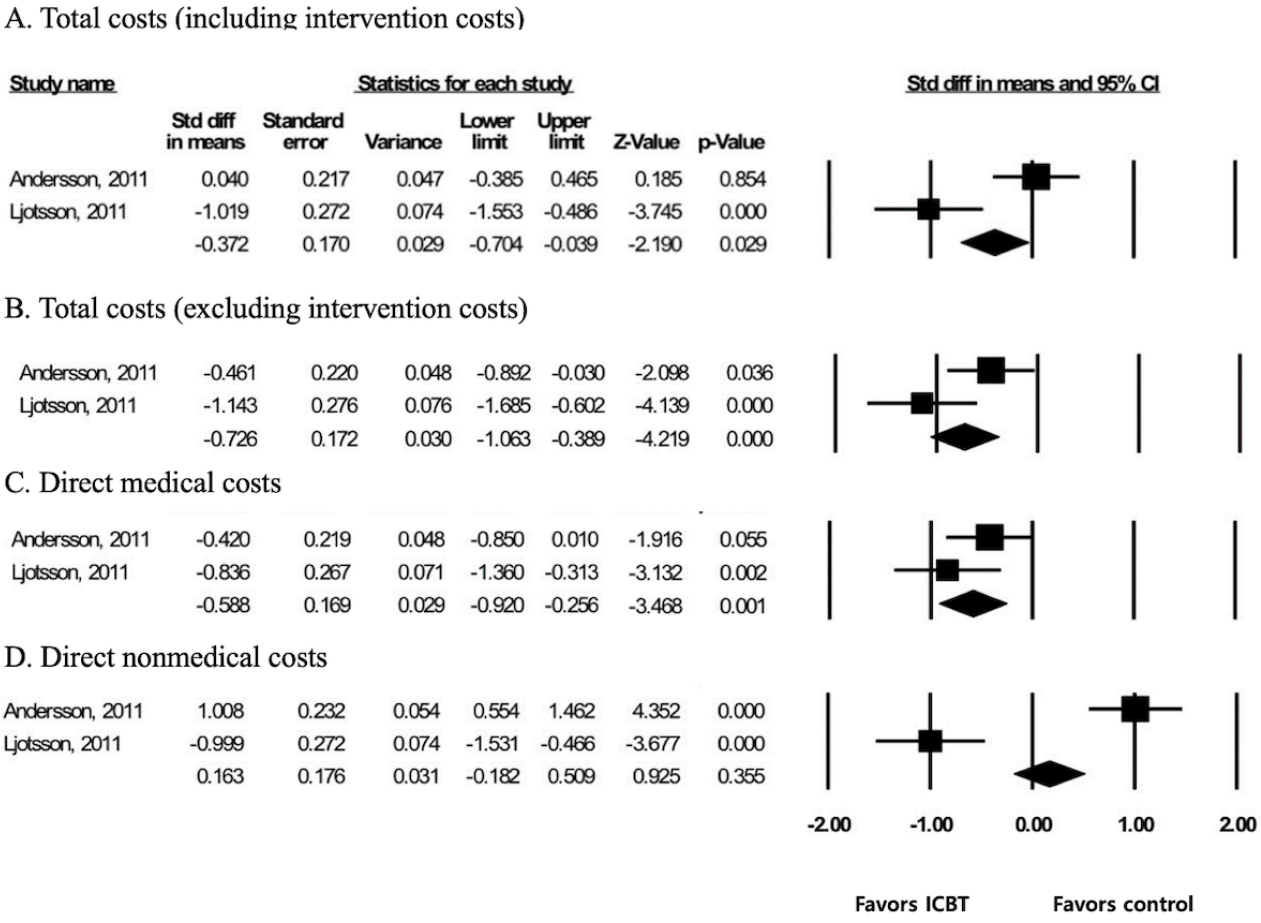
To evaluate the effects of the ICBT on psychological status, the effect sizes on depression and anxiety were analyzed (Figure 4B and 4C). Psychological status was reported in 7 studies. However, 2 studies that integrated depression and anxiety were excluded from the meta-analysis [15,31], and 1 study was excluded from the analysis for depression because it did not provide an accurate mean score for depression [32]. Therefore, depression was analyzed in 3 studies [14,33,34] and anxiety in 4 [11,14,32,34], but neither was significantly heterogeneous (depression: I^2=29.27 and P=.24; anxiety: I^2= 22.11 and P=.28), so a fixed model was adopted. There was no evidence that ICBT had any effect on depression (SMD –0.155; 95% CI –0.354 to 0.044) or anxiety (SMD 0.007; 95% CI –0.184 to 0.198).

Cost-Effectiveness

Two studies [8,35] assessed the cost-effectiveness of ICBT. When analysis was performed with fixed models, there were significant reductions in total costs including intervention costs (SMD –0.372; 95% CI –0.704 to –0.039) and in total costs excluding intervention costs (SMD –0.726; 95% CI –1.063 to –0.389). In addition, a significant effect was found on direct medical costs (SMD –0.588; 95% CI –0.920 to –0.256), but no effect was found on the reduction of direct nonmedical costs (SMD 0.163; 95% CI –0.182 to 0.509; Figure 5).

XSL•FO
RenderX

Figure 5. Effect size of ICBT on cost-effectiveness. ICBT: internet-delivered cognitive behavioral therapy.



Publication Bias

Funnel plots could not be used in this study to evaluate publication bias because these plots require at least 10 studies. Instead, bias was evaluated using Egger's regression intercept. The Egger intercepts were not significant in the analysis of outcome variables in this study, indicating that there was no risk for publication bias. Therefore, there was no need for a Duval and Tweedie's trim and fill.

Discussion

Principal Results

To our knowledge, this is the first meta-analysis to evaluate ICBT as an effective treatment for patients with IBS. Nine RCT studies were included in the analysis, and their quality was generally acceptable. As these studies were heterogeneous due to the use of different intervention methods and measurement tools, it may be difficult to definitively determine the results of this meta-analysis.

In this study, ICBT showed an overall medium-to-large effect size during the postintervention evaluation in patients with IBS. Specifically, there were significant effects on IBS symptom severity, QOL, and cost-effectiveness. However, ICBT did not have an effect on the psychological status of the treatment group compared with the waiting list or standard care controls. When stratified by the type of ICBT intervention, both exposure therapy and self-management interventions were effective compared to controls. In the follow-up studies, the effects of ICBT on the severity of IBS symptoms remained. These findings are consistent with the results of a meta-analysis in which CBT was effective in treating IBS bowel symptoms and improving the QOL of patients with IBS [7]. Although the therapist's contact is minimized in ICBT, our findings provide preliminary evidence that ICBT may be as effective as face-to-face CBT in patients with IBS.

Although only 2 RCTs among 9 studies reported the cost-effectiveness, the application of ICBT was found to improve clinical outcomes while reducing medical costs. Additional costs are required to provide ICBT, but the cost-reduction effect is maintained even after including the intervention costs. Furthermore, there was a significant effect on direct medical costs but not on nonmedical costs. Consistent with the McCrone et al [36] study, which evaluated CBT, there was no significant decrease in work days. Contrarily, in the treatment group, the improvement of IBS symptoms resulted in cost reduction compared with the control group [35]. This is consistent with our findings in the this study, in which ICBT showed significant effects on clinical outcomes.

Contrary to the results of this study, a recent CBT meta-analysis showed a significant improvement in psychological status [7]. However, in a recent review study of online psychological interventions in gastrointestinal disorders, a meta-analysis of 6 ICBT studies demonstrated no effect on stress, depression,

XSL•FO
RenderX

anxiety, or QOL in patients with IBS [37]. This discrepancy may be because psychological status is not the primary outcome of ICBT. Unlike CBT with face-to-face intervention, ICBT with minimal therapist contact might not have significant effects on psychological status. Although ICBT is effective because it is not limited by time or location, having direct contact with therapists may provide additional benefits [38]. Support from therapists could also help participants improve their motivation and adherence to therapy, which would further enhance the effectiveness of ICBT [32]. In particular, for patients who suffer from more severe symptoms, direct contact with therapists could be beneficial. In future studies, it is necessary to evaluate the extent, content, and type of contact that would improve the effectiveness of the therapy. Moreover, even though ITT analysis was conducted, the levels of attrition were high in several of the RCTs we analyzed. In particular, the attrition rate ranged from 30% to 55% in studies with long-term follow-up [15,31]. This high attrition rate might be reduced with encouragement or motivation from therapists [11]. Refractory IBS, defined as persisting symptoms in a patient even after treatment for IBS is received, requires patients to continue to manage their symptoms [31], as ensuring the long-term effects of treatment is essential. Our findings demonstrated that the effect of ICBT on IBS symptom severity persisted for a period of 12 to 24 months after the final ICBT session. This showed that ICBT is a cost-effective intervention for IBS symptoms without the need for a booster session for a long period of time. However, for other variables in this study, the effects could not be analyzed because of the small number of RCTs. Therefore, more well-designed RCTs are required to verify the long-term effects of these outcome variables. Furthermore, it is necessary to determine how long after ICBT intervention a booster session would be required to sustain the effects.

Limitations

This study has several limitations. First, there was a limited number of RCTs on ICBT for patients with IBS since research on ICBT only started recently. In particular, as 2 research teams conducted most of the ICBT trials on patients with IBS, there may be inherent biases in this meta-analysis. Our findings may be difficult to generalize to all IBS patients due to the low diversity in ethnicity and the similar characteristics of the participants. Certain limitations, such as Ljotsson's team being unable to control for the expectancy of improvement by using a waiting list as a control group and Everitt's team being unable to assess the treatment expectancy effects, indicate the importance of using an active control group. In addition, this meta-analysis may not reflect the effects of various ICBT programs or population groups. Therefore, our findings should be interpreted with caution, and further research on ICBT in different populations is needed. Second, some of the RCTs analyzed had small sample sizes, high attrition rates, and were heterogeneous, which may not substantially verify the effects of the interventions. Further research is warranted for RCTs through use of a large number of patients with IBS. Although a protocol was present for the ICBT used in each RCT, each protocol is different. Future studies should determine the effective content, frequency, and duration of ICBT.

Conclusions

In conclusion, this meta-analysis demonstrated that ICBT was superior to standard care or being on a waiting list with regard to improving IBS symptom severity, QOL, and cost-effectiveness. The effects on IBS symptom severity persisted for a long time after the intervention; that is, ICBT can be considered an effective intervention that can be provided to patients with IBS regardless of location and time. However, the number of RCTs concerning the provision of ICBT to patients with IBS is still limited, and the protocols for ICBT, including content, duration, and operators, are heterogeneous, requiring further research and standardization. Nevertheless, this meta-analysis provides the first comprehensive insight into how ICBT could be used to improve the clinical outcomes and QOL of patients with IBS while reducing treatment costs.

Acknowledgments

This work was supported by a National Research Foundation of Korea grant funded by the Korean Government (#2020R1A2C1011686).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Complete search strategy.
[DOCX File , 15 KB - jmir_v24i6e35260_app1.docx]

Multimedia Appendix 2
Summary of data extraction results.
[DOCX File , 19 KB - jmir_v24i6e35260_app2.docx]

References

1. El-Salhy M. Irritable bowel syndrome: diagnosis and pathogenesis. World J Gastroenterol 2012 Oct 07;18(37):5151-5163 [FREE Full text] [doi: 10.3748/wjg.v18.i37.5151] [Medline: 23066308]

XSL•FO
RenderX

2. Weaver K, Melkus G, Henderson W. Irritable bowel syndrome: a review. Am J Nurs 2017 Jun;117(6):48-55 [FREE Full text] [doi: 10.1097/01.NAJ.0000520253.57459.01] [Medline: 28541989]
3. Miller V, Hopkins L, Whorwell PJ. Suicidal ideation in patients with irritable bowel syndrome. Clinical Gastroenterology and Hepatology 2004 Dec;2(12):1064-1068. [doi: 10.1016/s1542-3565(04)00545-2]
4. Buono JL, Carson RT, Flores NM. Health-related quality of life, work productivity, and indirect costs among patients with irritable bowel syndrome with diarrhea. Health Qual Life Outcomes 2017 Feb 14;15(1):35 [FREE Full text] [doi: 10.1186/s12955-017-0611-2] [Medline: 28196491]
5. Jones M, Koloski N, Boyce P, Talley NJ. Pathways connecting cognitive behavioral therapy and change in bowel symptoms of IBS. J Psychosom Res 2011 Mar;70(3):278-285. [doi: 10.1016/j.jpsychores.2010.10.004] [Medline: 21334499]
6. Kim JH, Sung I. Current issues on irritable bowel syndrome: diet and irritable bowel syndrome. Korean J Gastroenterol 2014 Sep 25;64(3):142-147 [FREE Full text] [doi: 10.4166/kjg.2014.64.3.142] [Medline: 25252862]
7. Li L, Xiong L, Zhang S, Yu Q, Chen M. Cognitive-behavioral therapy for irritable bowel syndrome: a meta-analysis. J Psychosom Res 2014 Jul;77(1):1-12. [doi: 10.1016/j.jpsychores.2014.03.006] [Medline: 24913335]
8. Ljótsson B, Andersson G, Andersson E, Hedman E, Lindfors P, Andréewitch S, et al. Acceptability, effectiveness, and cost-effectiveness of internet-based exposure treatment for irritable bowel syndrome in a clinical sample: a randomized controlled trial. BMC Gastroenterol 2011 Oct 12;11(1):1-13. [doi: 10.1186/1471-230x-11-110]
9. Jang A, Hwang S, Kim E. The effects of cognitive behavioral therapy in female nursing students with irritable bowel syndrome: a randomized trial. Eur J Gastroenterol Hepatol 2014 Aug;26(8):918-926. [doi: 10.1097/MEG.0000000000000140] [Medline: 24999797]
10. Tang Q, Lin GY, Zhang MQ. Cognitive-behavioral therapy for the management of irritable bowel syndrome. World J Gastroenterol 2013 Dec 14;19(46):8605-8610 [FREE Full text] [doi: 10.3748/wjg.v19.i46.8605] [Medline: 24379577]
11. Hunt MG, Moshier S, Milonova M. Brief cognitive-behavioral internet therapy for irritable bowel syndrome. Behav Res Ther 2009 Sep;47(9):797-802. [doi: 10.1016/j.brat.2009.05.002] [Medline: 19570525]
12. Hedman E, Ljótsson B, Lindefors N. Cognitive behavior therapy via the Internet: a systematic review of applications, clinical efficacy and cost-effectiveness. Expert Rev Pharmacoecon Outcomes Res 2012 Dec;12(6):745-764. [doi: 10.1586/erp.12.67] [Medline: 23252357]
13. Andersson G. Using the Internet to provide cognitive behaviour therapy. Behav Res Ther 2009 Mar;47(3):175-180. [doi: 10.1016/j.brat.2009.01.010] [Medline: 19230862]
14. Lee T, Hsieh T, Sung H, Chen W. Internet-delivered cognitive behavior therapy for young Taiwanese female nursing students with irritable bowel syndrome-a cluster randomized controlled trial. Int J Environ Res Public Health 2019 Feb 27;16(5):708 [FREE Full text] [doi: 10.3390/ijerph16050708] [Medline: 30818837]
15. Everitt H, Landau S, O'Reilly G, Sibelli A, Hughes S, Windgassen S, ACTIB trial group. Assessing telephone-delivered cognitive-behavioural therapy (CBT) and web-delivered CBT versus treatment as usual in irritable bowel syndrome (ACTIB): a multicentre randomised trial. Gut 2019 Sep;68(9):1613-1623 [FREE Full text] [doi: 10.1136/gutjnl-2018-317805] [Medline: 30971419]
16. 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]
17. Francis C, Morris J, Whorwell P. The irritable bowel severity scoring system: a simple method of monitoring irritable bowel syndrome and its progress. Aliment Pharmacol Ther 1997 Apr;11(2):395-402 [FREE Full text] [doi: 10.1046/j.1365-2036.1997.142318000.x] [Medline: 9146781]
18. Wiklund IK, Fullerton S, Hawkey CJ, Jones RH, Longstreth GF, Mayer EA, et al. An irritable bowel syndrome-specific symptom questionnaire: development and validation. Scand J Gastroenterol 2003 Sep;38(9):947-954. [doi: 10.1080/00365520310004209] [Medline: 14531531]
19. Boyce P, Gilchrist J, Talley NJ, Rose D. Cognitive-behaviour therapy as a treatment for irritable bowel syndrome: a pilot study. Aust N Z J Psychiatry 2000 Apr;34(2):300-309. [doi: 10.1080/j.1440-1614.2000.00731.x] [Medline: 10789535]
20. Patrick D, Drossman DA, Frederick IO, DiCesare J, Puder KL. Quality of life in persons with irritable bowel syndrome: development and validation of a new measure. Dig Dis Sci 1998 Feb;43(2):400-411. [doi: 10.1023/a:1018831127942] [Medline: 9512138]
21. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
22. Spielberger C. Manual for the State-Trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press; 1983.
23. Svanborg P, Asberg M. A new self-rating scale for depression and anxiety states based on the Comprehensive Psychopathological Rating Scale. Acta Psychiatr Scand 1994 Jan;89(1):21-28. [doi: 10.1111/j.1600-0447.1994.tb01480.x] [Medline: 8140903]
24. Radloff LS. The CES-D Scale: a self-report depression scale for research in the general population. Applied Psychological Measurement 1977 Jun 01;1(3):385-401. [doi: 10.1177/014662167700100306]
25. Hakkaart-Van RL, van Straten A, Donker M. Timbos/iMTA Questionnaire for Coasts Associated with Psychiatric Illness (TIC-P). Rotterdam: Institute for Medical Technology Assessment, Erasmus University; 2002.

XSL•FO
RenderX

26. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
27. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 06;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
28. Kim YL, Jeong SH. Effects of nursing interventions for fall prevention in hospitalized patients: a meta-analysis. J Korean Acad Nurs 2015 Aug;45(4):469-482. [doi: 10.4040/jkan.2015.45.4.469] [Medline: 26364522]
29. Cohen J. A power primer. Psychological Bulletin 1992;112(1):155-159. [doi: 10.1037/0033-2909.112.1.155]
30. Sutton A, Duval SJ, Tweedie RL, Abrams KR, Jones DR. Empirical assessment of effect of publication bias on meta-analyses. BMJ 2000 Jun 10;320(7249):1574-1577 [FREE Full text] [doi: 10.1136/bmj.320.7249.1574] [Medline: 10845965]
31. Everitt HA, Landau S, O'Reilly G, Sibelli A, Hughes S, Windgassen S, et al. Cognitive behavioural therapy for irritable bowel syndrome: 24-month follow-up of participants in the ACTIB randomised trial. The Lancet Gastroenterology & Hepatology 2019 Nov;4(11):863-872. [doi: 10.1016/s2468-1253(19)30243-2]
32. Everitt H, Moss-Morris R, Sibelli A, Tapp L, Coleman N, Yardley L, et al. Management of irritable bowel syndrome in primary care: the results of an exploratory randomised controlled trial of mebeverine, methylcellulose, placebo and a self-management website. BMC Gastroenterol 2013 Apr 21;13:68 [FREE Full text] [doi: 10.1186/1471-230X-13-68] [Medline: 23602047]
33. Ljótsson B, Falk L, Vesterlund AW, Hedman E, Lindfors P, Rück C, et al. Internet-delivered exposure and mindfulness based therapy for irritable bowel syndrome--a randomized controlled trial. Behav Res Ther 2010 Jun;48(6):531-539. [doi: 10.1016/j.brat.2010.03.003] [Medline: 20362976]
34. Ljótsson B, Hedman E, Andersson E, Hesser H, Lindfors P, Hursti T, et al. Internet-delivered exposure-based treatment vs. stress management for irritable bowel syndrome: a randomized trial. Am J Gastroenterol 2011 Aug;106(8):1481-1491. [doi: 10.1038/ajg.2011.139] [Medline: 21537360]
35. Andersson E, Ljótsson B, Smit F, Paxling B, Hedman E, Lindefors N, et al. Cost-effectiveness of internet-based cognitive behavior therapy for irritable bowel syndrome: results from a randomized controlled trial. BMC Public Health 2011 Apr 07;11:215 [FREE Full text] [doi: 10.1186/1471-2458-11-215] [Medline: 21473754]
36. McCrone P, Knapp M, Kennedy T, Seed P, Jones R, Darnley S, et al. Cost-effectiveness of cognitive behaviour therapy in addition to mebeverine for irritable bowel syndrome. European Journal of Gastroenterology & Hepatology 2008;20(4):255-263. [doi: 10.1097/meg.0b013e3282f2519d]
37. Hanlon I, Hewitt C, Bell K, Phillips A, Mikocka-Walus A. Systematic review with meta-analysis: online psychological interventions for mental and physical health outcomes in gastrointestinal disorders including irritable bowel syndrome and inflammatory bowel disease. Aliment Pharmacol Ther 2018 Aug;48(3):244-259. [doi: 10.1111/apt.14840] [Medline: 29901820]
38. Versluis A, Verkuil B, Spinhoven P, van der Ploeg MM, Brosschot JF. Changing mental health and positive psychological well-being using ecological momentary interventions: a systematic review and meta-analysis. J Med Internet Res 2016 Jun 27;18(6):e152 [FREE Full text] [doi: 10.2196/jmir.5642] [Medline: 27349305]

Abbreviations

CBT: cognitive behavioral therapy
CES-D: Center for Epidemiological Studies Depression scale
GSRS: Gastrointestinal Symptom Rating Scale
HADS: Hospital Anxiety and Depression Scale
IBS: irritable bowel syndrome
IBS-SSS: Irritable Bowel Syndrome Symptom Severity Scale
ICBT: internet-delivered cognitive behavioral therapy
ITT: intention to treat
KISS: Korean Studies Information Service System
KMBASE: Korean Medical Database
MADRS: Montgomery Asberg Depression Rating Scale
MSIT: Ministry of Science and ICT
PICO: Participant, Intervention, Comparator, Outcome
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis
QOL: quality of life
RCT: randomized controlled trial
RISS: Research Information Sharing Service
SMD: standardize mean difference
STAI-S: State-Trait Anxiety Inventory
TIC-P: Trimbos/Institute of Medical Technology Assessment Cost Questionnaire for Psychiatry

XSL•FO
RenderX

Edited by T Leung; submitted 29.11.21; peer-reviewed by Y Hwang, A Crisafio, K Matsumoto; comments to author 31.01.22; revised version received 10.03.22; accepted 27.04.22; published 10.06.22.

Please cite as:
Kim H, Oh Y, Chang SJ
Internet-Delivered Cognitive Behavioral Therapy in Patients With Irritable Bowel Syndrome: Systematic Review and Meta-Analysis
J Med Internet Res 2022;24(6):e35260
URL: https://www.jmir.org/2022/6/e35260
doi:10.2196/35260
PMID:35687389

©Hyunjung Kim, Younjae Oh, Sun Ju Chang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.06.2022. 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

Assessing Progress Toward the Vision of a Comprehensive, Shared Electronic Care Plan: Scoping Review

Jenna M Norton[1], PhD, MPH; Alex Ip[2], MPH; Nicole Ruggiano[3], MSW, PhD; Tolulope Abidogun[4], MBBS; Djibril Souleymane Camara[5,6], MD, MPH; Helen Fu[7], MSN, PhD; Bat-Zion Hose[8], PhD; Saadia Miran[1], MS; Chun-Ju Hsiao[6], PhD; Jing Wang[9], PhD, MPH; Arlene S Bierman[6], MD, MS

[1]Division of Kidney, Urologic and Hematologic Diseases, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, United States

[2]School of Medicine and Health Sciences, George Washington University, Washington, DC, United States

[3]School of Social Work, University of Alabama, Tuscaloosa, AL, United States

[4]Office of Clinical Research Support, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, United States

[5]Public Health Informatics Fellowship Program, Center for Surveillance, Epidemiology, and Laboratory Services, Centers for Disease Control and Prevention, Atlanta, GA, United States

[6]Center for Evidence and Practice Improvement, Agency for Healthcare Research and Quality, Rockville, MD, United States

[7]Richard M Fairbanks School of Public Health, Center for Biomedical Informatics, Regenstrief Institute, Indiana University, Indianapolis, IN, United States

[8]Department of Anesthesiology and Critical Care, Perelman School of Medicine, University of Pennsylvania, Philadephia, PA, United States

[9]College of Nursing, Florida State University, Tallahassee, FL, United States

Corresponding Author:
Jenna M Norton, PhD, MPH
Division of Kidney, Urologic and Hematologic Diseases
National Institute of Diabetes and Digestive and Kidney Diseases
National Institutes of Health
6707 Democracy Blvd, 6th Floor
Bethesda, MD, 20817
United States
Phone: 1 301 928 5509
Email: jenna.norton@nih.gov

Abstract

Background: Care plans are central to effective care delivery for people with multiple chronic conditions. But existing care plans—which typically are difficult to share across care settings and care team members—poorly serve people with multiple chronic conditions, who often receive care from numerous clinicians in multiple care settings. Comprehensive, shared electronic care (e-care) plans are dynamic electronic tools that facilitate care coordination and address the totality of health and social needs across care contexts. They have emerged as a potential way to improve care for individuals with multiple chronic conditions.

Objective: To review the landscape of e-care plans and care plan–related initiatives that could allow the creation of a comprehensive, shared e-care plan and inform a joint initiative by the National Institutes of Health and the Agency for Healthcare Research and Quality to develop e-care planning tools for people with multiple chronic conditions.

Methods: We conducted a scoping review, searching literature from 2015 to June 2020 using Scopus, Clinical Key, and PubMed; we also searched the gray literature. To identify initiatives potentially missing from this search, we interviewed expert informants. Relevant data were then identified and extracted in a structured format for data synthesis and analysis using an expanded typology of care plans adapted to our study context. The extracted data included (1) the perspective of the initiatives; (2) their scope, (3) network, and (4) context; (5) their use of open syntax standards; and (6) their use of open semantic standards.

Results: We identified 7 projects for e-care plans and 3 projects for health care data standards. Each project provided critical infrastructure that could be leveraged to promote the vision of a comprehensive, shared e-care plan. All the e-care plan projects supported both broad goals and specific behaviors; 1 project supported a network of professionals across clinical, community,

XSL•FO
RenderX

and home-based networks; 4 projects included social determinants of health. Most projects specified an open syntax standard, but only 3 specified open semantic standards.

Conclusions: A comprehensive, shared, interoperable e-care plan has the potential to greatly improve the coordination of care for individuals with multiple chronic conditions across multiple care settings. The need for such a plan is heightened in the wake of the ongoing COVID-19 pandemic. While none of the existing care plan projects meet all the criteria for an optimal e-care plan, they all provide critical infrastructure that can be leveraged as we advance toward the vision of a comprehensive, shared e-care plan. However, critical gaps must be addressed in order to achieve this vision.

(J Med Internet Res 2022;24(6):e36569) doi:10.2196/36569

KEYWORDS

electronic care plan; care planning; care plan; care coordination; multiple conditions; multiple chronic conditions; chronic disease; chronic condition; electronic care; digital health; electronic tools; e-care; healthcare data; eHealth

Introduction

Multiple chronic conditions affect 1 in 3 American adults and 4 in 5 Medicare beneficiaries. It is the most common chronic condition seen in clinical practice, and while there is no standard definition or measure of multiple chronic conditions, it is generally understood to be the co-occurrence of 2 or more chronic mental or physical health conditions. Other impairments or disabilities are also sometimes included in the definition of multiple chronic conditions, as are syndromes such as frailty and social factors such as homelessness. Providing integrated person-centered care to people living with multiple chronic conditions is a major challenge [1,2]. People with multiple chronic conditions and their caregivers often experience significant burdens associated with coordinating care across multiple disease states, clinicians, and settings, including scheduling multiple medical appointments, managing complex drug and dietary regimens, and integrating multiple sources of (sometimes conflicting) medical advice [3-6]. Fragmentation of care for people living with multiple chronic conditions presents multiple challenges to clinicians and contributes to avoidable hospitalizations, duplication of services, adverse events, and higher health care costs [7]. Further, given the disproportionate prevalence of multiple chronic conditions in Black and Hispanic Americans [8], such fragmentation of care may exacerbate disparities in health outcomes.

Care plans are a central component of effective care delivery for people with multiple chronic conditions and other complex health care needs. Care plans, increasingly required by the Center for Medicare and Medicaid Services (CMS) in its programs, include written, comprehensive, patient-centered longitudinal plans of action that identify a patient's goals and health needs and the services and support required to meet them.

Existing care plans are largely paper based, and when electronic, often designed for a specific care setting or condition. They are often not interoperable and are difficult to share between providers, patients, and caregivers. People with multiple chronic conditions are more likely to have multiple care plans, which, rather than improving care coordination and integration, can instead lead to competing plans and increased fragmentation of care.

A comprehensive, shared electronic care (e-care) plan (CSeCP) that is also interoperable is a dynamic electronic tool that employs health information technology to facilitate collaboration between individuals and their clinical teams, with the goal of addressing the totality of their health and social needs across all care settings [9]. Ideally, a CSeCP would allow clinicians, patients, and caregivers to electronically view role-specific information [10]. A National Quality Forum report on care coordination recommended that an e-care plan should include the following sections, with data shared across all care settings: (1) prioritized health concerns, including social determinants of health (SDoH), (2) health and life goals, (3) interventions, and (4) health status of the individual [11]. Potential benefits of e-care plans include (1) improved quality and efficiency of care, (2) streamlined access to patient health records across the care team (including the patient), (3) coordinated medication and treatment management, and (4) improved care transitions [12-15]. E-care plans can also aid in the assessment, identification, and collection of information on SDoH for individuals and communities and inform practice and policy recommendations across health care settings [16].

The use of CSeCPs has emerged as a potential solution for improving and coordinating the care of individuals with multiple chronic conditions [17]. However, e-care plans that use different data standards cannot be easily shared across providers. Emerging standards combined with commonly used clinical terminology provide a foundation that makes the development of a comprehensive, interoperable e-care plan achievable. The Office of the National Coordinator for Health Information Technology has set a goal of nationwide interoperability by 2024 [17]. This has contributed to a rapid uptake of emerging health information technology data standards, such as the Fast Healthcare Interoperability Resources (FHIR) specification—a flexible standard for exchanging health care information electronically—and Substitutable Medical Applications, Reusable Technologies (SMART), an open, vendor-agnostic, standards-based technology platform that enables the development of applications that seamlessly and securely integrate with health information technology systems [18,19].

To advance toward a CSeCP, the Agency for Healthcare Research and Quality (AHRQ) and the National Institute for Diabetes and Digestive and Kidney Diseases (NIDDK) are collaborating to build interoperable, open-source, patient-, caregiver-, and clinician-facing e-care plan applications and a Health Level Seven (HL7) FHIR implementation guide to improve care coordination for people with multiple chronic

XSL•FO
RenderX

conditions across clinical and community settings. To inform this and other efforts in the field, we conducted a scoping review of past and current e-care plans and care plan–related initiatives, aiming to identify foundational projects and resources that could inform the multiple chronic conditions e-care plan project and other efforts in this space. This paper describes the process and results of our scoping review, as well as the functionality of existing e-care plans and the gaps that need to be addressed in order to advance toward a comprehensive e-care plan.

Methods

Using the scoping review methodology [20], we first searched Scopus, Clinical Key, and PubMed for articles featuring nonproprietary e-care plan projects; the reference list was reviewed to identify additional articles. We also searched the grey literature, used Google, and used others sources such as the US Office of the National Coordinator for Health Information Technology (ONC) Interoperability Proving Grounds [21] and Health Level Seven International [22], an organization accredited by the American National Standards Institute to develop health standards. All searches included combinations of the following terms: “interoperability,” “electronic care plan,” “care plan,” “SMART on FHIR,” “FHIR,” “C-CDA,” and “multiple chronic conditions.” Searches were limited to the years January 2015 to June 2020 to capture recent projects in a rapidly evolving field. In addition, we conducted discussions with expert informants across the federal government, academia, developer and vendor organizations, and industry (including HL7) to identify additional projects missed in the search of gray literature and published literature. Contact information for the included projects was used to identify the informants, who provided individual consultation about e-care plan development. Snowballing techniques [23] were used to add other relevant stakeholders.

Once an e-care plan project was identified, data were extracted, including (1) the implementation period, (2) project contact information, (3) the project description, (4) the population targeted, (5) fields and domains documented through the e-care plan, (6) standard technology features (eg, FHIR and HL7 Consolidated Clinical Document Architecture [C-CDA]), (7) current project activity, and (8) project results and outcomes. To determine how the identified e-care plan, including the multiple chronic conditions e-care plan project, contributes to the development of an interoperable CSeCP, we applied a recently developed typology of care plans by Burt and colleagues [24] that includes three domains: (1) perspective, indicating the degree to which the content and development of the care plan reflect a person- and patient-centered perspective rather than a professional-centered perspective, (2) scope, indicating the focus on discrete behaviors versus broad goals, with an optimal CSeCP including both, and (3) network, or the inclusion of broad care teams rather than patient-clinician dyads. We also expanded on Burt’s typology by adding three domains: (1) context, representing clinical versus SDoH data, with an optimal CSeCP including both, (2) the use of an open syntax (or format) standard (eg, C-CDA or FHIR), and (3) the use of open semantic standards (eg, clinical terminology value sets) to support interoperability. We assessed the degree to which each project met these optimal criteria for a CSeCP.

Results

Development of e-Care Plans

Table 1 shows the 7 existing nonproprietary e-care plan projects that we identified. These included (1) the Care Plan Domain Analysis Model (DAM) version 1.0, (2) the Care Plan DAM version 2.0 [9,25], (3) the Electronic Long-Term Services and Supports (eLTSS) plan [26], (4) the Pharmacist e-Care Plan (PeCP) [27], (5) the chronic kidney disease (CKD) e-care plan [28], (6) the Dynamic Care Planning (DCP) profile [29], and (7) the Omnibus Care Plan (OCP) [30,31]. Several of these care plans incorporated components of the Standards and Interoperability Framework developed by the National Quality Forum [11] and hence were useful to consider when developing a comprehensive, interoperable e-care plan for multiple chronic conditions. For example, the PeCP initiative includes prioritized health concerns, goals (ie, medication optimization), and interventions (eg, medication management) [27]. Table 1 provides an overview of the e-care plan projects. Figure 1 provides a visual description of the expected data flow for the e-care plan apps. A central FHIR server will aggregate data across multiple settings of care. SMART on FHIR e-care plan apps designed for key users (ie, patients, unpaid caregivers, and clinicians) will pull from the FHIR server to display aggregated patient data. In addition, the apps will collect novel person-centered data and share these data back to the FHIR server, where they will be available (along with comprehensive EHR data) back to clinical and research settings.

XSL•FO
RenderX

Table 1. Projects to develop e-care plans.

Organization	Project	Time frame	Description	Users/settings	Domains/features	Underlying standards	Outputs	Contributions to a CSeCP[a] and gaps
Health Level Seven	Care Plan DAM[b] 1.0	2011-2016	Provides industry with a set of comprehensive clinical requirement–driven use cases and logical information models to inform design, development, and implementation of care plan systems.	Hospitals; long-term care; home care; mental health	Health concerns (including risks/barriers); goals/preferences; intervention (care activity); outcomes	C-CDA[c]	C-CDA specification	Provides syntax for e-care plans; uses an interdisciplinary approach; allows for multiple, potentially uncoordinated disease/context-specific plans, which is not patient-centered; does not identify semantic standards or specific value sets; does not capture SDoH[d] data; document-based format limits real-time data updates
Health Level Seven	Care Plan DAM 2.0	2017-present	Uses iterative literature/use case reviews and industry engagement to provide an evidence-based and user-centered blueprint to inform a revision of the Care Plan DAM 1.0 C-CDA specification, develop a FHIR[e] care plan template, and improve related resources.	Hospitals; long-term care; home care; mental health	DAM 1.0 features plus possible additions: assessment; SDoH; protocol; order/order set (as intervention/care activity); advance directives; care coordination	C-CDA; FHIR	C-CDA specification; FHIR specification	Provides syntax structure for the e-care plan; uses an interdisciplinary approach; allows for multiple, potentially uncoordinated disease/context-specific plans, which is not patient-centered; does not identify semantic standards or specific value sets
Center for Medicare and Medicaid Services and Office of the National Coordinator for Health Information Technology	eLTSS[f] Initiative	2014-present	Working to identify and harmonize electronic standards to enable the creation, exchange, and reuse of interoperable service plans to improve the coordination of health and social services that support an individual's mental and physical health.	Long-term service providers (clinical and community); recipients of long-term care	Medicare/Medicaid beneficiary demographics; goals and strengths; person-centered planning; plan information; plan signatures; risks; service information; service provider information	C-CDA; FHIR; clinical terminology	C-CDA implementation guide; FHIR implementation guide; VSAC[g]	Provides semantic standards and value sets for inclusion in a multiple chronic condition e-care plan; provides a syntax for the exchange of data among long-term services and support providers; discipline-specific approach may limit application in the multiple chronic conditions context
Pharmacy Health Information Technology Collaborative	Pharmacist e-Care Plan	2015-present	Provides a standard for interoperable exchange of consensus-driven, prioritized, medication-related activities, plans, and goals for enhanced medication management, specified through Health Level Seven C-CDA and FHIR implementation guides.	Pharmacists; people receiving care in the community; family caregivers; pharmacies; hospitals; long-term care facilities	Patient goals; health concerns; active medication list; drug therapy problems; laboratory results; vitals; payer information; billing for services	C-CDA; C-CDA on FHIR; clinical terminology	C-CDA implementation guide; FHIR implementation guide; VSAC	Provides value sets for inclusion in a multiple chronic conditions e-care plan; provides a syntax for exchange with community-based settings; the discipline-specific approach may limit application in the multiple chronic condition context; document-based format limits real-time data updates

XSL•FO
RenderX

Organization	Project	Time frame	Description	Users/settings	Domains/features	Underlying standards	Outputs	Contributions to a CSeCP[a] and gaps
National Institute of Diabetes and Digestive and Kidney Disease	CKD[h] e-Care Plan	2016-2019	Aimed to facilitate the longitudinal transfer of key patient data among the patient, family caregivers, and the clinical care team across settings by identifying and prioritizing a comprehensive set of clinical and contextual data elements and associated data standards from widely used clinical terminologies.	People with CKD; family caregivers; diverse clinicians providing care for people with CKD; primary care; specialty practices; hospitals	Header (person and plan information); health and social concerns; patient and clinician goals; interventions; health status evaluation and outcomes	Clinical terminology	Value sets specifying more than 300 data elements	Provides value sets for inclusion in a multiple chronic conditions e-care plan; disease-specific approach is of limited use in the context of multiple chronic conditions
Integrating the Healthcare Enterprise	Dynamic Care Planning Profile	2016-present	Provides the structures and transactions for care planning, creating, dynamically updating, and sharing care plans. This profile does not define or assume a single care plan for a patient, but rather depicts how multiple care plans can be shared and used to coordinate care.	Clinicians; patients; payers	Health issues; goals; interventions; outcomes	FHIR; care plan DAM	Supplement to the Integrating the Healthcare Enterprise Patient Care Coordination Technical Framework (Standard for Trial Use 4)	Interdisciplinary approach; allows for multiple, potentially uncoordinated disease/context-specific plans, which is not patient-centered; does not identify specific value sets
SAMHSA[i]	Omnibus Care Plan	2018	Developed SMART[j] on FHIR, a browser-based (desktop or mobile), patient-centered care coordination application designed to share information with multiple care providers. It is built on existing SMART applications which determine consent, explanation of benefits, and clinical value sets, some of which are proprietary.	Clinicians	Opioid management; suicide prevention; care coordination; alerts/notifications; consent management; task/activity management; referral management; scheduling/ appointments	FHIR; SMART on FHIR	SMART on FHIR application	Provides an open-source SMART on FHIR application for use by clinicians; addresses SDoH and behavioral considerations; use of proprietary tools and applications creates a barrier to implementation and interoperability

XSL•FO
RenderX

Organization	Project	Time frame	Description	Users/settings	Domains/features	Underlying standards	Outputs	Contributions to a CSeCP[a] and gaps
Agency for Healthcare Research and Quality, National Institute of Diabetes and Digestive and Kidney Disease, and Assistant Secretary for Planning and Evaluation	Multiple chronic conditions e-care plan	2019-2023	Developing patient- and clinician-facing, interoperable e-care plan applications and a FHIR implementation guide to facilitate aggregation and sharing of critical patient-centered data across home, community, clinic, and research-based settings by extracting data from point-of-care health systems and allowing transfer of that data across settings.	People with multiple chronic conditions, including CKD, type 2 diabetes, cardiovascular disease, and chronic pain; family caregivers; diverse clinicians providing care for people with multiple chronic conditions; home and community-based providers	Person/plan information; health and social concerns; patient and clinician goals; interventions; health status evaluation and outcomes	FHIR; SMART on FHIR; clinical terminology	FHIR implementation guide; clinician SMART on FHIR app; patient mobile SMART on FHIR app	Provides syntax and semantic standards for the exchange of patient data across multiple users/settings; provides a proof-of-concept of a single comprehensive shared care plan; will require expansion to additional disease states

[a]CSeCP: comprehensive shared electronic (e-)care plan

[b]DAM: domain analysis model

[c]C-CDA: consolidated clinical document architecture

[d]SDoH: social determinants of health

[e]FHIR: Fast Healthcare Interoperability Resources

[f]eLTSS: electronic long-term services and supports

[g]VSAC: Value Set Authority Center [32]

[h]CKD: chronic kidney disease

[i]SAMHSA: Substance Abuse and Mental Health Services Administration

[j]SMART: substitutable medical applications, reusable technologies

XSL•FO
RenderX

Figure 1. Multiple chronic conditions e-care plan data flow. FHIR: fast healthcare interoperability resources; SMART: substitutable medical applications, reusable technologies; EHR: electronic health records; API: Application Programming Interface.

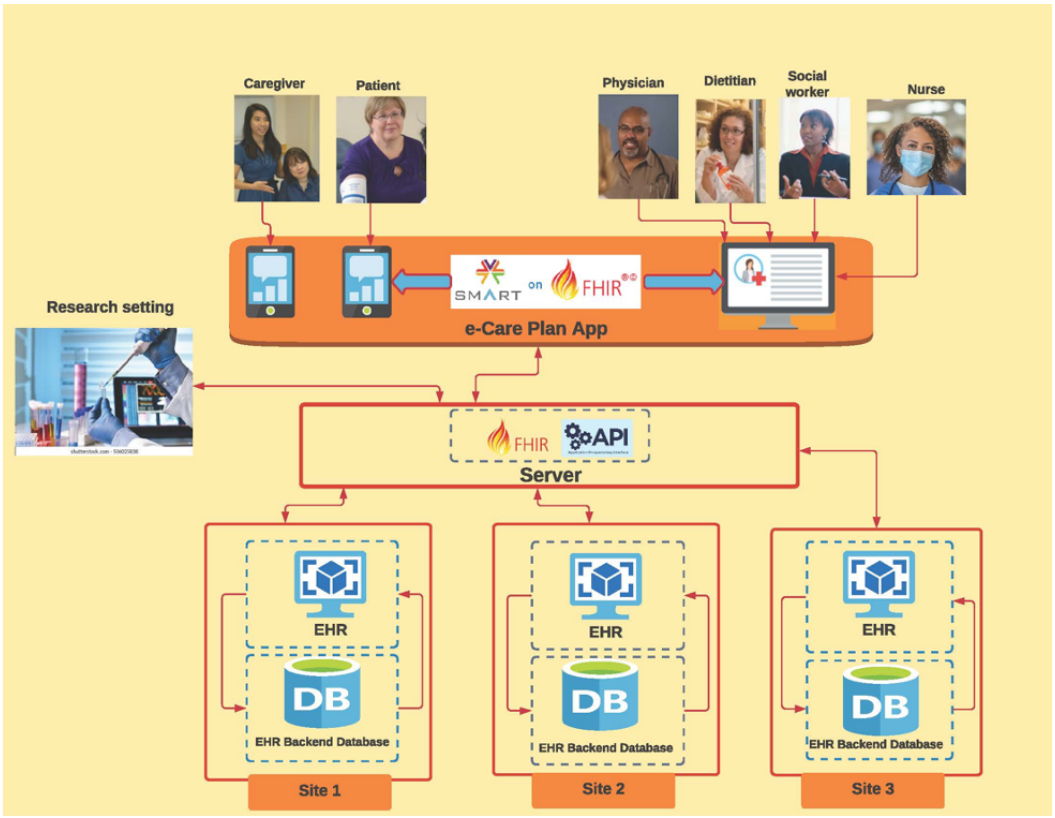


Figure 2 shows a schema of the degree to which each project aligned with the optimal CSeCP criteria. We determined that—in the context of multiple chronic conditions—4 of the 7 care plan projects reflected the perspective of the professional rather than that of the person, because they either supported only a single disease (eg, a CKD e-care plan), or they allowed for multiple, distinct and potentially uncoordinated disease or context-specific plans (DAM 1.0, DAM 2.0, and DCP). Either situation—a single disease care plan or multiple uncoordinated care plans—would not meet the needs of a person with multiple chronic conditions who must manage their conditions simultaneously in their day-to-day life, and thus does not reflect such a person's perspective. All plans supported both broad goals and specific behaviors. Only the OCP supported a network of professionals across the clinical, community, and home-based networks, while 4 projects supported the entire clinical team, 1 supported the entire LTSS team, and 1 focused primarily on pharmacist care.

Four of the care plan projects (DAM 2.0, eLTSS, PeCP, and CKD) included SDoH data. All but the CKD care plan specified either C-CDA, FHIR, or both as syntax standards, while only 3 projects (eLTSS, PeCP, and CKD) specified open-source clinical terminology value sets (eg, Logical Observation Identifiers, Names, and Codes [LOINC], Systematized Nomenclature of Medicine Clinical Terms [SNOMED-CT], International Classification of Diseases 10th Revision [ICD-10], Current Procedural Terminology [CPT], or RxNORM). For instance, the CKD e-care plan project identified data standards from common clinical terminology for more than 300 prioritized data elements, and partnered with the Regenstrief Institute to develop new LOINC codes for the data elements lacking existing data standards [28]. The OCP uses proprietary tools to identify specific condition value sets, creating a barrier for potential implementation and interoperability [30].

XSL•FO
RenderX

Figure 2. Alignment of identified care plan projects with comprehensive, shared electronic care plan criteria. Red indicates suboptimal alignment with a criterion, yellow indicates partial alignment, and green indicates optimal alignment. DAM: domain analysis model; C-CDA: consolidated clinical document architecture; CKD: chronic kidney disease; eLTSS: electronic long-term services and supports; FHIR: fast healthcare interoperability resources; SDoH: social determinants of health; SMART: substitutable medical applications, reusable technologies; PeCP: pharmacist e-care plan; DCP: dynamic care planning; OCP: omnibus care plan; CSeCP: comprehensive, shared electronic care plan.

Care Plan	Perspective	Scope	Network	SDoH data	Format standard	Terminology standards/ value sets
DAM 1.0	Professional	Behaviors and goals	Clinical team	No	C-CDA	Not specified
DAM 2.0	Professional	Behaviors and goals	Clinical team	Yes	C-CDA and FHIR	Not specified
eLTSS	Person	Behaviors and goals	eLTSS team	Yes	C-CDA and FHIR	Publicly available value sets
PeCP	Person	Behaviors and goals	Pharmacy centered	Yes	C-CDA and FHIR	Publicly available value sets
CKD	Professional	Behaviors and goals	Clinical team	Yes	None	Publicly available value sets
DCP	Professional	Behaviors and goals	Clinical team	No	FHIR	Not specified
OCP	Person	Behaviors and goals	Clinical, community & home-based teams	No	FHIR	Proprietary value sets

Development of Key Health Care Data Standards for People With Multiple Chronic Conditions

Table 2 shows the 3 projects we identified that are developing clinical terminology and coding harmonization that can be leveraged in the development of interoperable e-care plans. These projects included (1) the Data Element Library (DEL) [33], (2) the Gravity Project [34], and (3) the Post-Acute Care Interoperability (PACIO) project [31]. The DEL specifies data elements and standards for the data that the CMS requires postacute care facilities to collect as part of patient health assessments. The Gravity Project, led by the Social Interventions Research and Evaluation Network at the University of California, San Francisco, is a national collaborative to harmonize documentation of SDoH data in electronic health record (EHR) systems. The PACIO project aims to identify data standards to advance interoperable health data exchange between postacute care providers, other health care providers, patients, and key stakeholders through a consensus-based, case-driven approach. Their initial efforts have focused on data standards relating to cognitive and functional status.

XSL•FO
RenderX

Table 2. Development of key health care data standards for people with multiple chronic conditions.

Organization	Project	Time frame	Description	Intended users	Fields/domains	Standards	Outputs
Center for Medicare and Medicaid Services	Data Element Library	2018-present	Centralized resource for Center for Medicare and Medicaid Services assessment instrument data elements (eg, questions and responses) and their associated health information technology standards.	Inpatient rehabilitation facilities, home health agencies, long-term care hospitals, skilled nursing facilities, hospice care, home and community-based services	IRF[a] Patient Assessment Instrument, Outcome and Assessment Information Set; LTCH[b] Continuity Assessment Record and Evaluation Data Set; SNF[c] Minimum Data Set; Hospice Item Set; Functional Assessment Standardized Items	Clinical terminology	Standardized data elements relevant to postacute care
Social Interventions Research and Evaluation Network	Gravity Project	2019-present	Develop structured data standards to reduce barriers to documentation and exchange of social determinants of health data, including social risks and protective factors	Health care	Food insecurity, housing instability and homelessness, inadequate housing, transportation access; additional domains to be determined	FHIR[d]; clinical terminology	Social determinants of health FHIR implementation guide
Center for Medicare and Medicaid Services and The Alliance to Modernize Healthcare	Post-Acute Care Interoperability Project	2019-present	Advance interoperable health data exchange between postacute care and other providers, patients, and key stakeholders across health care.	Postacute care, long-term care hospitals, home health agencies, skilled nursing facilities, inpatient rehabilitation facilities	Cognitive status; functional status; additional domains to be determined	FHIR; clinical terminology	Cognitive status, FHIR implementation guide, functional status FHIR implementation guide

[a]IRF: Inpatient Rehabilitation Facility

[b]LTCH: Long-Term Care Hospital

[c]SNF: Skilled Nursing Facility

[d]FHIR: Fast Healthcare Interoperability Resources

Discussion

Most care plans in use today are paper-based and localized or limited to a specific discipline, disease, or care setting. An electronic, interoperable CSeCP has the potential to greatly improve the quality of care for individuals with multiple chronic conditions, who see numerous providers across multiple care settings, and overcome barriers faced by these providers to accessing and sharing person-centered health information across settings. The burden of multiple chronic conditions is increasing in the United States as its population ages, warranting a redoubled focus on care coordination and the interoperable exchange of health information for people with multiple chronic conditions. Greater interoperability across all health care settings may improve health outcomes, increase clinician workflow efficiency, decrease redundant services, minimize searching for clinical information, and reduce health care costs. This need is heightened in the wake of the ongoing COVID-19 pandemic, which has increased the use of virtual care and, given evidence of potential long-term complications among COVID 19 survivors, may result in individuals with underlying chronic conditions carrying a heavy burden of multiple chronic conditions, in addition to creating a new cohort of people with multiple chronic conditions in previously healthy populations.

Prior efforts to develop e-care plans [9,14,25,27,29,35,36] and data standards [31,33,34] have provided a solid foundation that makes the realization of a CSeCP more feasible. While none of the existing care plan projects identified by our review met all our criteria for an optimal CSeCP, each provides critical infrastructure that can be leveraged as we advance toward the vision of the CSeCP. The multiple chronic conditions e-care plan project [37] aims to build on the identified e-care plan and standards efforts to bring us closer to a CSeCP. The multiple chronic conditions e-care plan project will support the aggregation and sharing of person-centered data through identification of key data elements and clinical terminology standards, specification of an HL7 FHIR implementation guide, and development of clinician-, patient-, and caregiver-facing SMART on FHIR e-care plan applications. The multiple chronic conditions e-care plan project takes a person-centered approach, aggregating person-important health and social data—including patient-reported outcomes—across numerous chronic conditions, beginning initially with CKD, a subset of cardiovascular diseases (ischemic heart disease, hypertension, and heart failure), type 2 diabetes, and chronic pain. With these conditions as a use case, the project will provide an extensible framework for a CSeCP upon which additional disease- and condition-specific value sets and FHIR profiles can be added. To curate a holistic set of data elements for exchange, data element selection and prioritization are informed by broad stakeholder input through

XSL•FO
RenderX

technical expert panels. These technical expert panels consist of people with multiple chronic conditions, their caregivers, clinicians from diverse disciplines, community organizations, clinical informaticists, EHR vendors, and developers, among others. The project focuses not only on the core patient–primary care provider dyad but also on a wide, multidisciplinary care team network across the clinical, community, and home-based settings of care. The draft implantation guide and multiple chronic conditions e-care plan project app is being tested during multiple HL7 Connectathons and implemented and tested across real-world clinical and community-based settings of care, with the goal of balloting through HL7 as a standard for trial use in September 2022.

While we anticipate that the multiple chronic conditions e-care plan project will bring us closer to the vision of a CSeCP, much work will be necessary beyond the scope of this project. Key data elements and corresponding value sets and FHIR profiles must be identified and specified for numerous additional chronic conditions. Many data elements known to be important for care—including SDoH—are currently not supported by semantic standards or clinical terminology. While efforts to build these standards are underway [34], widespread implementation may take years. “Writing back” consolidated care plan data to individual EHRs will be necessary to achieve the full interoperability benefits of the e-care plan; however, writing back remains a widely recognized policy challenge, as many EHR systems are reluctant to write back data from external systems. While standard practices are in place for patient authorization of data exchange on a broad scale, additional work is needed to determine whether and how individuals may wish to specify data access privileges on the individual data element level—and to determine the implications this may have for individual privacy and care coordination. Such data element–level specification may be particularly important for potentially stigmatizing information (eg, sexually transmitted diseases, mental health conditions, or addiction). In addition to the semantic and syntax standards included in this scoping review, the realization of a CSeCP will require a comprehensive reference architecture outlining the structures and integrations of the various information technology products and systems, such as EHRs and health information exchanges, potentially involved in the exchange of e-care plan data. The Centers for Disease Control and Prevention’s Making EHR Data More Available for Research and Public Health (MedMorph) project [38] aims to develop a reliable, scalable, and interoperable reference architecture and demonstrated implementation to access and share EHR data across multiple public health and research scenarios. However, many home- and community-based providers have information technology systems that are distinct from the traditional health information technology infrastructure and do not have health information exchange access. Further, unaffiliated EHRs are more widely used in rural settings, creating a barrier to implementation of e-care plans in areas that are already disproportionately affected by poor health outcomes [39]. Additional work will be necessary to ensure equitable application of the CSeCP and other health information technology solutions regardless of location.

The purposes of an interoperable shared e-care plan are, first, to improve the quality and outcomes of care delivery by improving communication, coordination, and information sharing across clinical teams, patients, and caregivers. The second purpose is to provide comprehensive data on clinical conditions and management, as well as patient-reported outcomes, social factors, and patient goals and preferences in order to conduct real world research on people living with multiple chronic conditions. Clinical research on the management of different constellations of disease and health service research on the most effective models of care delivery are both needed [1]. Furthermore, since common risk factors such as smoking, physical inactivity, and unhealthy diets lead to multiple conditions, research on reducing the risk of developing multiple chronic conditions is also needed. Our study has several strengths. This is the first review to identify and assess the numerous ongoing activities in the dynamic field of care plan development. Our data collection and search strategies were broad. In addition to searching academic literature, we reviewed the gray literature, including the use of search engines and a review of government and standards development organization websites, and conducted stakeholder interviews. However, we must acknowledge certain limitations. Information gathered from websites may not be frequently updated, which could have limited our understanding of specific aspects of the sampled e-care plans and related projects. However, this may have been mitigated by our strategy of interviewing stakeholders. Our focus was limited to nonproprietary plans and e-care plan–related initiatives that have developed data standards to support interoperability. Several identified care plan and standards projects are ongoing, and thus their final outputs and success remain to be seen.

A CSeCP has the potential to greatly improve the quality of care for individuals with multiple chronic conditions who see multiple providers across multiple care settings. Prior efforts to develop e-care plans [9,14,25,27,29,35,36] and data standards [31,33,34] provide a solid foundation that makes the realization of a CSeCP feasible. The multiple chronic conditions e-Care Plan is building on these efforts to advance toward a CSeCP. However, critical gaps must be addressed in order to achieve a person-centered, interdisciplinary, and interoperable CSeCP.

Acknowledgments

The findings and conclusions in this document are those of the authors, who are responsible for its content, and do not necessarily represent the views of the Agency for Healthcare Research and Quality (AHRQ), the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), or the National Institutes of Health (NIH). No statement in this report should be construed as an official position of the AHRQ, NIDDK, NIH or the US Department of Health and Human Services.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Bierman AS, Wang J, O'Malley PG, Moss DK. Transforming care for people with multiple chronic conditions: Agency for Healthcare Research and Quality's research agenda. Health Serv Res 2021 Oct 06;56 Suppl 1(S1):973-979 [FREE Full text] [doi: 10.1111/1475-6773.13863] [Medline: 34378192]
2. Suls J, Bayliss E, Berry J, Bierman A, Chrischilles E, Farhat T, et al. Measuring Multimorbidity: Selecting the Right Instrument for the Purpose and the Data Source. Med Care 2021 Aug 01;59(8):743-756. [doi: 10.1097/MLR.0000000000001566] [Medline: 33974576]
3. Eton DT, Yost KJ, Lai J, Ridgeway JL, Egginton JS, Rosedahl JK, et al. Development and validation of the Patient Experience with Treatment and Self-management (PETS): a patient-reported measure of treatment burden. Qual Life Res 2017 Feb 26;26(2):489-503 [FREE Full text] [doi: 10.1007/s11136-016-1397-0] [Medline: 27566732]
4. Chronic Conditions among Medicare Beneficiaries. Centers for Medicare and Medicaid Services. URL: https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/chronic-conditions/downloads/2012chartbook.pdf [accessed 2021-05-20]
5. Ioakeim-Skoufa I, Poblador-Plou B, Carmona-Pírez J, Díez-Manglano J, Navickas R, Gimeno-Feliu LA, et al. Multimorbidity Patterns in the General Population: Results from the EpiChron Cohort Study. Int J Environ Res Public Health 2020 Jun 14;17(12):4242 [FREE Full text] [doi: 10.3390/ijerph17124242] [Medline: 32545876]
6. Smith SM, Wallace E, O'Dowd T, Fortin M. Interventions for improving outcomes in patients with multimorbidity in primary care and community settings. Cochrane Database Syst Rev 2021 Jan 15;1:CD006560 [FREE Full text] [doi: 10.1002/14651858.CD006560.pub4] [Medline: 33448337]
7. Hempstead K, Delia D, Cantor J, Nguyen T, Brenner J. The fragmentation of hospital use among a cohort of high utilizers: implications for emerging care coordination strategies for patients with multiple chronic conditions. Med Care 2014 Mar;52 Suppl 3:S67-S74. [doi: 10.1097/MLR.0000000000000049] [Medline: 24561761]
8. Quiñones AR, Botoseneanu A, Markwardt S, Nagel CL, Newsom JT, Dorr DA, et al. Racial/ethnic differences in multimorbidity development and chronic disease accumulation for middle-aged adults. PLoS One 2019 Jun 17;14(6):e0218462 [FREE Full text] [doi: 10.1371/journal.pone.0218462] [Medline: 31206556]
9. Care Plans 2.0: Consumer Principles for Health and Care Planning in an Electronic Environment. Consumer Partnership for eHealth. URL: https://www.nationalpartnership.org/our-work/resources/health-care/digital-health/consumer-principles-for-1.pdf [accessed 2021-05-20]
10. Baker A, Cronin K, Conway P, DeSalvo K, Rajkumar R, Press M. Making the comprehensive shared care plan a reality. NEJM catalyst. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.16.0838 [accessed 2021-05-20]
11. Critical Paths for Creating Data Platforms: Care Coordination. National Quality Forum. 2012. URL: https://www.qualityforum.org/Publications/2012/11/Critical_Paths_for_Creating_Data_Platforms__Care_Coordination.aspx [accessed 2021-05-20]
12. Ogle SM, Cooke CE, Brandt NJ. Medication Management and e-Care Planning: What are the Opportunities for the Future? J Gerontol Nurs 2015 Oct;41(10):13-17. [doi: 10.3928/00989134-20150915-02] [Medline: 26488251]
13. de Jong CC, Ros WJ, van Leeuwen M, Schrijvers G. How Professionals Share an E-Care Plan for the Elderly in Primary Care: Evaluating the Use of an E-Communication Tool by Different Combinations of Professionals. J Med Internet Res 2016 Nov 24;18(11):e304 [FREE Full text] [doi: 10.2196/jmir.6332] [Medline: 27884811]
14. Norton JM, Ketchum CJ, Narva AS, Star RA, Rodgers GP. Complementary Initiatives from the NIDDK to Advance Kidney Health. Clin J Am Soc Nephrol 2017 Sep 07;12(9):1544-1547 [FREE Full text] [doi: 10.2215/CJN.02120217] [Medline: 28716859]
15. Cullen TA, Kasthurirathne SN, Norton JM, Narva AS. Personalizing Longitudinal Care Coordination for Patients with Chronic Kidney Disease. Stud Health Technol Inform 2017;245:1354. [doi: 10.3233/978-1-61499-830-3-1354] [Medline: 29295433]
16. Chen C, Weider K, Konopka K, Danis M. Incorporation of socioeconomic status indicators into policies for the meaningful use of electronic health records. J Health Care Poor Underserved 2014 Feb;25(1):1-16 [FREE Full text] [doi: 10.1353/hpu.2014.0040] [Medline: 24509007]
17. Connecting Health and Care for the Nation: A Shared Nationwide Interoperability Roadmap. Office of the National Coordinator for Health Information Technology. URL: https://www.healthit.gov/sites/default/files/hie-interoperability/nationwide-interoperability-roadmap-final-version-1.0.pdf [accessed 2022-04-22]
18. Overview - FHIR v4.0.1. HL7.org. URL: https://www.hl7.org/fhir/overview.html [accessed 2021-05-20]
19. SMART on FHIR. Computational Health Informatics Program, Boston Children's Hospital. URL: https://docs.smarthealthit.org/ [accessed 2021-05-20]
20. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb 20;8(1):19-32. [doi: 10.1080/1364557032000119616]

XSL•FO
RenderX

21. Interoperability Proving Ground. The U.S. Office of the National Coordinator for Health Information Technology (ONC). URL: https://www.healthit.gov/techlab/ipg/ [accessed 2021-05-20]
22. Homepage | HL7 International. Health Level Seven International. URL: https://www.hl7.org/ [accessed 2021-05-20]
23. Robson C, McCartan K. Real World Research, 4th ed. West Sussex, UK: John Wiley & Sons Ltd; 2016.
24. Burt J, Rick J, Blakeman T, Protheroe J, Roland M, Bower P. Care plans and care planning in long-term conditions: a conceptual model. Prim Health Care Res Dev 2014 Oct;15(4):342-354 [FREE Full text] [doi: 10.1017/S1463423613000327] [Medline: 23883621]
25. Chu S. Care Plan Domain Analysis Model (DAM) 2020. Health Level Seven International. URL: https://confluence.hl7.org/display/PC/Care+Plan+Projects [accessed 2021-05-20]
26. Reineke N, White G. eLTSS Home: ONC Tech Lab Standards Coordination 2020. Office of the National Coordinator for Health Information Technology. URL: https://oncprojectracking.healthit.gov/wiki/display/TechLabSC/eLTSS+Home [accessed 2021-05-20]
27. Gaunt S, Gonzaja Z. Pharmacist Care Plan Document. Health Level Seven International. URL: https://confluence.hl7.org/pages/viewpage.action?pageId=31690434 [accessed 2021-05-20]
28. Development of an Electronic CKD Care Plan 2019. National Institute of Diabetes and Digestive and Kidney Diseases. URL: https://www.niddk.nih.gov/research-funding/research-programs/kidney-clinical-research-epidemiology/health-information-technology/development-electronic-ckd-care-plan?dkrd=hisce0104#draftSet [accessed 2021-05-20]
29. Dynamic Care Planning 2020. Integrating the Healthcare Enterprise. URL: https://wiki.ihe.net/index.php/Dynamic_Care_Planning [accessed 2021-05-20]
30. Salyards K. Omnibus Care Plan - OCP. GitHub. URL: https://github.com/petercyli/omnibus-care-plan [accessed 2021-05-20]
31. About PACIO Project. PACIO Project. URL: http://pacioproject.org/ [accessed 2021-05-20]
32. Value Set Authority Center. National Library of Medicine. URL: https://vsac.nlm.nih.gov [accessed 2022-05-13]
33. CMS Data Element Library (DEL). Centers for Medicare & Medicaid Services. URL: https://del.cms.gov/DELWeb/pubHome [accessed 2021-05-20]
34. Gravity Project 2020. Health Level Seven International. URL: https://www.hl7.org/gravity/ [accessed 2021-05-20]
35. Development of an Electronic CKD Care Plan. National Institute of Diabetes and Digestive and Kidney Disease. URL: https://www.niddk.nih.gov/research-funding/research-programs/kidney-clinical-research-epidemiology/health-information-technology/development-electronic-ckd-care-plan?dkrd=hisce0104 [accessed 2021-05-20]
36. Project Summary for Electronic Long-Term Services and Supports (eLTSS) Service Plan. Health Level Seven International. URL: http://www.hl7.org/special/Committees/projman/searchableProjectIndex.cfm?action=edit&ProjectNumber=1431 [accessed 2021-05-20]
37. Chu S, Shell K. Multiple Chronic Conditions (MCC) eCare Plan 2020. Health Level Seven International. URL: https://confluence.hl7.org/display/PC/Multiple+Chronic+Conditions+%28MCC%29+eCare+Plan [accessed 2021-05-20]
38. Making EHR Data More Available for Research and Public Health. Centers for Disease Control and Prevention. URL: https://aspe.hhs.gov/making-electronic-health-record-ehr-data-more-available-research-public-health [accessed 2021-05-20]
39. Gill E, Dykes PC, Rudin RS, Storm M, McGrath K, Bates DW. Technology-facilitated care coordination in rural areas: What is needed? Int J Med Inform 2020 May;137:104102 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104102] [Medline: 32179256]

Abbreviations

AHRQ: Agency for Healthcare Research and Quality
C-CDA: Consolidated Clinical Document Architecture
CKD: chronic kidney disease
CMS: Center for Medicare and Medicaid Services
CSeCP: comprehensive shared electronic (e-)care plan
DAM: Domain Analysis Model
DCP: Dynamic Care Planning
DEL: Data Element Library
EHR: electronic health record
eLTSS: Electronic Long-Term Services and Supports
FHIR: Fast Healthcare Interoperability Resources
HL7: Health Level Seven
NIDDK: National Institute for Diabetes and Digestive and Kidney Diseases
OCP: Omnibus Care Plan
ONC: US Office of the National Coordinator for Health Information Technology
PACIO: Post-Acute Care Interoperability
PeCP: Pharmacist e-Care Plan
SDoH: social determinants of health

XSL•FO
RenderX

SMART: Substitutable Medical Applications, Reusable Technologies

Edited by A Mavragani; submitted 24.01.22; peer-reviewed by J Wolff, W Zhang, I Ioakeim-Skoufa; comments to author 04.03.22; revised version received 28.03.22; accepted 06.04.22; published 10.06.22.

Please cite as:
Norton JM, Ip A, Ruggiano N, Abidogun T, Camara DS, Fu H, Hose BZ, Miran S, Hsiao CJ, Wang J, Bierman AS
Assessing Progress Toward the Vision of a Comprehensive, Shared Electronic Care Plan: Scoping Review
J Med Internet Res 2022;24(6):e36569
URL: https://www.jmir.org/2022/6/e36569
doi:10.2196/36569
PMID:35687382

©Jenna M Norton, Alex Ip, Nicole Ruggiano, Tolulope Abidogun, Djibril Souleymane Camara, Helen Fu, Bat-Zion Hose, Saadia Miran, Chun-Ju Hsiao, Jing Wang, Arlene S Bierman. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.06.2022. 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

User Requirements for Comanaged Digital Health and Care: Review

Chaloner Chute[1], MSc; Tara French[2], PhD; Sneha Raman[2], MA; Jay Bradley[2], PhD

[1]The Digital Health & Care Innovation Centre, University of Strathclyde, Glasgow, United Kingdom

[2]The Glasgow School of Art, Forres, United Kingdom

Corresponding Author:
Chaloner Chute, MSc
The Digital Health & Care Innovation Centre
University of Strathclyde
1st Floor, Suite B, Inovo Building, 121 George Street
Glasgow, G1 1RD
United Kingdom
Phone: 44 01414447074
Email: chalonerchute@dhi-scotland.com

Abstract

Background: The sustainability of health and social care has led to an imperative to shift the balance of care to communities and support person-centered, integrated, preventive, comanaged, and sustainable care. The digital tool set can support this shift; however, it must extend beyond a clinical focus to include broader personal, social, and environmental needs, experiences, and outcomes. The existing digital health and care design and user requirements literature focuses mainly on specific digital products or design methods. There is little whole-system or whole-of-life consideration, which is crucial to enacting more significant transformations that span different groups and domains.

Objective: This study aimed to present a set of recurring user requirements and themes for comanaged digital health and care services derived from the body of co-design projects within a digital health and care program. This study aimed to enable people and organizations looking to reorient their approach to health and care research and delivery from a system-led and condition-specific approach to a more person-centric, whole-of-life model.

Methods: Participatory design formed the core methodological approach in underlying the design research, from which user requirements were derived. The process of surfacing requirements involved a selection framework for the identification of eligible projects and a structured review process to consolidate user requirements.

Results: This paper presents a set of 14 common user requirements that resulted from a review of co-design projects. The findings demonstrate overlapping and reinforcing sets of needs from citizens and care professionals related to how data are comanaged to improve care and outcomes. This paper discusses the alignment, contrasts, and gaps with broader, comparable literature. It highlights consensus around requirements for personal health storytelling, sharing data on care experiences and how this can support personalized guidance, visualize trends to support decision-making, and generally improve dialog between a citizen and care professionals. These findings identify gaps around how groups and networks of people engage, posing difficult questions for people designing support services as some of the user requirements are not easily met by organizations operating in silos.

Conclusions: This study proposes future recommendations for citizens as active, informed, and consenting partners using new forms of privacy-preserving digital infrastructure that puts the citizen in firm control. It is also recommended that these findings be used by people developing new digital services to ensure that they can start with knowledge of the broader user requirement context. This should inform domain-specific research and development questions and processes. Further work is needed to extend these common requirements to more explicitly consider the trust framework required when citizens comanage their data and care across a broad range of formal and informal actors. Consideration of how authority, delegation, and trust function between members of the public will be critical.

(J Med Internet Res 2022;24(6):e35337) doi:10.2196/35337

XSL•FO
RenderX

KEYWORDS

delivery of health care; integrated; patient-centered care; digital technology; decision-making; health services accessibility; trust; mHealth; eHealth; telehealth

Introduction

Background

Health care systems worldwide face unprecedented sustainability challenges that further exacerbate the impact of the COVID-19 pandemic [1,2]. A shifting political landscape and growing recruitment crisis further affect the United Kingdom's health care service delivery and staff well-being [3]. In parallel, there is an increasing policy and practice imperative to shift the balance of care to communities and enable a system that supports person-centered, integrated, preventive, comanaged, and sustainable care [4,5]. Scotland's strategy recognizes digital technology as a critical asset for delivering changes at scale [6].

In previous work, the authors have proposed that the digital health and social care tool set must help systems understand people's lived experiences [7]. This study defined individuals' health in terms beyond what a clinical record system holds to include broader personal, social, and environmental needs, experiences, and outcomes. The authors also argued for a balance between a health care system's need for controlled, governed, and secure record systems and a person's need for agency, trust, choice, and the ability to connect their data across agencies, informal care circles, and communities. The authors' *systems of record* arguments are nested inside the need for broader changes to culture and practice, from a focus on transactional relationships between citizens and systems to a more personalized and collaborative approach to care and support. In addition, the study proposed that a care system must use any formal or informal assets to sustain engagement, care interactions, and experiences on a comanaged basis. This position recognizes the complexity of the digital health and care ecosystem among the stakeholders involved, the continued exploration required to understand the efficacy of digital methods, and the challenges of various digital tools and products [7].

Therefore, this concept of the comanagement of health and social care emphasizes working in partnership with citizens to organize multiple relationships and assets to deliver person-centered care. This approach will create more sustainable methods to meet citizens' support and self-care needs and wishes through mutual discussion and decision-making. This previous work concluded by contrasting these principles with other approaches that focused on organization-centric needs, practices, and efficiencies [7].

The literature on digital health and care design and user requirements focuses mainly on either digital products [8-16] or design methods [17-22]. This focus limits digital health and care domain knowledge to silos, such as individual clinical conditions, care groups, clinical specialties, or domains of influence (eg, health care, social care, housing, and social security). There is no whole-system or whole-of-life consideration, indicating that people looking to enact more significant transformations have no common reference to span these groups and domains. This lack of whole-system thinking makes it difficult to formulate strategies, policies, and digital architectures to satisfy the person-centered, integrated, and comanaged care ambitions set out in government health care transformation strategies described previously.

Other attempts have been made to close similar gaps through frameworks and guidance, such as the National Institute for Health and Care Excellence Evidence Standards Framework for Digital Health and Care Technologies [23] and the World Health Organization guidance on digital health for researchers [24]. However, these were limited to one system satisfying requirements related to safety and did not acknowledge the user requirements and benefits of involving citizens to create value for both the people and the system. There is a need to move beyond acceptability and feasibility to ensure that the future introduction and development of digital tools meet the identified needs.

On the basis of an analysis of insights from a co-design program, this study proposes a set of recurring user requirements and themes for comanaged digital health and care services. The findings present a starting point for further development and future research on digital health and care support options. It provides evidence of unmet needs in a whole-system context to support more holistic and integrated care led by the person.

Related Literature

This study categorized the digital health and care user requirements literature into three main areas: (1) co-design methods to change services or elicit user requirements; (2) topic-, condition-, or product-specific design exercises; and (3) a cross-cutting review or discussion of general user requirements.

Greenhalgh et al [25] reviewed the cocreation literature, identifying different threads across disciplines, including business studies (*value cocreation*), design science (*experience - based co - design*), computer science (*technology co - design*), and community development (*participatory research*). They noted commonalities across the methods that determined success. These were (1) systems thinking, (2) focus on creativity and human experience, and (3) emphasis on process, (eg, relationships, governance, leadership, and conflict). The broader literature reflects this characterization with evidence of value c-creation [17], experience-based co-design [18], technology co-design [19-21], and participatory research [22]. Additional studies reinforce the need to create end-to-end co-design frameworks that look beyond technological cocreation [26,27].

The literature demonstrates a diverse range of co-design projects covering healthy aging [8], palliative care [9], cancer care [10], medicine adherence [11], reablement [12], and a range of long-term conditions [13-16]. The precise methods vary, from those undertaking user research through semistructured interviews [8,9] to those focusing on group-based activity

XSL•FO
RenderX

following user-centered design principles [10,13-15] to more product-oriented approaches [11,12,16].

The peer-reviewed literature for a cross-cutting approach to user requirements is limited, focusing on common perceptions and insights rather than user requirements [28,29]. The gray literature provides the most cross-cutting user requirements curation and analysis. This material is provided by professional membership bodies [30,31], regulators [32], or innovation and delivery agencies [33,34].

Overall, the literature focuses on design methods and co-design to elicit user requirements for specific technologies or services. However, although there are calls for more focus on whole systems, human experience, and processes, the literature does not yet fully define common user requirements across diverse groups of people, conditions, services, and technologies.

Objective

This study aimed to address this gap by sharing a set of common user requirements based on lived experiences from a range of co-design projects across the health care continuum, which were undertaken within a digital health and care program spanning 7 years [35]. Through this, the authors sought to inform the future development of digital health and care interventions based on human experiences that span whole systems. The requirements shared in this paper have enabled the program to evolve to undertake rapid coproduction-based service modeling, prototyping and integration, and possible deployment exercises. Therefore, this study intended to support knowledge sharing to enable others to develop similar infrastructure and methods to help move beyond a purely digital product focus and satisfy the cross-cutting data-sharing, coordination, and integration needs described hereafter.

Methods

Overview

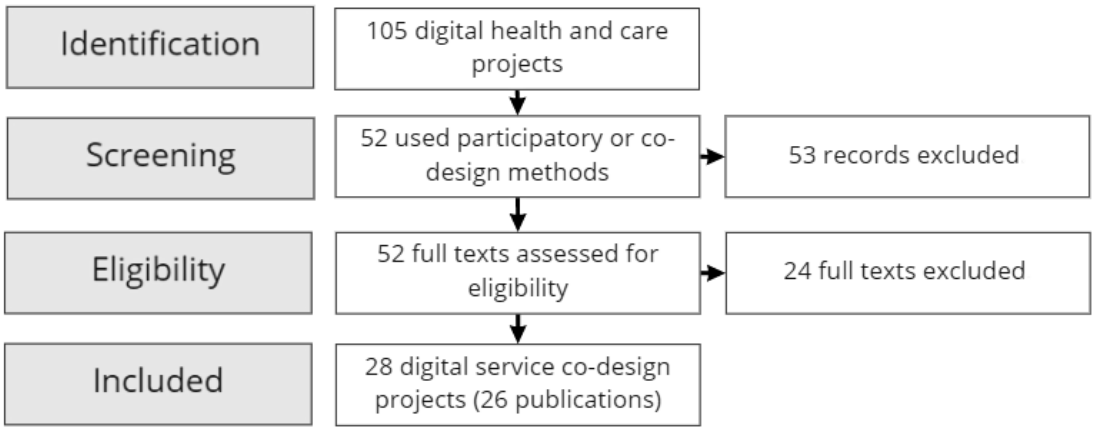
Before providing details regarding the specific process of identifying user requirements, it is important to provide the methodological context and the range of projects from which the requirements were derived. The reviewed projects were undertaken as part of a large-scale digital health and care program in Scotland. The Digital Health & Care Innovation Centre (DHI) was established in 2013 as a response to the need to support collaborative approaches in research and innovation across academia, civic organizations, and industry partners. The initial model of the DHI recognized the value of academic research in evidencing and testing ideas for innovation, particularly in the digital health and care context where the previous introduction of technological solutions failed to meet the needs of health care services. In addition, design-led approaches that supported rapid prototyping and testing of solutions provided the opportunity to learn quickly and iterate with the benefit of involving key stakeholders in the co-design process. During the first phase, the DHI commissioned and delivered 105 projects over 3 years. The projects reached varying stages of maturity, with some intended for concept exploration only, whereas others went into live clinical and academic trials.

Participatory design is the core methodological approach and design research practice of academics working as part of the DHI to engage a diverse range of participants in the co-design of digital health and care projects. The approach across the projects from which the user requirements were derived involved a range of methods to engage people in co-design, such as interviews, workshops, experience prototyping, creation of lived experience personas, speculative design, and the wizard of oz techniques. The methods applied within each project were bespoke to the people, topics, and outcomes in question.

Surfacing Requirements

The process of surfacing user requirements, outlined in Figure 1, involved reviewing 52 co-design projects over 7 years, working with >3500 citizens, >1000 health care professionals, 16 health boards, 15 charities, and 10 social care providers. These projects supported diverse groups, generating insights across several areas, including healthy aging, mental health, and long-term condition management.

Figure 1. Initial study selection.



The eligibility criteria used to filter projects were as follows:

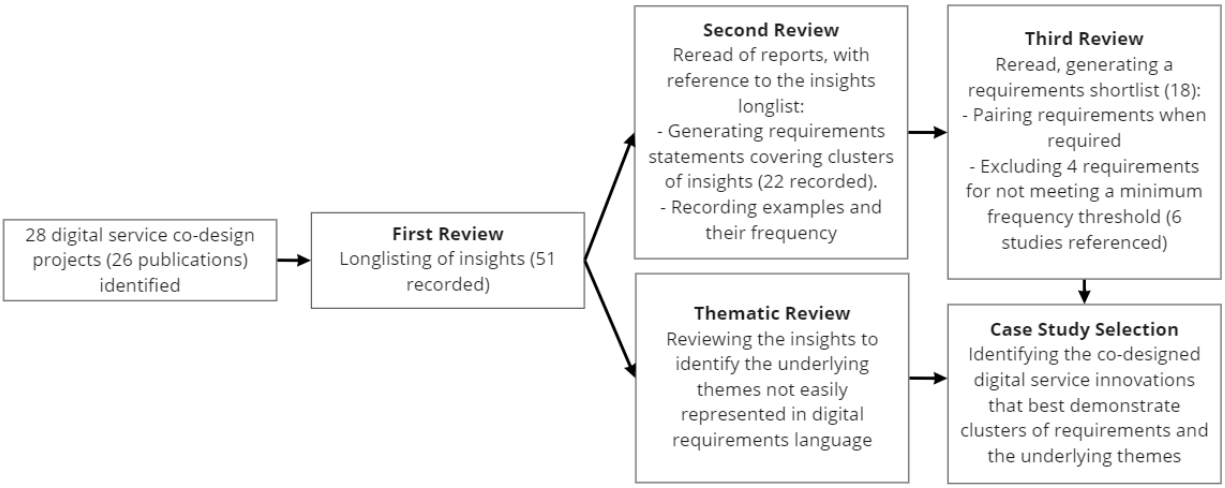
- The project focused on services where citizens were engaged in the co-design process (eg, diabetes or multiple sclerosis). Other projects were excluded if citizens were not directly engaged; for example, ambulance clinician decision support.

XSL•FO
RenderX

- The project explored communication, decision-making, and planning, which could be improved through digital tools, such as blood donation services. Other excluded projects focused on tactile, physical, and rapport-based interactions; for example, music therapy and objects to help young people communicate their needs.
- The projects were from the first 4 years of the co-design portfolio work. This criterion was introduced as the second half of the program involved co-design work that was informed by learning and experiences to date. This minimized the risk of design insights being overtly influenced by the design team because of their accrued knowledge and experiential learning.

Following project selection, the team undertook 3 reviews, as shown in Figure 2, of project reporting to gather and summarize co-design insights.

Figure 2. Review process.



The insights were then clustered and synthesized into a set of common user requirements (>6 project references) in an appropriate user design format: "As a [type of person or role] I want to be able to [do something] to [achieve a goal]." Insights that did not fit within the user requirements language and format were analyzed thematically. Case study insights and innovative ideas from the co-design activities that showed clusters of requirements and themes in context were selected. Where possible, the outlined user requirements and themes were curated using language that does not heavily lean toward any given domain; for example, using generic language around people, personal data, dialog, and care rather than medical language around patients, clinical data, appointments, and treatment.

Results

Common Citizen Needs

The findings were organized into 4 parts. Tables 1-5 are a summary of the user requirements mapped to the projects that generated insights. Example quotes from the co-design participants were included to illustrate the original insights. The second part is a small section summarizing the professional needs that arose in parallel through the same co-design projects to demonstrate that in most cases, the professionals were asking for the same tools, both for the citizen and to improve activities and outcomes when providing care. The third section describes the emergent themes that could and should not be translated into the required language. The fourth section offers a case study that draws together many themes and requirements to illustrate user needs in context.

Table 1. Requirements: telling my story once.

Codes	Requirements (as a citizen comanaging health care services, I want to be able to)	Example quotes	Co-design studies
P1	Hold and share my personal health story and have services use this to personalize my care	• "Different people every time...It can be a bit annoying I think for anyone, if you have one main doctor and you're seeing ten other different ones, feel like you're telling the same story over and over and over again." [Person living with diabetes] [36] • "...it would be really nice if there was a little bubble with my story there without me having to say it again and again." [Person living with multiple sclerosis] [37]	[36-54]
P2	Share my experience and outcomes and for this to improve care for myself and others in the future	• "Perhaps when I am sending notes to you, you can see, 'yes, she cycles once a week'—or 'she works seven days a week on her back-side!' I think [the consultant] needs to know that people are doing some level of exercise." [Person living with diabetes] [40]	[37, 41, 43, 48, 50, 51, 53-57]

XSL•FO
RenderX

Table 2. Requirements: meaningful dialog with professionals.

Codes	Requirement (as a citizen comanaging health care services, I want to be able to)	Example quotes	Co-design studies
P3	Have conversations with professionals that focus on my priorities	• "It's just trying to balance up what the patient's needs are, versus your own agenda with them." [Care professional] [39] • "In the holistic needs assessment, the client will tick what concerns they have and will also score them out of ten. If someone's scored something ten then that's a really high concern for them, and that to me would be a priority" [Care professional] [48]	[38-40, 43, 48, 51, 52, 54-56, 58, 59]
P4	Have conversations with professionals who have the necessary information or test results available and gathered ahead of time	• "...before I come in you would be reading [my] notes, and I'll have a wee drop-down box with the questions I would like to ask you about my blood sugar levels, so you have [time] to think 'oh that is what she wants to discuss today'" [Person living with diabetes] [40]	[38-43, 51, 53, 54, 56, 57]
P5	Have an ongoing dialog with professionals outside of formal appointments, allowing me to ask questions on my own terms	• "...you always forget everything. The number of times I go to a clinic appointment, and I think 'oh, I must ask them this,' and then afterwards you go out and my mum's like, 'did you ask about...?'" [Person with asthma] [54]	[43,50-52,55,56]

Table 3. Requirements: access and understand data.

Codes	Requirements (as a citizen comanaging health care services, I want to be able to)	Example quotes	Co-design studies
P6	Access personalized guidance, signposting, and navigation support based on my personal health story	• "How important is it that you can personalise the system? 100% That's how you make it work for you." [Older adult] [60] • "That's one of the challenges for patients, if clinical staff potentially aren't aware of the service, it could take somebody a long time to then get engaged" [Care professional] [48]	[36, 38, 40, 42, 50-52, 54, 55, 61, 62]
P7	Have joint visualizations of clinical and personal data available to help me and others to see patterns and trends over time	• "It's all about constant monitoring and recording and using previous experience." [Person living with diabetes] [41]	[36, 37, 39-43, 51, 53, 54, 56, 63]
P8	See a timeline or route map of my care interactions and understand their content and purpose	• "I wouldn't know who to contact or even if you phone the MS nurse, you leave a message, and they'll get back to you but even that gets lost in translation...I do tend to write things down...I must get a book because bits of paper just go missing, I know it's my biggest problem." [Person living with multiple sclerosis] [38] • "And also we said about having the care package—how much care is coming in and what times they are going in, because often we'd be the same—we do joint visits with carers, and you are running around trying to find out what times carers are coming in." [Professional supporting someone living with multiple sclerosis] [38]	[38, 40, 50, 51, 55, 56, 59]

XSL•FO
RenderX

Table 4. Requirements: do things on my own terms.

Codes	Requirements (as a citizen comanaging health care services, I want to be able to)	Example quotes	Co-design studies
P9	Use my technology to access services and monitor myself to support my care	• "Fitbits are quite trendy but [anon] is wearing something here, she's wearing something here, she walks about with a bottle of Lucozade and sweeties so something else would drive her nuts, she just wants to fit in and be normal. A Fitbit is a good example because everybody wears one now..." [Carer of a person living with diabetes] [41]	[36, 37, 39-43, 49, 53, 54, 56, 60, 63]
P10	Manage my circle of care, communicating and sharing data with my peers, family, friends, care professionals, and community organizations	• "I quite like to get advice from other mums as professionals so it's like real-life experiences, even if those professionals have fed themselves, it's nice to have some mums that are going through it at that particular point" [Mother] [61]	[36-39, 42, 45, 47, 50, 51, 55, 58, 60-63]
P11	Jointly manage personal, "whole-of-life" care plans with my circle of care, agreeing to actions, access rights and triggers in advance	• "So there's a team of support there but I kind of needed to hold in my head that these are all people that can be accessed. But I'm quite motivated and articulate so I have pieced together the system that works for me, and the journey has meant that different people have taken centre-stage at different times." [Person living with multiple sclerosis] [38] • "Things need to be kept local—once it goes to a big organisation they might use as evidence to say you need to go to a home...makes it more personal—a friend, a neighbour." [Older adult] [60]	[41, 42, 45, 51-53, 55, 56, 59, 60]

Table 5. Requirements: use my data to unlock care.

Codes	Requirements (as a citizen comanaging health care services, I want to be able to)	Example quotes	Co-design studies
P12	Trust in how others use my personal information	• "The client needs to be able to trust us to be able to get the information from them" [Care professional] [48] • "...to get a hold of all of these powerful things that are in the room takes understanding and skill and compassion and it needs somebody to make it safe." [Person living with multiple sclerosis] [38]	[36, 40, 51, 54, 58, 60, 63]
P13	Share relevant, trusted data with people who can help me	• "If I need help, privacy goes out of the door" [Older adult] [58] • "So to be able to have a once and for all, okay, it's not going to be once and for all because it's changing all the time, but a template for my story of MS with all the awful bits remembered but without having to keep on doing it with each agency you engage with, having to prove yourself." [Person living with multiple sclerosis] [38]	[37-40,51,54,63]
P14	Have the authority to activate services that I am entitled to myself	• "Although there might be things there, there was no trigger mechanism to trigger services happening" [Older adult] [62] • "If it's combined with respiratory infection, I know that so I can go to the hospital, but if it's just cold air or air quality I'm more towards staying at home than going to the hospital." [Person with asthma] [54] • "Once (the GP) has made a diagnosis that someone has MS, that can be represented by a letter or it can be represented by a digital letter or it can be represented by a digital token and the person could carry that with them, as they do, or it could be live with (local) Council that when this person rings there's a token that comes up to say that this is who they are." [Person living with multiple sclerosis] [38]	[38, 47, 48, 50-54, 58, 62]

Common Professional Needs

The co-design process often included citizens, carers, and health care professionals. Although not the focus of this paper, this section outlines the most common professional requirements and co-design projects generating these insights. This high-level summary illustrates that citizen requirements do not exist in a vacuum. In most cases, carers and professionals want the same types of data-sharing and navigation tools to help citizens and professional teams better coordinate care together (Textbox 1).

Further review is underway to explore professional comanagement needs in more detail.

XSL•FO
RenderX

Textbox 1. Care provider requirements.

Care provider requirements (as a care provider comanaging health and care services, I want to be able to)

Care provider requirement 1

Access and contribute to an individual's personal health story so that I can deliver more personalized care and enhance dialog and joint decision-making [38-43,46,48,50,53-58,63]

Care provider requirement 2

Share and visualize where the individual is on their current care pathway, personalized to their story to help us both manage and prepare [39,40,42,48,51,53,55,56,59]

Care provider requirement 3

Help me and the individual understand their condition better through the joint recording of, and access to, personal symptoms, triggers, medications, and test results [37,38,40,41,43,51,54-57]

Care provider requirement 4

Empower an individual with the knowledge and assets to either self-manage or escalate to other people or services [37-42,46-52,54-56,58,61]

Additional Recurring Themes

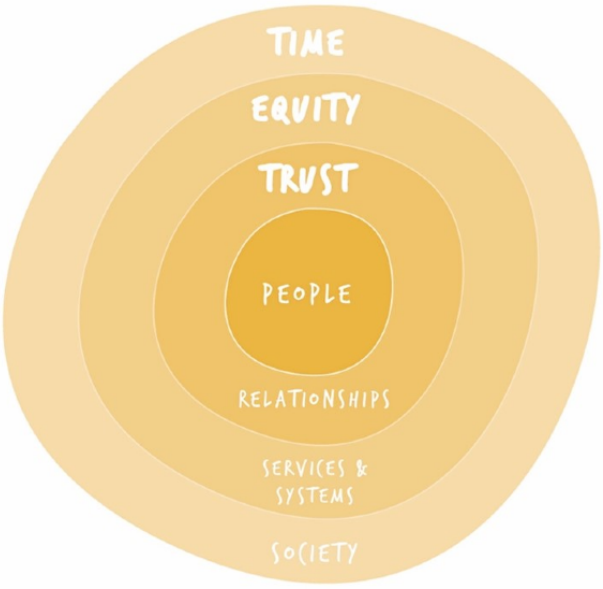
In addition to the user requirements discussed in this paper, shared themes related to emerging principles and visions for future health care emerged during the review of the design research team's co-design work. Although it is not within the scope of this paper to discuss these in detail, this section presents an overview of these themes to contextualize user requirements within broader transformations that are required socially, culturally, and politically to guide future innovation in health and care.

The emerging principles and visions for the future, as depicted in Figure 3, focused on the following:

- Enabling a person-centered focus on understanding the whole person rather than their health condition with systems built around people's holistic needs and what matters to them
- Trust across all levels of the health care system, with a key focus on interpersonal and professional relationships
- Equity of access to information, services, and systems, revealing a tension between the need for standardization and tailoring of care
- Ensuring citizens and health care professionals have time to care for themselves and others

This paper presents an evidence base of user requirements for the future development of digital health and care interventions. However, their use will only lead to ethical and meaningful care experiences and outcomes if systemically and culturally underpinned by the values of trust, equity, and time.

Figure 3. Co-design themes (image courtesy: author SR).



Case Study: Backpack

The *Backpack* research project provides an illustrative use case for this paper to contextualize the user requirements. The project aimed to explore how people living with multiple sclerosis would like to manage their personal information to improve their experience of accessing services and understand the potential of a person-owned data store (or digital *Backpack*) in the delivery of integrated and person-centered care [38]. The project involved engaging people living with multiple sclerosis in a focus group and a co-design workshop with health and social care professionals.

XSL•FO
RenderX

Findings from the focus group revealed the need for people to do the following:

- Retell their health stories repeatedly when accessing services and across multiple interactions with different parts of health and social care
- Understand what health and social care services are doing, including knowing about and navigating available services and keeping track of the people within their care team
- Cope with transitions and the requisite change in their care

In the first workshop, the participants designed their own physical *backpack*, which served as a relatable analogy for a person-owned record. This process included considering what information they would store and how, why, and with whom they would share it. In the final workshop, health and social care professionals set typical health care tasks to explore how person-owned records might change the way they work through paper-based and digital prototypes. Through these activities, 4 concepts emerged regarding the future use of person-owned records. Table 6 outlines these concepts and relates them to the defined common user requirements.

The findings from the Backpack project are evident in many of the requirements shared in this study, as illustrated in Table 6.

Table 6. Backpack innovation concepts.

Co-designed innovation concept		Related user requirements
Concept	Description	
Circle of care	• This is a digital means of mapping interactions with formal and informal health and care systems. • The complex, multi-organization services accessed by people living with multiple sclerosis (MS) means that it can be challenging to understand who they are seeing or have seen and for what purpose. The same issue occurs for the health and social care professionals who provide care, leading to unneeded or duplicated referrals, poor resource use for the health care system, and unhelpful interactions for the citizen. • Participants discussed "building your own care team"—mapping interactions with the person at the center of a connected network of professionals and a timeline showing who they saw or will see. • By mapping what has happened and what will happen, the "backpack" should create a space for shared and shared decision-making, leading to a more equitable relationship between care providers and people who access support. • Circle of care technologies are beginning to emerge in practice, for example, to support care for parents and children [64].	P8, P10, P11, and CP2
Health story	• Citizens can tell their own health stories using their choice of format and content. • Participants often found that professionals lacked even basic knowledge about them but did not wish to recount their stories repeatedly. • Professionals saw value in understanding the citizen and their needs before meeting for the first time. • A health story might be text or video, contain key dates (eg, diagnosis or change in personal circumstance), and can be updated. It may also contain other suitable information, such as a video of their home environment. • A health story would be shared by consent from specific organizations or publicly (which might help others in similar situations). These preshared health stories have been shown to improve communication between care professionals and the people they are supporting [65].	P1, P3, and CP1
Automatic form filling	• This is a digital means of avoiding repeated form filling. • Existing data in a person's backpack could autofill many form fields. • In particular, the data could be used to identify eligibility criteria quickly and easily. • This method would replace the need to "make yourself known" to various health care providers to find out what services are available to an individual. • This mechanism would gradually fill up, collecting data as it went, and would not require a massive data entry exercise at the beginning. • The stored data could be used, with the person's consent, to find eligible services automatically. • Participants also suggested an "in case of emergency" feature to share their backpacks with nominated people if necessary.	P6, P13, and P14
Responsive case management	• This includes digital tools that help professionals in sharing information and caring for people with person-owned records. • A regional MS nurse does not necessarily know about any change in circumstance for the people they support. They may not be informed of hospital admissions, deterioration in a condition, or even death. This mechanism was visualized as a list of all people with MS in the region. • The MS nurse can send messages to individuals or a subset that matches chosen criteria. • The MS nurse could order the patients according to criteria and be alerted to any change of circumstances entered in the "backpack" by citizens. • This mechanism would allow the MS nurse to effectively help many people with MS. • Comanaged digital tools are now being studied, which use wearable data and patient-reported outcome measures to help clinical teams identify and respond to change [66].	P2, P5, P6, CP1, and CP3

XSL•FO
RenderX

Discussion

Principal Findings

This study provides a robust and systematic analysis of common user requirements for the digital comanagement of care. On the basis of a diverse body of co-design work, it provides a starting point for people and organizations looking to reorient their approach to health care data sharing from an organization-centric to a person-centric model. This study set out to create an initial frame of reference for *whole-system* service and system design, underpinned by insights generated through co-design with a wide range of user groups across multiple domains. These findings demonstrated a consistent set of user requirements that look beyond individual technologies and processes specific to one type or domain of care. Through the active participation of both citizens and care professionals in the underlying design research, the findings also demonstrated overlapping and mutually reinforcing sets of needs from both groups related to how data are comanaged to improve care and outcomes.

Comparison With Prior Work

The peer-reviewed literature focuses mainly on co-design methods and technologies for individual health and care services. The work comparable with this synthesis was extremely limited, with some studies focusing on common perceptions and insights [28]. Other studies elicited more definitive requirements for more complex needs but still with an individual product focus [24]. More systematic user requirements curation and analysis were primarily found in the gray literature provided by professional membership bodies, regulators, or innovation and delivery agencies. As a result, although these pieces were broader in scope, they were still tied to individual domains, mainly medical [30,33] and social care record keeping [32]. Table 7 maps these 5 comparators against the common requirements described in this study. It identifies the areas of complete alignment and partial alignment.

In contrast to the approach shared in this paper, which focused on co-design insights generated separately from any given product, Vo et al [28] analyzed 43 studies reviewing mobile health (mHealth) apps providing commentary on existing apps. However, the analysis rose above individual products or methods, focusing on the common strengths and weaknesses of mHealth apps. This study's concepts map well relative to our findings, particularly regarding personalization, meaningful dialog, and citizen participation. Without digital tools that rebalance the power dynamic between citizens and professionals, citizens may otherwise "resign themselves to receiving care without taking up the possibility to engage in active participation" [28]. Both their work and ours found that citizens wish to use these digital tools to facilitate relationships and not to replace them. Finally, the study identified concerns about the scientific validity of some mHealth apps, which were not covered within our user requirements' elicitation [28].

The cross-cutting needs for patients identified by Bhattacharyya et al [29] arrived at a broadly comparable set of user requirements. The key features were again mapped to the most common requirements in this study, focusing on elements relating to trend analysis, navigation, and guidance. All key features were covered by the common requirements presented in this study [29].

Further overlap was evident in the set of benefits and other supporting materials of interviews and focus groups on the topic of personal health records. However, the findings focused on transactional National Health Service (NHS) service access (beyond the scope of this work), such as reminders for medications or access to test results [30].

In the context of previous work on user needs relating to personal health records, there was further alignment with the user requirements, particularly the data-sharing relationship between a patient and clinical team, but with an additional focus on trust and privacy not readily evident in the broader user requirements literature. Another parallel was the account of complementary professional needs in the comanagement of data. A wider NHS work identified additional requirements beyond the findings in this paper around the delegation of authority over data and ethical limitations to medical data sharing [31,33]. Equivalent exercises in the social care domain focused on general shared care record methods, prioritizing more joint care team capabilities to improve citizen outcomes [32].

Overall, the requirements aligned with previous research and strongly in the case of requirements for personal health storytelling, sharing data on health experiences and how this can support personalized guidance, visualizing trends to support decision-making, and generally improving dialog between a citizen and a care professional (a *vertical* relationship). However, there were notable differences where this study makes key contributions. The first contribution was the new knowledge curated, with common requirements identified in this paper, extending to cover more *horizontal* relationships and more holistic needs beyond dialog with any one professional; for example, the need to create care plans and manage care circles involving multiple professionals, informal carers, agencies, and technologies. Another example was the ability to comanage the data itself, with personally held data being trusted by professionals and, in turn, professionals being trusted by citizens to use their personal data appropriately.

There were some contrasts and gaps; for example, only this study identified that citizens needed to have authority granted to them and the data they hold. This authority was crucial to creating a more effortless experience in demonstrating eligibility when moving between professional domains (eg, for a benefit or being able to access rationed specialist services directly). Overall, the differences were related to the scope of the different pieces of work. The review of projects was concerned with the whole of life and integrated services and, thus, reflected requirements that spanned domains. Previous studies, acting out of only one domain, tended to reflect requirements that optimized citizen-professional dialog within that domain, service, or specialty [30-33].

The second key contribution was related to this method. This study and the summarized evidence covered 3 main elements that were not entirely paralleled by any of the previous key work comparators. For example, this paper has reviewed a large body of co-design projects, considered citizen and professional needs

XSL•FO
RenderX

in tandem, and generated specific user requirements through this review (Table 8).

This finding points to the need for robust literature that summarizes and translates large bodies of co-design input into requirements language to support the comanagement of care at the whole-system level.

Table 7. Common requirements comparisons across publications.

Common requirement (authors)	Vo et al [28]	Bhattacharyya et al [29]	Royal College of Physicians [30]	NHS[a] Digital [33]	Care Quality Commission [32]
Hold and share my personal health story and have services use this to personalize my care	Partial alignment	Complete alignment	Complete alignment	Complete alignment	Partial alignment
Share my experience and outcomes—and for this to improve care for myself and others in the future	Complete alignment	Complete alignment	Complete alignment	Complete alignment	Complete alignment
Have conversations with professionals that focus on my priorities	Complete alignment	Partial alignment	Complete alignment	N/A[b]	Complete alignment
Have conversations with professionals who have the necessary information or test results available and gathered ahead of time	Partial alignment	N/A	Complete alignment	Complete alignment	Complete alignment
Have an ongoing dialog with professionals outside of formal appointments, allowing me to ask questions on my own terms	Complete alignment	N/A	Complete alignment	Complete alignment	N/A
Access personalized guidance, signposting, and navigation support based on my personal health story	Complete alignment	Complete alignment	Complete alignment	Complete alignment	Complete alignment
Have joint visualizations of clinical and personal data available to help me and others to see patterns and trends over time	N/A	Complete alignment	Complete alignment	Complete alignment	Complete alignment
See a timeline or route map of my care interactions and understand their content and purpose	N/A	Complete alignment	N/A	Partial alignment	N/A
Use my technology to access services and monitor myself to support my care	Complete alignment	N/A	Complete alignment	N/A	N/A
Manage my circle of care and communicate and share data with my peers, family, friends, care professionals, and community organizations	N/A	N/A	Partial alignment	N/A	N/A
Jointly manage personal, "whole-of-life" care plans with my circle of care, agreeing to actions, access rights and triggers in advance	N/A	Complete alignment	N/A	Partial alignment	N/A
Trust in how others use my personal information	Complete alignment	N/A	Partial alignment	Complete alignment	Complete alignment
Share relevant, trusted data with people who can help me	N/A	N/A	Complete alignment	Complete alignment	Complete alignment
Have the authority to activate services that I am entitled to myself	Partial alignment	N/A	N/A	N/A	N/A

[a]NHS: National Health Service.

[b]N/A: not available.

Table 8. Comparison of study elements.

Study element	Chute et al [7]	Vo et al [28]	Bhattacharyya et al [29]	Royal College of Physicians [30]	NHS[a] Digital [33]	Care Quality Commission [32]
Reviewed large body of design studies	Yes	Yes	No	No	No	Yes
Considered citizens and professionals in tandem	Yes	No	No	No	Yes	Yes
Generated specific user requirements	Yes	No	Yes	Yes	Yes	No

[a]NHS: National Health Service.

XSL•FO
RenderX

Implications for Practice

The requirements summarized in this paper pose difficult questions for people designing health care, social care, and broader support services as they are not easily met by organizations operating in silos. For example, the most universal of all the studied citizens' needs was that of people wanting to *tell their story once* and not repeat themselves across different parts of the system. Although there were numerous initiatives to create a joint approach, they rarely looked across domain boundaries. This problem is best illustrated by the ongoing pursuit of a single clinical record and the domain-specific goal of aggregating all clinical data to drive improved care and outcomes. This single medical record would undoubtedly help health services and individual patients have more joint medical care. However, it would not meaningfully change the way citizens transact with social security and housing or empower their informal circle of care or a third or independent sector organization to support their nonmedical needs. A record dictated by a medical model and associated standards and governance is not likely to tolerate new forms of data generated by citizens, broader organizations, and other sources that would increasingly allow for more context-rich, whole-of-life outcomes to be pursued through greater personalization and prevention.

However, a citizen-comanaged, holistic story would be heavily dependent on the quality of the organizational systems and the data it must synchronize with. A prerequisite for more citizen control and reuse of their medical records requires that those records be well-defined and structured and that those supplying health care software conform to modern, standards-based practices. In technical terms, these challenges are beginning to be met at scale by the proliferation of application programming interfaces, messaging standards (eg, Fast Healthcare Interoperability Resources [FHIR]), and data storage models (eg, OpenEHR). The technical barriers are increasingly surmountable; however, a more significant effort will be required in evolving the culture, commissioning, and supplier practices to adhere to standards and separate the data from software products to enable its reuse.

The routes available to meet the user requirements outlined in this paper will almost certainly do so with the citizen as an active, informed, and consenting partner using new forms of privacy-preserving digital infrastructure that puts the citizen in firm control. Only through this kind of comanagement of data can comanagement of care that respects *whole-of-life* needs and satisfies whole-system governance and trust be achieved. The findings in this study can be used by people developing new digital health and care services to ensure that they can start with knowledge of the broader user requirement context. This should inform domain-specific research and development questions and processes. For example, when creating a shared care record between health and social care, the citizens' requirements in this study may encourage system designers to consider how the record needs to be viable beyond either of those 2 domains. This additional consideration may help us collectively move toward systems that support citizens to *tell their story once* and reuse the record to access broader support services.

Limitations and Future Work

There are three main limitations to this study and several ways the authors attempted to mitigate them.

First, all source projects were undertaken by the same design research team from one institution (with other universities occasionally collaborating). Therefore, although diverse, the methods and results came from the same collective approach to co-design, which may have limited the general applicability of the outputs. A comparative review of the broader literature, as documented in the *Introduction* section, aimed to test the findings in a broader context to mitigate this.

Second, as a design research group, knowledge and design experience grew. Therefore, later work was often informed by earlier work, which may have created patterns based on the interventions. This study was limited to reviewing the first 4 years of work to mitigate this effect. Further ≥30 recent projects have not been included to avoid more recent work artificially inflating the perceived commonality of these requirements.

Finally, most of the design research was commissioned by the NHS. Although many participants were in the social care and third sector, the overall tone of most of the work was health (clinical care) focused. The team strove to take a step back from the initial commissions and used design methods to ensure that the project asked the right questions across diverse populations in a broader context of health, care, and well-being. However, the overall tone is undeniably still clinical care focused; thus, future work is needed to expand the understanding of nonclinical common user requirements to complement the findings of this paper.

Future research on co-design and requirements elicitation could build on this foundation and address several gaps. For example, although *circle of care* and *joint care planning* were common requirements, they are both concepts that span many people and organizations. Therefore, more work is required to harmonize requirements and data sets across multiple actors. It is also unclear where peer networks (eg, diabetes management community networks) end and circles of care (eg, friends, family, and carers) begin and the level of data sharing and privacy that relates to these different types of relationships. Finally, the concept of delegation of authority has begun to emerge as health care systems become more digitally enabled. To support equity of access and maintain interpersonal care relationships, some groups will need to name and delegate authority to trusted people, who can then act on their behalf with digital services.

Conclusions

This paper demonstrated common user requirements relating to the comanagement of care between citizens and their circles of care. The common requirements relating to *vertical* relationships between a citizen and a professional were corroborated by comparator literature. The common requirements extended to cover the *horizontal* relationships between people and their broader support networks across services and agencies and their informal circles of care. Further work is needed to extend these common requirements to more explicitly consider the trust framework required when citizens

XSL•FO
RenderX

comanage their data and care across a broad range of actors. Consideration of how authority, delegation, and trust function among members of the public will be critical. The authors propose that these user requirements can inform service design and data-sharing infrastructure across organizations involved in providing health, social care, and well-being support. We welcome further dialog on how these requirements can drive a person-centered integration agenda that brings value to people and the system.

Acknowledgments

The authors would like to thank all participants and partners who were part of the co-design projects shared in this paper. The authors are grateful to their colleagues across the Digital Health and Care Innovation Center and their respective institutions—The Glasgow School of Art and the University of Strathclyde. The Digital Health and Care Innovation Center is funded by grants from the Scottish Funding Council and the Scottish Government Health and Care Department.

Conflicts of Interest

None declared.

References

1. Organisation for Economic Co-operation and Development. Fiscal Sustainability of Health Systems: Bridging Health and Finance Perspectives. Paris, France: OECD Publishing; 2015.
2. Moynihan R, Sanders S, Michaleff ZA, Scott AM, Clark J, To EJ, et al. Impact of COVID-19 pandemic on utilisation of healthcare services: a systematic review. BMJ Open 2021 Mar 16;11(3):e045343 [FREE Full text] [doi: 10.1136/bmjopen-2020-045343] [Medline: 33727273]
3. Charles A, Ewbank L. The road to renewal: five priorities for health and care. The Kings Fund. 2021 Apr 8. URL: https://www.kingsfund.org.uk/publications/covid-19-road-renewal-health-and-care [accessed 2001-12-21]
4. Kringos DS, Boerma WG, Hutchinson A, Saltman RB. Building primary care in a changing Europe. European Observatory of Health Systems and Polocies. 2015. URL: https://www.euro.who.int/en/publications/abstracts/building-primary-care-in-a-changing-europe-2015 [accessed 2001-12-21]
5. Healthcare Quality Strategy for NHS Scotland. Scottish Government. 2010 May 10. URL: https://www.gov.scot/publications/healthcare-quality-strategy-nhsscotland/ [accessed 2012-11-21]
6. Digital health and care strategy: enabling, connecting and empowering. Scottish Government. 2018 Apr 25. URL: https://www.gov.scot/publications/scotlands-digital-health-care-strategy-enabling-connecting-empowering/ [accessed 2018-11-21]
7. Chute C, French T. Introducing care 4.0: an integrated care paradigm built on industry 4.0 capabilities. Int J Environ Res Public Health 2019 Jun 25;16(12):2247 [FREE Full text] [doi: 10.3390/ijerph16122247] [Medline: 31242687]
8. Elers P, Hunter I, Whiddett D, Lockhart C, Guesgen H, Singh A. User requirements for technology to assist aging in place: qualitative study of older people and their informal support networks. JMIR Mhealth Uhealth 2018 Jun 06;6(6):e10741 [FREE Full text] [doi: 10.2196/10741] [Medline: 29875083]
9. Nkhoma KB, Ebenso B, Akeju D, Adejoh S, Bennett M, Chirenje M, et al. Stakeholder perspectives and requirements to guide the development of digital technology for palliative cancer services: a multi-country, cross-sectional, qualitative study in Nigeria, Uganda and Zimbabwe. BMC Palliat Care 2021 Jan 04;20(1):4 [FREE Full text] [doi: 10.1186/s12904-020-00694-y] [Medline: 33397321]
10. Harder H, Holroyd P, Burkinshaw L, Watten P, Zammit C, Harris PR, et al. A user-centred approach to developing bWell, a mobile app for arm and shoulder exercises after breast cancer treatment. J Cancer Surviv 2017 Dec;11(6):732-742 [FREE Full text] [doi: 10.1007/s11764-017-0630-3] [Medline: 28741202]
11. Mehta P, Moore SL, Bull S, Kwan BM. Building MedVenture - a mobile health application to improve adolescent medication adherence - using a multidisciplinary approach and academic-industry collaboration. Digit Health 2021 May 22;7:20552076211019877 [FREE Full text] [doi: 10.1177/20552076211019877] [Medline: 34104467]
12. Chantler T, Paton C, Velardo C, Triantafyllidis A, Shah SA, Stoppani E, et al. Creating connections - the development of a mobile-health monitoring system for heart failure: qualitative findings from a usability cohort study. Digit Health 2016 Oct 10;2:2055207616671461 [FREE Full text] [doi: 10.1177/2055207616671461] [Medline: 29942568]
13. Pal K, Dack C, Ross J, Michie S, May C, Stevenson F, et al. Digital health interventions for adults with type 2 diabetes: qualitative study of patient perspectives on diabetes self-management education and support. J Med Internet Res 2018 Jan 29;20(2):e40 [FREE Full text] [doi: 10.2196/jmir.8439] [Medline: 29463488]
14. Floch J, Zettl A, Fricke L, Weisser T, Grut L, Vilarinho T, et al. User needs in the development of a health app ecosystem for self-management of cystic fibrosis: user-centered development approach. JMIR Mhealth Uhealth 2018 May 08;6(5):e113 [FREE Full text] [doi: 10.2196/mhealth.8236] [Medline: 29739742]
15. Schimmer R, Orre C, Öberg U, Danielsson K, Hörnsten Å. Digital person-centered self-management support for people with type 2 diabetes: qualitative study exploring design challenges. JMIR Diabetes 2019 Sep 19;4(3):e10702 [FREE Full text] [doi: 10.2196/10702] [Medline: 31538941]

XSL•FO
RenderX

16. Giunti G, Guisado Fernández E, Dorronzoro Zubiete E, Rivera Romero O. Supply and demand in mHealth apps for persons with multiple sclerosis: systematic search in app stores and scoping literature review. JMIR Mhealth Uhealth 2018 May 23;6(5):e10512 [FREE Full text] [doi: 10.2196/10512] [Medline: 29792295]
17. Lupton D. Digital health now and in the future: findings from a participatory design stakeholder workshop. Digit Health 2017 Nov 9;3:2055207617740018 [FREE Full text] [doi: 10.1177/2055207617740018] [Medline: 29942616]
18. Woods L, Roehrer E, Duff J, Walker K, Cummings E. Co-design of a mobile health app for heart failure: perspectives from the team. Stud Health Technol Inform 2019 Aug 08;266:183-188. [doi: 10.3233/SHTI190792] [Medline: 31397321]
19. Fricker SA, Thümmler C, Gavras A. Requirements Engineering for Digital Health. Cham, Switzerland: Springer; 2015.
20. Mincolelli G, Imbesi S, Marchi M, Giacobone GA. New domestic healthcare. Co-designing assistive technologies for autonomous ageing at home. Design J 2019 May 31;22(sup1):503-516. [doi: 10.1080/14606925.2019.1595435]
21. Gradinar A, Davenport J, Hill H, Coulton P. Improving the visualisation of renal blood test results to enhance patient – clinician communication. Design J 2017 Sep 06;20(sup1):S2363-S2374. [doi: 10.1080/14606925.2017.1352751]
22. Jull J, Giles A, Graham ID. Community-based participatory research and integrated knowledge translation: advancing the co-creation of knowledge. Implement Sci 2017 Dec 19;12(1):150 [FREE Full text] [doi: 10.1186/s13012-017-0696-3] [Medline: 29258551]
23. Unsworth H, Dillon B, Collinson L, Powell H, Salmon M, Oladapo T, et al. The NICE Evidence Standards Framework for digital health and care technologies - developing and maintaining an innovative evidence framework with global impact. Digit Health 2021 Jun 24;7:20552076211018617 [FREE Full text] [doi: 10.1177/20552076211018617] [Medline: 34249371]
24. Jandoo T. WHO guidance for digital health: what it means for researchers. Digit Health 2020 Jan 8;6:2055207619898984 [FREE Full text] [doi: 10.1177/2055207619898984] [Medline: 31949918]
25. Greenhalgh T, Jackson C, Shaw S, Janamian T. Achieving research impact through co-creation in community-based health services: literature review and case study. Milbank Q 2016 Jun;94(2):392-429 [FREE Full text] [doi: 10.1111/1468-0009.12197] [Medline: 27265562]
26. Bird M, McGillion M, Chambers EM, Dix J, Fajardo CJ, Gilmour M, et al. A generative co-design framework for healthcare innovation: development and application of an end-user engagement framework. Res Involv Engagem 2021 Mar 01;7(1):12 [FREE Full text] [doi: 10.1186/s40900-021-00252-7] [Medline: 33648588]
27. Papoutsi C, Wherton J, Shaw S, Morrison C, Greenhalgh T. Putting the social back into sociotechnical: case studies of co-design in digital health. J Am Med Inform Assoc 2021 Feb 15;28(2):284-293 [FREE Full text] [doi: 10.1093/jamia/ocaa197] [Medline: 33043359]
28. Vo V, Auroy L, Sarradon-Eck A. Patients' perceptions of mHealth apps: meta-ethnographic review of qualitative studies. JMIR Mhealth Uhealth 2019 Jul 10;7(7):e13817 [FREE Full text] [doi: 10.2196/13817] [Medline: 31293246]
29. Bhattacharyya O, Mossman K, Gustafsson L, Schneider EC. Using human-centered design to build a digital health advisor for patients with complex needs: persona and prototype development. J Med Internet Res 2019 May 09;21(5):e10318 [FREE Full text] [doi: 10.2196/10318] [Medline: 31094334]
30. Personal health record (PHR) - user insights: final report. Royal College of Physicians. 2017 Jan. URL: https://www.rcplondon.ac.uk/file/5653/download [accessed 2012-12-21]
31. Information and Digital Technologies: Clinical Requirements 2020. Academy of Medical Royal Colleges. 2017 Feb. URL: https://www.aomrc.org.uk/wp-content/uploads/2017/03/IDT_Clinical_Requirements_2020_140317-2.pdf [accessed 2014-12-21]
32. What good looks like for digital records in adult social care. Care Quality Commission. 2020. URL: https://www.cqc.org.uk/guidance-providers/adult-social-care/what-good-looks-digital-records-adult-social-care [accessed 2021-08-28]
33. Personal Health Record user needs, accessibility and buy-in. National Health Service Digital. 2021. URL: https://tinyurl.com/yk7z7zsd [accessed 2021-08-28]
34. Weidberg E. Lessons from designing digital health for patients, with patients. UX Collective. 2020 Jan 2. URL: https://uxdesign.cc/insights-from-designing-digital-health-for-patients-with-patients-31d975f4b326 [accessed 2021-08-28]
35. Evashwick C. Creating the continuum of care. Health Matrix 1989;7(1):30-39. [Medline: 10293297]
36. Teal G, Thorup T, Baillie J, Johnson M, Crooks G. Digital diabetes dudes. Digital Health & Care Innovation Centre. 2018. URL: https://pureportal.strath.ac.uk/en/publications/digital-diabetes-dudes [accessed 2010-11-21]
37. Angoshtan M, Bradley J, Thorup T, Crooks G. Better Back. Digital Health & Care Innovation Centre. 2015. URL: https://pureportal.strath.ac.uk/en/publications/better-back [accessed 2012-11-21]
38. Teal G, French T, Bradley J, Crooks G. Backpack. Digital Health & Care Innovation Centre. 2016. URL: https://pureportal.strath.ac.uk/en/publications/backpack [accessed 2010-10-21]
39. Johnson M, Teal G, Thorup T, Ballie J. Digital Diabetes - Communication Between Communicators (CBC): Supporting Behaviour Change. The Digital Health & Care Innovation Centre. 2017. URL: http://radar.gsa.ac.uk/7583/ [accessed 2022-05-30]
40. Teal G, Thorup T, Baillie J, Johnson M, Crooks G. Digital Diabetes Dashboard. Digital Health & Care Innovation Centre. 2018. URL: https://pureportal.strath.ac.uk/en/publications/digital-diabetes-dashboard [accessed 2012-12-21]
41. Teal G, Thorup T, Baillie J, Johnson M, Crooks G. Digital Diabetes IDDEAS and GDS. Digital Health & Care Innovation Centre. 2018. URL: https://pureportal.strath.ac.uk/en/publications/digital-diabetes-iddeas-and-gds [accessed 2010-11-21]

XSL•FO
RenderX

42. Teal G, Baillie J, Johnson M, Thrrup T, Crooks G. Digital Diabetes. The Digital Health & Care Innovation Centre. 2018. URL: https://pureportal.strath.ac.uk/en/publications/digital-diabetes [accessed 2018-11-21]
43. Raman S, Blom J, Bradley J. Digital Empathy. Digital Health & Care Innovation Centre. 2016. URL: https://www.dhi-scotland.com/projects/digital-empathy/ [accessed 2010-11-21]
44. Raman S, French T, Crooks G. Game Jam : Co-designing a Game-based Learning Tool on Internet and Social Media Safety with Young People with Learning Disabilities. Digital Health & Care Innovation Centre. 2015. URL: https://strathprints.strath.ac.uk/65083/ [accessed 2010-10-21]
45. Bradley J, Johnson M, Crooks G. Glimpse - Movement for Wellbeing in the Workplace. Digital Health & Care Innovation Centre. 2017. URL: https://pureportal.strath.ac.uk/en/publications/glimpse-movement-for-wellbeing-in-the-workplace [accessed 2018-10-21]
46. Brooks E, Stengs G. Community Mental Health. The Digital Health & Care Innovation Centre. 2020. URL: http://radar.gsa.ac.uk/7296/ [accessed 2022-05-30]
47. Brooks E. Insights: Type 2 Diabetes. The Digital Health & Care Innovation Centre. 2020. URL: https://www.dhi-scotland.com/projects/type-2-diabetes-transforming-the-diagnosis-conversation/ [accessed 2012-11-21]
48. Hepburn LA, McIntyre D. Exploring the Now to Design for Next: Future Cancer Care Services. The Digital Health & Care Innovation Centre. 2019. URL: http://radar.gsa.ac.uk/7584/ [accessed 2022-05-30]
49. Teal G, Rice G. Melanoma. Glasgow School of Art. 2013. URL: http://radar.gsa.ac.uk/6252/ [accessed 2022-05-30]
50. Raman S, Angela T. Ritual Respect: Co-designing Care and Emotional Support around Miscarriage. The Digital Health & Care Innovation Centre. 2018. URL: https://pureportal.strath.ac.uk/en/publications/ritual-respect-co-designing-care-and-emotional-support-around-mis [accessed 2008-11-21]
51. Teal G, Crooks G. Modern Outpatient Care. The Digital Health & Care Innovation Centre. 2018. URL: https://strathprints.strath.ac.uk/66130/ [accessed 2004-12-21]
52. French T, Raman S. Virtual Hospice – Executive Summary. The Digital Health & Care Innovation Centre. 2015. URL: https://strathprints.strath.ac.uk/66133/1/French_Raman_DHI_2015_Virtual_Hospice.pdf [accessed 2004-12-21]
53. Angoshtan M, Thorup T, Baillie J, Crooks G. Well Connected Blood Donation. The Digital Health & Care Innovation Centre. 2016. URL: https://pureportal.strath.ac.uk/en/publications/well-connected-blood-donation [accessed 2012-11-21]
54. Chute C, Hepburn LA, Rooney L. Next Generation Asthma Care: Position Paper. The Digital Health & Care Innovation Centre. 2019. URL: https://pureportal.strath.ac.uk/en/publications/next-generation-asthma-care-position-paper [accessed 2012-11-21]
55. Raman S, Teal G. Transforming Conversation About Type 2 Diabetes. The Digital Health & Care Innovation Centre. 2019 Dec. URL: https://static1.squarespace.com/static/5ab504e77c9327e5eed2778a/t/5e2accb5db08282f9aa5657e/1579863239655/lay+summary_A5_web+and+print+version.pdf [accessed 2011-12-21]
56. Rimpilainen S. Revolutionising the outpatient care two day exploratory. The Digital Health & Care Innovation Centre. 2015. URL: https://strathprints.strath.ac.uk/65958/1/Rimpilainen_DHCI2015_Revolutionising_the_outpatient_care_two_day_exploratory.pdf [accessed 2010-10-21]
57. Bradley J, Tulloch A, Crooks G. Nursing Records and Open Innovation. The Digital Health & Care Innovation Centre. 2017. URL: https://pureportal.strath.ac.uk/en/publications/nursing-records-and-open-innovation [accessed 2022-05-30]
58. Hepburn LA, Tulloch A, French T, Crooks G. Crossreach Confidential Connections. The Digital Health & Care Innovation Centre. 2017. URL: https://pureportal.strath.ac.uk/en/publications/experience-lab-crossreach [accessed 2001-11-21]
59. Hepburn LA, Jaatun E, Crooks G. Language of Pain : Merging Multiple Voices for Improved Chronic Pain Management. The Digital Health & Care Innovation Centre. 2018. URL: https://strathprints.strath.ac.uk/66142/ [accessed 2002-12-21]
60. French T, Teal G, Sneha R, Blom J. Notification System. The Digital Health & Care Innovation Centre. 2014. URL: http://radar.gsa.ac.uk/6248/ [accessed 2022-05-30]
61. French T, Hepburn LA. Breast Feeding. The Digital Health & Care Innovation Centre. 2016. URL: https://www.dhi-scotland.com/projects/breastfeeding/ [accessed 2014-11-21]
62. Van Labeke N, Blom J, Crooks G. Digital Brokering: Project Report. The Digital Health & Care Innovation Centre. 2014. URL: https://pureportal.strath.ac.uk/en/publications/digital-brokering-project-report [accessed 2006-12-21]
63. Blom J, Crooks G. LED-based indoor tracking of people living with dementia. The Digital Health & Care Innovation Centre. 2015. URL: https://pureportal.strath.ac.uk/en/publications/led [accessed 2017-11-21]
64. Ranade-Kharkar P, Weir C, Norlin C, Collins SA, Scarton LA, Baker GB, et al. Information needs of physicians, care coordinators, and families to support care coordination of children and youth with special health care needs (CYSHCN). J Am Med Inform Assoc 2017 Sep 01;24(5):933-941 [FREE Full text] [doi: 10.1093/jamia/ocx023] [Medline: 28371887]
65. Holt JM, Cusatis R, Winn A, Asan O, Spanbauer C, Williams JS, et al. Impact of pre-visit contextual data collection on patient-physician communication and patient activation: a randomized trial. J Gen Intern Med 2021 Nov;36(11):3321-3329. [doi: 10.1007/s11606-020-06583-7] [Medline: 33559067]
66. Taylor A, Lowe DJ, McDowell G, Lua S, Burns S, McGinness P, et al. Remote-management of COPD: evaluating the implementation of digital innovation to enable routine care (RECEIVER): the protocol for a feasibility and service adoption observational cohort study. BMJ Open Respir Res 2021 Aug;8(1):e000905 [FREE Full text] [doi: 10.1136/bmjresp-2021-000905] [Medline: 34462271]

XSL•FO
RenderX

Abbreviations

DHI: Digital Health & Care Innovation Centre
FHIR: Fast Healthcare Interoperability Resources
mHealth: mobile health
NHS: National Health Service

Edited by A Mavragani; submitted 01.12.21; peer-reviewed by W Chapman, H Ayatollahi, H Mehdizadeh; comments to author 21.12.21; revised version received 02.03.22; accepted 08.05.22; published 10.06.22.

Please cite as:
Chute C, French T, Raman S, Bradley J
User Requirements for Comanaged Digital Health and Care: Review
J Med Internet Res 2022;24(6):e35337
URL: https://www.jmir.org/2022/6/e35337
doi:10.2196/35337
PMID:35687379

©Chaloner Chute, Tara French, Sneha Raman, Jay Bradley. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.06.2022. 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

Technology-Based Compensation Assessment and Detection of Upper Extremity Activities of Stroke Survivors: Systematic Review

Xiaoyi Wang[1], MSc; Yan Fu[1], PhD; Bing Ye[2,3], MSc; Jessica Babineau[4], MLiS; Yong Ding[5], MSc; Alex Mihailidis[2,3], PhD, PEng

[1]School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan, China

[2]KITE - Toronto Rehabilitation Institute, University Health Network, Toronto, ON, Canada

[3]Department of Occupational Science and Occupational Therapy, University of Toronto, Toronto, ON, Canada

[4]Library and Information Services, University Health Network, Toronto, ON, Canada

[5]Department of Rehabilitation Medicine, Hubei Provincial Hospital of Traditional Chinese Medicine, Wuhan, China

Corresponding Author:
Yan Fu, PhD
School of Mechanical Science and Engineering, Huazhong University of Science and Technology
Room 519, East 8th Building, Huazhong University of Science and Technology
Wuhan
China
Phone: 86 15071320900
Email: laura_fy@mail.hust.edu.cn

Abstract

Background: Upper extremity (UE) impairment affects up to 80% of stroke survivors and accounts for most of the rehabilitation after discharge from the hospital release. Compensation, commonly used by stroke survivors during UE rehabilitation, is applied to adapt to the loss of motor function and may impede the rehabilitation process in the long term and lead to new orthopedic problems. Intensive monitoring of compensatory movements is critical for improving the functional outcomes during rehabilitation.

Objective: This review analyzes how technology-based methods have been applied to assess and detect compensation during stroke UE rehabilitation.

Methods: We conducted a wide database search. All studies were independently screened by 2 reviewers (XW and YF), with a third reviewer (BY) involved in resolving discrepancies. The final included studies were rated according to their level of clinical evidence based on their correlation with clinical scales (with the same tasks or the same evaluation criteria). One reviewer (XW) extracted data on publication, demographic information, compensation types, sensors used for compensation assessment, compensation measurements, and statistical or artificial intelligence methods. Accuracy was checked by another reviewer (YF). Four research questions were presented. For each question, the data were synthesized and tabulated, and a descriptive summary of the findings was provided. The data were synthesized and tabulated based on each research question.

Results: A total of 72 studies were included in this review. In all, 2 types of compensation were identified: disuse of the affected upper limb and awkward use of the affected upper limb to adjust for limited strength, mobility, and motor control. Various models and quantitative measurements have been proposed to characterize compensation. Body-worn technology (25/72, 35% studies) was the most used sensor technology to assess compensation, followed by marker-based motion capture system (24/72, 33% studies) and marker-free vision sensor technology (16/72, 22% studies). Most studies (56/72, 78% studies) used statistical methods for compensation assessment, whereas heterogeneous machine learning algorithms (15/72, 21% studies) were also applied for automatic detection of compensatory movements and postures.

Conclusions: This systematic review provides insights for future research on technology-based compensation assessment and detection in stroke UE rehabilitation. Technology-based compensation assessment and detection have the capacity to augment rehabilitation independent of the constant care of therapists. The drawbacks of each sensor in compensation assessment and detection are discussed, and future research could focus on methods to overcome these disadvantages. It is advised that open data together with multilabel classification algorithms or deep learning algorithms could benefit from automatic real time compensation detection. It is also recommended that technology-based compensation predictions be explored.

(J Med Internet Res 2022;24(6):e34307) doi:10.2196/34307

XSL•FO
RenderX

KEYWORDS

stroke; upper extremity rehabilitation; UE rehabilitation; compensation; assessment; technology; sensor; artificial intelligence; AI

Introduction

Background

Stroke occurs almost every 2 seconds worldwide, affecting 13.7 million people each year [1]. Approximately 80% of stroke survivors are affected by upper extremity (UE) motor impairment, and 50% have UE motor dysfunction even 4 years after stroke [2]. Poststroke UE rehabilitation plays an important role in UE motor function recovery. Current research has shown that 2 competing mechanisms may occur simultaneously during the UE function recovery process: motor recovery and compensation. Motor recovery is defined as the "reappearance of elemental motor patterns presents prior to central nervous system injury," whereas compensation is defined as "the appearance of new motor patterns resulting from the adaptation of remaining motor elements or substitution" [3]. Common compensatory strategies include excessive trunk displacement during reaching movement [3]. Recent research argues that the frequent use of compensation may lead to long-term chronic pain in overused joints, limited function in the impaired muscles, suboptimal motor recovery in the impaired arm, and an abnormal UE movement pattern in activities of daily living [3-5]. Therefore, timely detection and appropriate correction of compensation are important The mechanism underlying UE rehabilitation is neuroplasticity, which refers to the rewiring or reorganization of the brain by creating new connections between brain cells after a stroke [6]. More specifically, to realize brain plasticity, extensive, intensive, task-oriented UE movement repetitions must be performed [7]. Traditionally, UE rehabilitation is completed in a hospital under the supervision of a therapist, in which case some compensatory behaviors can be avoided or corrected under the guidance of the therapist [8]. However, not all compensation can be observed in a timely manner by therapists [9]. Moreover, the UE rehabilitation protocol is labor-intensive for therapists, and there are not enough skilled therapists to support such huge demands [10]. Technology-based therapies, such as robot-assisted therapy and virtual reality (VR) therapy [11], have been used to facilitate UE rehabilitation after stroke in recent years. However, an important prerequisite for taking full advantage of these technology-based therapies is that stroke survivors can correctly perform the therapy exercises as intended, which means that compensation should be automatically detected and corrected in these therapy systems [12]. Technologies could provide more fine-grained automatic compensation monitoring in less-supervised UE therapies so that stroke survivors could continue with the required exercises independent of therapists. Despite the recent increase in attention given to technology for automatic compensation assessment and detection, no systematic reviews have been conducted in this area.

Objectives

The main goal of this review was to explore how technology-based methods were used to assess and detect compensation without the constant care of therapists.

Our research questions (RQs) are as follows:

1. What models are used to assess and detect compensation in poststroke UE activities?
2. What measurements are used to evaluate compensatory movements?
3. What types of sensor technology are used for compensation assessment and detection?
4. Which statistical or artificial intelligence (AI) methods are used for compensation assessment and detection?

Methods

The systematic review was performed according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Multimedia Appendix 1).

Information Sources and Search Strategy

A comprehensive search strategy was developed and executed by an information specialist (JB). The search strategy was originally developed in MEDLINE ALL (Ovid), in consultation with the research team. The search results were then translated into other databases and study registries. The following electronic databases were searched: MEDLINE (R) ALL (Ovid), Embase and Embase Classic (Ovid), Cochrane Central Register of Controlled Trials (CENTRAL, Ovid), Health Technology Assessment (Ovid), SPORTDiscus (EBSCO), Scopus, Compendex (Engineering Village), INSPEC (Engineering Village), IEEE Xplore, and ACM Digital Library. Dissertations and Theses Global (ProQuest) were searched to identify dissertations or theses. The study registries searched were ClinicalTrials.gov and World Health Organization International Clinical Trials Registry Platform.

Search strategies included the use of text words and subject headings (eg, Medical Subject Headings and Emtree) related to five concepts: (1) stroke, (2) rehabilitation, (3) UE, (4) compensation, and (5) robotics or technology. The search was limited to English. Cochrane search filters were applied to exclude animal-only studies when possible [13]. All databases and registers were searched from the inception of resources. Searches were conducted on May 26, 2020. Searches were updated by rerunning all search strategies on July 23, 2021, and exporting new results. The full search strategies for each database and registry are provided in Multimedia Appendix 2.

Study Selection

All search results were first imported into EndNote software, where duplicates were removed. The remaining results were imported into Covidence. A total of 2 screening steps were conducted: title and abstract screening and full-text screening. In all, 2 researchers (XW and YF) independently conducted

XSL•FO
RenderX

title and abstract screening as well as full-text screening using the same inclusion and exclusion criteria. Disagreements between the 2 researchers were discussed and resolved between the 2 researchers. A third researcher (BY) was involved when an agreement could not be reached.

The inclusion and exclusion criteria used for the screening process are presented in Textbox 1.

Textbox 1. The inclusion and exclusion criteria used for the screening process.

Inclusion criteria

- Stroke survivors or healthy participants enrolled in the intervention.
- The study involves upper extremity rehabilitation.
- Compensation was assessed using technology (ie, information and communication technologies, sensors, cameras, wearables, or artificial intelligence).
- The study involves compensation assessment or detection.
- The study involves compensation measurements: kinematic parameters (speed, angle, angular speed, etc), electromyogram, or compensatory posture or pattern classification.

Exclusion criteria

- Studies involving nonhuman participants.
- Studies about stroke neural recovery, stroke prevalence, and pathological analysis.
- Studies that do not use technology-based measurement methods.
- Studies on activity logs, functional electrical stimulation, gravity compensation in robotics, and effects of virtual therapy.
- Studies are not about upper extremity rehabilitation.
- Qualitative, usability, or nonacademic studies.
- Review studies such as systematic reviews.
- Case reports and letters.
- Studies are not written in English.

After the screening stage, studies were rated for their level of evidence based on the Centre for Evidence-Based Medicine (CEBM) [14] criteria. According to the CEBM, 4 clinical scales were used as reference standards, which included the compensation assessment scale—the Reaching Performance Scale [15], Motor Activity Log [16], Actual Amount of Use Test [17], and Chedoke-McMaster Stroke Assessment [18]. We used CEBM criterion 2b as a reference. The study would be regarded as having good reference standards if it had the same training task from any of the aforementioned 4 scales or if it had the same or partially the same evaluation criteria as any of the 4 scales.

Results

Overview

A total of 1584 records were retrieved from the search. After removing duplicates, 69.51% (1101/1584) of records were screened at the title and abstract stage. In the first stage, 84.29% (928/1584) of the records were removed. The remaining 15.71% (173/1584) articles were subjected to full-text screening. A total of 76 studies were included after both screening stages. Figure 1 shows the PRISMA [19] flow diagram. After studies were rated based on the CEBM criteria, 72 (range from 1b to 2b in CEBM criteria) of the 76 (95%) studies were included in the final analysis; Table 1 shows the relationships between the included studies and the reference standards. In all, 67 papers were published after 2010, 69% (50/72) of which were published between 2015 and 2021.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram illustrating the screening process for papers included in this study.

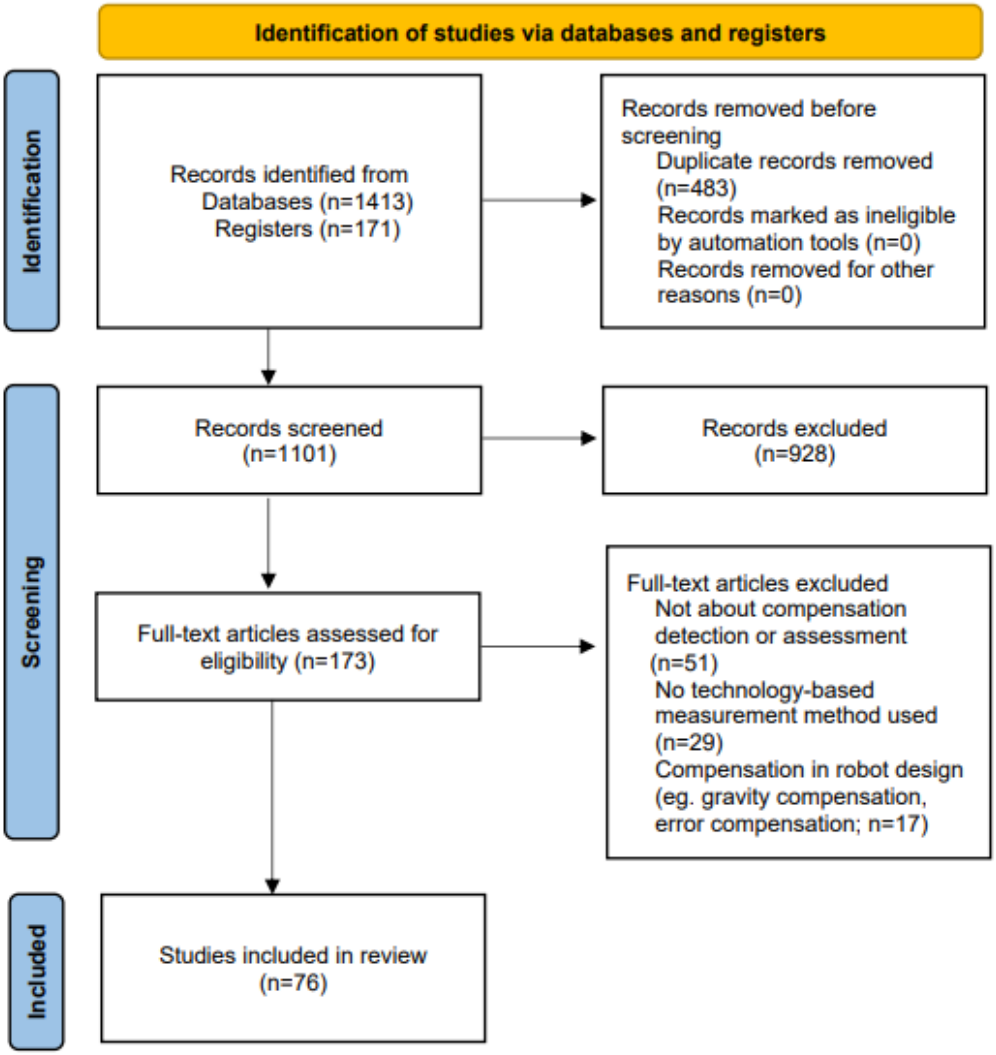


Table 1. Correlation with reference standards.

Reference standard	Correlated references	Example
Reaching Performance Scale	[20-58]	[23]; task: reaching tasks; evaluation criteria: trunk displacement
Chedoke-McMaster Stroke Assessment and Reaching Performance Scale	[33,59-80]	[69]; task: a set of upper extremity exercises (Chedoke-McMaster Stroke Assessment); evaluation criteria: trunk displacement and shoulder movements (Reaching Performance Scale)
Motor Activity Log or Actual Amount of Use Test	[81-89]	[82]; task: activities of daily living; evaluation criteria: arm use

Study Characteristics

Of the 72 studies, 38 (53%) recruited only stroke survivors, 9 (13%) included only healthy participants, and the remaining studies (n=25, 34%) recruited both (Table 2). Both men and women were included in most (48/72, 67%) studies. The age range of stroke survivors was 21 to 92 years and that of healthy participants was 18 to 85 years. For stroke survivors, the stage of recovery included subacute (between 1 and 6 months after stroke; 4 studies), chronic (>6 months after stroke; 36 studies), or both (18 studies). The sample size varied from 1 to 119 (Table 3).

XSL•FO
RenderX

Table 2. Characteristics of the studies (N=72).

	References
Participants	
Stroke survivors	[21,23,25,26,28-30,32-34,38,40,44-47,51,52,54,56,58,60,62,65,67,69,71,73,74,76,77,79,80,82,85,86,90]
Healthy participants	[20,31,39,42,43,55,59,61,78]
Stroke survivors and healthy participants	[22,24,27,35-37,41,48-50,53,57,63,64,66,68,70,72,75,81,83,84,87-89]
Stage of recovery	
Subacute	[35,73,82,89]
Chronic	[21,23,26-30,32-34,38,40,41,44,46,47,51,52,56,58,60,62,65,68,69,71,74,75,79,80,83-85,87,88,90]
Subacute and chronic	[22,24,25,36,37,45,50,53,54,57,63,64,66,67,72,77,86,91]

Table 3. The sample size distribution (N=72).

Sample size	Studies, n (%)
0-18	46 (64)
19-36	13 (18)
37-54	9 (13)
55-72	2 (3)
73-90	0 (0)
91-108	1 (1)
109-126	1 (1)

RQ1: What Models Have Been Established to Assess and Detect Compensation?

Types of Compensation

Overview

Two types of compensation were identified according to the study by Miller et al [81]: (1) *disuse of the affected UE*, and (2) *use of the affected UE in an awkward manner to adjust for limited strength, mobility, and motor control.* We refer to the second type of compensation as awkward use of the affected UE for the remainder of this paper. Table 4 presents the compensation types, models, and measurements.

XSL•FO
RenderX

Table 4. Compensation type, model, and measurements.

Compensation type, model, and measurements		References
Disuse of the affected upper limb		
Arm use		
	The ratio between the duration of movement in the least and less affected arm	[82]
	Mean squared sum of the acceleration over a 1-minute epoch of the arm	[83,85]
	Torques due to the measured tangential forces on the split-steering wheel	[84]
Arm nonuse		
	The difference of the Euclidean distance between the trunk and hand to the target	[86]
	Movement time, peak velocity, total displacement, and movement smoothness	[87]
	Root mean square of the rotation angle of the steering wheel	[88]
	Total movement duration of each limb and the ratio between the movement duration in the paretic and nonparetic limb	[89]
Interlimb coordination	Amplitude, time, and frequency data from inertial sensors on upper body	[81]
Awkward use of the affected UE[a]		
Trunk compensation		
	Trunk movements in the sagittal plane: trunk lean forward, trunk displacement, trunk flexion, trunk anteriorization, and trunk lean backward	
	Trunk angular displacement	[20,22,23,25-27,39,47,60,63,68,90]
	Trunk linear displacement	[24,30,40,41,51,52,58,66]
	Trunk contribution slope	[37,38,62]
	Acceleration of trunk motion	[28,64]
	sEMG[b] signal	[39,77]
	Face orientation	[27]
	Measurements for AI[c]-based compensatory posture classification	[43-45,48-50,53,54,61,74,77,91]
	N/A[d]	[56,65,69]
	Trunk movements in the transverse plane: trunk rotation and trunk twist	
	Trunk angular displacement	[22,25,26,39,47,68,90]
	Acceleration of trunk motion	[27,28,64]
	Trunk linear displacement	[34,40]
	sEMG signal	[39]
	Measurements for AI-based compensatory posture classification	[45,48-50,53,54,61,74,77,91]
	N/A	[65,69]
	Trunk movements in the coronal plane: trunk leans from side to side, trunk contralateral and ipsilateral flexion, trunk lateral bending, and trunk lateral shift	
	Trunk angular displacement	[21,22,47,60,68,90]
	Trunk linear displacement	[34,41]
	Measurements for AI-based compensatory posture classification	[61]
	Unspecified	
	Trunk movement time, trunk distance, trunk peak velocity, and maximal angle of trunk flexion	[46]
	Position and angle	[42,59]
Shoulder compensation		
	Shoulder elevation	

XSL•FO
RenderX

Compensation type, model, and measurements			References
		Elevation angle of scapula, acromion, or acromio-clavicular joint	[26,39,57,66]
		Acceleration of shoulder joint motion	[27,28]
		Shoulder vertical translated distance	[37]
		sEMG signal	[39]
		Measurements for AI-based compensatory posture classification	[45,48-50,53,54,74,77,91]
		N/A	[69]
	Shoulder abduction		
		Acceleration of shoulder joint motion	[27,28,64]
		Shoulder abduction angle	[70]
		fMRI[e]	[71]
	Shoulder girdle compensatory movements		
		Acceleration of shoulder joint motion	[29,64]
		sEMG signal	[30]
		Shoulder position	[31]
		The coefficient of the elbow joint extension to the shoulder joint flexion ratio	[72]
	Shoulder forward	Shoulder forward liner displacement	[26,32]
	Shoulder overflexion	Shoulder flexion angle	[22,33,73]
	Unspecified	Shoulder position	[59]
Elbow compensation: insufficient elbow extension			
	Elbow extension angle		[66,74,90]
	Acceleration of elbow joint motion		[27,28]
	N/A		[65]
Finger compensation			
	Individual finger compensation	Finger extension angle	[34]
	Multiple fingers adaptive compensation		
		The covariance of individual finger impulses across multiple pulses	[75]
		Pressure force of fingers	[76]
Joint coordination			
	Scapula, shoulder, elbow and wrist joint angles, movement time, goal-equivalent variance, non-goal-equivalent variance		[36]
	Joint angles		[35]
Muscle synergy	sEMG signal		[33]
Slouching	Joint position		[69]

[a]UE: upper extremity.

[b]sEMG: surface electromyogram.

[c]AI: artificial intelligence.

[d]N/A: not applicable.

[e]fMRI: functional magnetic resonance imaging.

XSL•FO
RenderX

Disuse of the Affected UE

A total of 9 studies assessed this type of compensation, and 3 models were discussed: the arm use model [82-85], arm nonuse model [86-89] and interlimb coordination model [81]. The arm use model measured the actual use of the impaired arm either in activities of daily living [82,83,85] or in bilateral and unilateral steering tasks [84]. The arm nonuse model was used to quantify the difference between the actual use of the impaired arm and its performance measured using standard clinical scales in reaching tasks [86,87], bilateral and unilateral steering tasks [88], and occupational therapy [89]. The interlimb coordination model was used to detect the reduction in interlimb coordination in stroke survivors compared with healthy participants in unimanual and bimanual activities of daily living [81]. In 4 studies [84,87-89], the tasks were completed using robot-assisted devices.

Awkward Use of the Affected UE

Most (63/72, 88%) studies assessed this type of compensation. The main models were (1) trunk compensation, (2) shoulder compensation, (3) elbow compensation, (4) finger compensation, and (5) others.

Trunk Compensation Model

This model (46/63, 73% studies) measures the awkward movements of the trunk for the affected UE [15]. Trunk compensatory movements can occur in 3 anatomical planes (sagittal, transverse, and coronal) of the human body. The sagittal plane (41/46, 89% studies) was the most common, which was described as trunk lean forward, trunk lean backward, trunk displacement, trunk flexion, and trunk anteriorization. The transverse plane (24/46, 52% studies) included trunk rotation and trunk twist. A total of 9 (20%) studies discussed the coronal plane, including trunk leans from side to side, trunk contralateral and ipsilateral flexion, trunk lateral bending, and trunk lateral shift (Table 3). The most common task (35/46, 76%) was the reaching task, followed by shoulder, elbow, and wrist exercises (4/46, 9%) [20,59-61], daily life activities [62-65], drinking tasks [66], simulated therapy activities [67], instrumented trunk impairment scale (version 2) tasks [21], Fugl–Meyer Assessment (FMA) items [22], occupational therapy tasks [68], and the Graded Repetitive Arm Supplementary Program (GRASP) [69], which is a set of UE exercises completed without the presence of a therapist. In 13 (28%) studies, tasks were completed using a robot-assisted device. In 4 (9%) studies, the tasks were conducted using VR [23-25] and mixed reality (MR) training systems [26].

Shoulder Compensation Model

This model (29/63, 46% studies) measures awkward movements of the shoulder of the affected UE [15], involving complex movements of the shoulder girdle and shoulder joint. The most observed shoulder compensation was shoulder elevation (17/29, 59% studies), followed by shoulder abduction [27,28,64,70,71], shoulder girdle compensatory movement [29-31,64,72], shoulder forward (protraction) [26,32], and shoulder overflexion [22,33,73]. The most commonly used task was reaching task (18/29, 62% studies). Other tasks involved hand-to-mouth tasks [33,70,73], drinking tasks [66,72], elbow flexion-extension task [59,71], daily life activities [64,65], counterclockwise cyclic motions [31], FMA items [22], and GRASP [69]. In all, 12 studies were conducted using a robot-assisted device and 1 with an MR training system [26].

Elbow Compensation Model

This model measures awkward elbow movements of the affected UE [15]. A total of 6 studies found that stroke survivors had insufficient elbow extension during reaching tasks [27,28,74,90], drinking tasks [66], or daily life activities [65].

Finger Compensation Model

This model (3/63, 5% studies) measures the compensation among finger joints [34,75,76]. A study assessed the compensation among the joints in a finger in reaching tasks [34], whereas 2 other studies assessed compensation among multiple fingers in repetitive force-pulse tasks [75] and index finger movements [76].

Other Types of Compensation Models

Other types of compensation models included joint coordination [35,36], slouching [69], and muscle synergies [33], which were measured in reaching tasks, GRASP, and hand-to-mouth tasks, respectively.

RQ2: What Measurements Are Used to Evaluate Compensatory Movements?

Disuse of the Affected UE

No standard measurement has been applied across studies for this type of compensation. For the arm use model, Ballester et al [83] and Hung et al [85] computed the mean squared sum of the acceleration over 1 minute. Thrane et al [82] calculated the arm movement ratio, that is, the ratio of arm use duration between the impaired arm and less impaired arm. Johnson et al [84] quantified the arm use by comparing the torque generated by the tangential forces of the 2 arms on the steering wheel.

For the arm nonuse model, Bakhti et al [86] computed proximal arm nonuse, which was the difference between the Euclidean distance between the trunk and hand to the target during the reaching movement in both spontaneous and maximal proximal arm use conditions. Johnson et al [87] used 4 kinematic metrics, including movement time, peak velocity, total displacement, and movement smoothness, to predict learned nonuse (LNU). Johnson et al [88] compared the root mean square of the rotation angle of the wheel in steering tasks in 3 different steering modes (unilateral nondominant, unilateral dominant, and bilateral) to quantify LNU. Barth et al [89] computed the total movement duration of each limb and the activity ratio, which was the movement duration of the paretic limb to the nonparetic limb to assess LNU.

Miller et al [81] created an array of numerical values, including amplitude, time, and frequency data from acceleration signals on the sternum, right wrist, left wrist, right upper limb, and left upper limb to characterize the interlimb coordination model.

Awkward Use of the Affected UE

The parameters for measuring trunk compensation in the sagittal plane included trunk angular displacement (12/41, 29% studies),

XSL•FO
RenderX

trunk linear displacement (8/41, 20% studies), trunk contribution slope [37,38,62], acceleration of trunk motion [28,64], surface electromyogram (sEMG) signals [39,77], and face orientation [67]. The parameters used to measure trunk compensation in the transverse plane included trunk angular displacement (7/24, 29% studies), acceleration of trunk motion [27,28,64], trunk linear displacement [34,40], and sEMG signal [39]. A total of 2 parameters, trunk angular displacement (6/9, 67% studies) and trunk linear displacement [34,41] were measured to assess trunk compensation in the coronal plane.

The shoulder elevation compensation measurements included the elevation angle of the scapula, acromion, or acromioclavicular joint (4/17, 24% studies); acceleration of shoulder joint motion [27,28]; shoulder vertical translated distance [37]; and sEMG signal [39]. Shoulder abduction compensation was assessed by acceleration of shoulder joint motion [27,28,64], shoulder abduction angle [70], and functional magnetic resonance imaging (fMRI) [71]. Shoulder girdle compensatory movement measurements included acceleration of shoulder joint motion [29,64], sEMG signals [30], shoulder position [31], and the coefficient of the elbow joint extension to the shoulder joint flexion ratio [72]. A total of 3 studies [22,33,73] used the shoulder flexion angle to assess the shoulder overflexion compensation. The parameter for measuring shoulder forward compensation was shoulder forward liner displacement [26,32].

The elbow extension angle [66,74,90] and acceleration of elbow joint motion [27,28] were used to measure elbow compensation.

In all, 3 kinds of measurements were used to assess finger compensation. Fluet et al [34] measured the finger extension angle to assess the individual finger compensation. Kim et al [75] measured the covariance of individual finger impulses across multiple pulses, and Furudate et al [76] measured the pressure force of fingers to assess the compensation among multiple fingers.

As for other compensation models, Reisman and Scholz [36] measured multiple parameters including joint angles (ie, scapula, shoulder, elbow, and wrist), movement time, goal-equivalent variance, and nongoal-equivalent variance to evaluate joint coordination; Nibras et al [35] measured only joint angles to assess joint coordination. Lin et al [69] captured joint positions to assess slouching. Belfatto et al [33] measured sEMG signals to assess the muscle synergy.

RQ3: What Types of Sensor Technology Are Used for Compensation Assessment and Detection?

Overview

A total of 6 types of sensors were identified as shown in Tables 5 and 6.

Table 5. Studies classified by sensor type (N=72).

Sensor type	Studies, n (%)
Body-worn sensor technology	25 (35)
Marker-based motion capture system	24 (33)
Marker-free vision sensor	16 (22)
Physiological signal sensing technology	10 (14)
Sensors embedded in rehabilitation training system	8 (11)
Ambient sensor	5 (7)

XSL•FO
RenderX

Table 6. Studies classified by sensor type (N=72).

Sensor type		Sensor measurement	Application settings		References
			Technology-based therapy setting	Home setting	
Body-worn sensor					
	Accelerometer	Acceleration of upper limb segments and trunk	[89]	[60,64,65,83]	[27,28,60,64,65,82,83,85,89]
	IMU[a]	Original IMU signals or Euler angles of upper limb segments and trunk	[25,31,42,59,70]	[25,43,44,81]	[21,22,25,31,42-44,59,68,70,73,81,90]
	Strain sensors	Electrical resistance of sensors printed on the stretched parts	N/A[b]	[78,79]	[78,79]
	CyberGlove	Finger angles	N/A	N/A	[34]
Marker-based motion capture system					
	Optical motion capture system	3D coordinates of the markers placed on the upper body	[24,26,31-33,38,39,42,45-47,59,62,63,87]	N/A	[24,26,31-33,36,38-40,42,43,45-47,57-59,62,63,66,72,87]
	Electromagnetic motion capture system	3D coordinates of the markers placed on the upper body	N/A	N/A	[34]
	Ultrasound 3D motion capture system	3D coordinates of the markers placed on the upper body	N/A	N/A	[86]
Marker-free vision sensor					
	Microsoft Kinect depth sensor	Upper body joint positions in 3D space (x-y-z) coordinates	[23,41,48-52]	N/A	[20,23,41,48-52,61,69,86]
	Simple camera	Video	[74,84]	[67]	[27,28,67,69,74,84]
Physiological signal sensing technology					
	EMG[c]	sEMG[d] signals of upper limb and trunk muscles	[29,30,33,39,54,84]	N/A	[29,30,33,39,53,54,77,84]
	EEG[e]	EEG signals	[33]	N/A	[33,80]
	fMRI[f]	fMRI images	N/A	N/A	[72]
Sensors embedded in the training system	Force sensor or piezoelectric sensor or others	Force exerted by upper limbs, finger force, upper limb joint position, or orientation	[31,35,42,59,75,76,84,88]	N/A	[31,35,42,59,75,76,84,88]
Ambient sensor					
	Pressure distribution mattress	Force distribution	[45,54,91]	[45,54,91]	[45,54,91]
	Position sensor	Upper limb and trunk position	N/A	[55]	[55,56]

[a]IMU: inertial measurement unit.

[b]N/A: not applicable.

[c]EMG: electromyogram.

[d]sEMG: surface electromyogram.

[e]EEG: electroencephalogram.

XSL•FO
RenderX

[f]fMRI: functional magnetic resonance imaging.

Body-Worn Sensor Technology

Body-worn sensors (25/72, 35% studies) were the most commonly used technology to detect compensatory movements, including accelerometers, inertial measurement units (IMUs), strain sensors, and CyberGlove. In all, 9 studies used accelerometers, including uniaxial [82] and triaxial [27,28,60,64,65,83,85,89]. Accelerometers were attached to different parts of the body. Some were worn on the wrists of both arms [82,83,89] or only on the wrist of the affected arm [85] to measure arm movement quantity, such as movement duration and acceleration magnitude, to evaluate arm use. Some were placed on the trunk (chest, middle back, or T12 vertebrae) [27,28,60,64], shoulder [27,28,64], elbow [27,28], and wrist [65] to measure time and movement variables, such as accelerations and joint angles of UE movement to detect trunk, shoulder, and elbow compensation. Among these studies, Antonio et al [27] and Carla et al [28] compared quantitative detection results using accelerometers with therapist-based visual analysis of video records. The results showed that the compensatory movements detected by the accelerometers, including shoulder abduction and elevation, insufficient elbow extension, and trunk forward displacement and rotation, were consistent with the therapists' observations.

In all, 13 used IMUs. Each IMU typically consists of 1 or 2 triaxial accelerometers, a triaxial gyroscope, and a triaxial magnetometer [21,22,25,31,42,59,68,70,73,81]. The magnetometer was not included in some cases [43,44,90]. Accordingly, each IMU yielded 3D measurements of acceleration, angular velocity, and magnetic field vector (when using a magnetometer) in its intrinsic coordinate system [59]. In the reviewed studies, 1 to 9 IMUs were placed on the upper body parts, including the sternum [21,25,43,44,68,81,90], spine [21,22], pelvis [22], scapula [70,90], upper arms [22,25,31,42-44,59,70,73,81,90], forearms [22,25,43,44,59,70,90], wrists [73,81], and hands [22,70,90]. The original IMU signals [43,44,81] representing the movements of these body segments or the orientation in the form of Euler angles [21,22,25,31,42,59,68,70,73,90], of these body segments were output for compensation monitoring. It has been reported that IMUs can be used to detect trunk [21,22,25,42-44,59,68,90], shoulder [22,31,59,70,73], and elbow [22] compensation, as well as the interlimb coordination [81]. Furthermore, Ranganathan et al [43] proved that using IMUs could effectively detect compensatory trunk movements (approximately 90% accuracy) when compared with using an 8-camera motion capture system (Motion Analysis Corporation) as ground truth.

Moreover, 2 studies used changes in the electrical resistance of strain sensors printed on a garment [78,79] to identify different compensatory postures during UE movements. A study used CyberGlove to assess finger compensation by measuring the angles of finger joints [34]. Overall, 4 studies were conducted using robot-assisted therapies [42,59,70,87], 1 [25] using VR therapy, and 10 were conducted in homes [25,43,44,60,64,65,78,79,81,83].

Marker-Based Motion Capture System

The second most commonly used technology was the marker-based motion capture system (24/72, 33% studies). A total of 3 types of marker-based motion capture systems were used: an optical motion capture system (22/24, 92% studies), electromagnetic motion capture system [34], and ultrasound 3D motion capture system [86]. For this sensor, markers were attached to the upper body landmarks, which traditionally included the sternum, spinal process (C7 and T4), acromion processes, middle part of the humeri, lateral epicondyle, styloid process of the ulna, and bilateral thumbnails [62,63,87]. The participants were asked to perform the tasks while the positions of the markers were captured. The position and orientation of the trunk, shoulder, and elbow were then calculated according to the joint coordinate system method and used to characterize different compensation models.

For a long time, marker-based motion capture systems have been used as gold standard motion capture devices for clinical motion analysis [86]. Similarly, in the reviewed studies, marker-based motion capture systems were proven to be able to effectively identify compensation. In all, 5 studies have been used as the ground truth for the measurement of the effectiveness of other sensors in compensation detection [42,43,45,86,87]. Several interesting findings were reported using marker-based motion capture systems: (1) pre- and posttests showed that both robotic [32,38,63] and MR therapies [26] elicited benefits on reducing trunk compensatory movements, and stroke survivors showed less trunk compensatory movements during VR reaching [24]. However, Belfatto et al [33] argued that robotic therapy promoted the adoption of compensatory movements when stroke survivors performed training tasks; (2) therapist-based therapy [62] or a combination of robotic therapy and constraint-induced therapy [46] demonstrated more significant improvements in reducing trunk compensatory movements compared with robot-assisted therapy; (3) trunk displacement and shoulder elevation compensatory movements could discriminate between mild and moderate stroke paresis [66], whereas shoulder girdle compensatory movement could differentiate between mild or moderate and severe or pronounced stroke impairments [72].

Among all the studies, 13 [31-33,38,39,42,45-47,59,62,63,87] monitored compensation with robot-assisted upper limb devices, one [24] was conducted in VR therapy and one [26] was in MR therapy.

Marker-Free Vision Sensor

A total of 16 studies used marker-free vision sensors, including Microsoft Kinect depth sensors (versions 1 and 2) and a simple simple camera, as motion capture tools. Most (11/16, 69%) of these studies used Kinect, which is usually placed approximately 2.0 meters in front of the user to capture the 3D space (x-y-z) coordinates of 20 (version 1) or 25 (version 2) skeleton joint positions in the user's body at 30 frames per second. In the reviewed studies, the locations and orientations of the upper body parts (ie, hip, spine, shoulder, elbow, wrist, and hand) [48-50,61,69,86] or spine [20,23,41,51,52] were recorded, and 2 studies have verified the effectiveness of this sensor

XSL•FO
RenderX

technology for monitoring compensation. Bakhti et al [86] proved that Kinect can be used to accurately assess proximal arm nonuse when compared with an ultrasound 3D motion capture system (CMS20s, Zebris). The agreement between Kinect and CMS20s was measured using intraclass correlation coefficients (0.96), linear regression (r^2=0.92), and Bland and Altman plots (Kinect: −4.25, +6.76 to –6.76); CMS20s: −4.71, +7.88 to –7.88). Lin et al [69] found substantial agreement of detected compensation, such as shoulder elevation and hip extension, between annotated videos and Kinect (Cohen κ 0.60-0.80) and almost perfect agreement for trunk rotation and flexion (Cohen κ 0.80-1).

Overall, 6 studies used RGB cameras and 2 (33%) of them [67,74] used a camera to collect motion images to extract kinematic data through third-party data extraction algorithms for quantitative compensation assessment. The other 4 (67%) studies collected motion videos for clinicians' visual evaluation of compensation.

In all, 8 studies [41,48-52,74,84] were conducted using robot-assisted upper limb devices, 2 studies [23,69] were conducted using VR therapy, and 1 study [67] was conducted in a home using a single low-cost camera.

Physiological Signal Sensing Technology

Physiological signal sensing technologies include electromyogram (8/72, 11% studies), electroencephalogram (EEG) [33,80], and fMRI [71] systems. According to the reviewed studies, sEMG signals of upper limb muscles (including, but not limited to, biceps, triceps, upper trapezius, pectoralis major, brachioradialis, anterior, middle, and posterior deltoids) and trunk muscles (left or right rectus abdominis, left or right obliquus externus abdominis, left or right thoracic erector spinae, left or right lumbar erector spinae, and descending part of the trapezius) not only helped to discriminate true recovery and compensation [29,30,33,84] but also could be used as features for automatic compensation detection [39,53,77]. Chen et al [77] confirmed that using a generative adversarial network with sEMG signals as features could achieve excellent detection performance (accuracy=94.58%, +1.15% to –1.15%) of trunk compensatory movements.

A study used fMRI [71] to study the cortical activation pattern of compensatory movements and demonstrated that compensatory movements require a greater recruitment of cortical neurons. A total of 2 studies [33,80] showed that brain scalp EEG signals could help researchers gain more insight into the relationship between motor compensation and underlying brain activities. Among all studies, electromyogram [29,30,33,39,54,84] and EEG [33] systems were used along with robot-assisted devices for compensation assessment.

Sensors Embedded in the Rehabilitation Training System

In all, 8 studies directly selected sensors embedded in the rehabilitation training system as compensation evaluation tools. Nibras et al [35] used the measurement information in an exoskeleton to distinguish between recovery and compensation in stroke survivors. They found 2 compensatory patterns in stroke survivors: atypical decoupling of the shoulder elevation and forearm joints and atypical coupling of the shoulder horizontal rotation and elbow joints, by analyzing 4 ArmeoSpring angles when stroke survivors performed reaching movements with the ArmeoSpring exoskeleton. In contrast, a simpler and less complex UE rehabilitation robot, such as an end-effector robot, may not have the capacity to provide detailed UE measurement information as the exoskeleton. Therefore, additional sensors, such as inertial sensors [31,42,59], are required with the sensors in the end-effector robot to satisfy compensation assessment needs. In addition, Johnson et al [84,88] used the force sensors of a UE rehabilitation system, a driver simulation system, to quantify impaired arm activity [84] and LNU [88]. Kim et al [75] and Furudate et al [76] used force sensors in hand rehabilitation systems to evaluate the compensation among individual fingers.

Ambient Sensors

A total of 5 studies used ambient sensors, including a pressure distribution mattress (Body Pressure Measurement System, Model 5330, Tekscan, Inc) [45,54,91] and position measurement sensors [55,56]. A pressure distribution mattress was used to measure a person's body pressure distribution in a seated position for the automatic detection of compensatory postures [45,54,91]. Cai et al [45] verified the effectiveness of using pressure distribution data together with machine learning (ML) algorithms to detect compensatory patterns using a 3D motion capture system (VICON, Oxford Metrics) as the ground truth. When using a pressure mattress or VICON, the average F_1 score (an evaluator of the ML algorithm performance) was >0.95.

The position measurement sensors used were either a force sensor placed anterior to the back of the chair [56] or a contactless first-reflection ultrasonic echolocation sensor placed on the edge of a table [55] to monitor the position of the trunk. As only the trunk position was monitored, the researchers only realized a rough detection of compensatory trunk flexion movement. In addition, 3 studies [45,54,91] used robot-assisted upper limb devices, and 4 studies have proposed that these systems could be used in a home environment [45,54,55,91].

RQ4: Which Statistical or AI Methods Have Been Used for Compensation Assessment and Detection?

Overview

Overall, 56 studies used statistical methods and 15 adopted AI-based methods as shown in Table 7 and Table 8, respectively.

XSL•FO
RenderX

Table 7. Studies classified by statistical methods (N=56).

Data analysis scenario and statistical method	References
Differences among groups	
ANOVA	[24,36-38,41,46,75,84,88]
Mean and SD	[20,24,27,37,40,89]
Mann-Whitney *U* test	[36,40,66,85]
Wilcoxon test	[40,66]
Paired-sample *t* test	[88], 1-tailed; [66], 2-tailed; [73], 2-tailed
Principal components analysis	[35,36,66]
Regression analysis	[40,73,76]
Tukey honestly significant difference post hoc analysis	[24,37]
Tukey-Kramer tests	[84]
Scheffé test	[75]
Log-modulus transformation methods	[75]
Nonparametric Friedman test	[40]
Independent-samples *t* test	[66]
Kolmogorov-Smirnov normality test	[37]
Spearman rank correlations	[90]
Pearson correlations	[85]
Bonferroni corrections	[85]
Chi-square test	[85]
Graph learning theory	[22]
Differences before and after the intervention	
Wilcoxon signed-rank test	[32,33,52,72]
Mean and SD	[58,62,63]
ANOVA	[29,34,56]
Spearman rank correlation coefficient	[63,72]
Tukey HSD[a] test	[29,56]
Analysis of covariance	[52,62]
2-sample and paired *t* tests	[52], 1-tailed; [58], 2-tailed
Pearson correlation coefficient	[33]
Kolmogorov-Smirnov test	[58]
Mann-Whitney *U* test	[72]
Real time changes	
Canonical correlation analysis	[26,70,80]
Mean and SD	[21,23,28,31,42,47,51,57,59,60,64,68,69,82,83,86]
ANOVA	[25]
Spearman correlation test	[65,81,82,86,87]
Logistic regression	[65,82]
Paired *t* test	[87], 2-tailed
Associations of physiological signals with compensation parameters	
Spearman rank correlation coefficient test	[71]
Pearson correlation test	[39]
ANOVA	[30]

XSL•FO
RenderX

Data analysis scenario and statistical method	References
Post hoc contrasts	[30]

[a]HSD: honestly significant difference.

XSL•FO
RenderX

Table 8. Studies classified by machine learning (ML) algorithms (N=15).

ML algorithm and accuracy		References
Linear SVM[a]		
	Health: trunk compensation in 3 directions (AUC)[b]=99.15%	[61]
	Stroke (F_1 score): NC[c]=0.88; SE[d]=0.86; TR[e]=0.80; LF[f]=0.81	[77]
	Healthy group (AUC): NC=0.86; SE=0.68; TR=0.74; LF=0.98 and stroke group (AUC): NC=0.63; SE=0.27; TR=0.82; LF=0.92	[48]
	Healthy participant (F_1 score): NC=0.87; SE=0.15; TR=0.5; LF=0.74 and stroke survivor (F_1 score): NC=0.94; SE=0; TR=0; LF=0	[49]
	Healthy group (AUC): NC=0.98; SE=1.00; TR=0.99; LF=0.97 and stroke group (AUC): NC=1.00; SE=0.98; TR=0.85; LF=0.90	[53]
	Stroke (F_1 score): NC=0.990; SE=0.975; TR=0.983; LF=0.975	[54]
	Stroke: offline (F_1 score): NC=0.984; SE=1.000; TR=0.995; LF=0.963 and on the web: participant 1 (F_1 score): NC=0.978; SE=1.000; TR=0.929; LF=1.000; participant 2 (F_1 score): NC=0.994; SE=1.000; TR=1.000; LF=0.984	[45,91]
	Stroke: trunk flexion (AUC)=78.2%	[55]
***k*-NN[g]**		
	Health: trunk compensation in 3 directions (AUC)=97.9%	[61]
	Stroke (F_1 score): NC=0.79; SE=0.78; TR=0.70; LF=0.73	[77]
	Health: correct vs incorrect (involving typical compensatory movements) upper limb exercises (sensitivity and specificity): garment 1: 86%, +6% to –6% vs 79%, +7% to –7%; garment 2: 89%, +6% to –6% vs 93%, +5% to –5%; garment 3: 87%, +4% to –4% vs 84%, +4% to –4%	[78]
	Health: 3 incorrect compensatory positions (not specified) in UE[h] adduction exercise (*k* value): pos_run1=0.78, pos_run2=0.82, pos_run3=0.79, pos_run4=0.81	[79]
	Stroke (F_1 score): NC=0.989; SE=0.970; TR=0.983; LF=0.981	[54]
Naïve Bayes		
	Health: trunk displacement (precision and Recall)—non-compensatory=92.7% and 90.5% and compensatory=88.6% and 91.2%	[43]
	Stroke: trunk compensatory movements in anterior and posterior direction (precision)—Horizontal Reach: unaffected arm=100%, affected arm=87.5%; Vertical Reach: unaffected arm=87.5%, affected arm=100%; Card Flip: unaffected arm=62.5%, affected arm=66.7%; Jar Open: unaffected arm=71.4%, affected arm=71.4%	[44]
Logistic regression		
	Healthy: trunk compensation in 3 directions (AUC)=83%	[61]
	Health: 3 incorrect compensatory positions (not specified) in UE adduction exercise (*k* value): pos_run1=0.82, pos_run2=0.85, pos_run3=0.88, pos_run4=0.89	[79]
Random Forest	Healthy: trunk compensation in 3 directions (AUC)=96%	[61]
Multilabel *k*-NN	Stroke (F_1 score): NC=0.73; SE=0.53; TR=0.67; LF=0.69; insufficient elbow extension=0.73	[74]
Multilabel decision tree	Stroke (F_1 score): NC=0.69; SE=0.50; TR=0.60; LF=0.68; insufficient elbow extension=0.80	[74]
Generative adversarial network *k*-NN	Stroke (F_1 score): NC=0.94; SE=0.95; TR=0.93; LF=0.96	[77]
Sequential minimal optimization	Stroke: trunk compensatory movements in anterior and posterior direction (precision)—horizontal reach: unaffected arm=85.7%, affected arm=87.5%; vertical reach: unaffected arm=100%, affected arm=100%; Card Flip: unaffected arm=62.5%, affected arm=66.7%; Jar Open: unaffected arm=57.1%, affected arm=57.1%	[44]
Decision tree J48	Health: 3 incorrect compensatory positions (not specified) in UE adduction exercise (*k* value): pos_run1=0.64, pos_run2=0.81, pos_run3=0.82, pos_run4=0.81	[79]
Recurrent Neural Network	Healthy group (AUC): NC=0.87; SE=0.79; TR=0.84; LF=0.98 and stroke group (AUC): NC=0.66; SE=0.27; TR=0.81; LF=0.77	[48]

XSL•FO
RenderX

ML algorithm and accuracy		References
Weighted random Forest	Healthy participant (F_1 score): NC=0.87; SE=0.15; TR=0.5; LF=0.74 and stroke survivor (F_1 score): NC=0.94; SE=0; TR=0; LF=0	[49]
Cost sensitive	Healthy participant (F_1 score): NC=0.83; SE=0.09; TR=0.19; LF=0.68 and stroke survivor (F_1 score): NC=0.94; SE=0; TR=0; LF=0	[49]
Random Undersampling	Healthy participant (F_1 score): NC=0.71; SE=0.29; TR=0.48; LF=0.72 and stroke survivor (F_1 score): NC=0.69; SE=0.04; TR=0.20; LF=0.07	[49]
Tomek links	Healthy participant (F_1 score): NC=0.79; SE=0; TR=0; LF=0 and stroke survivor (F_1 score): NC=0.94; SE=0; TR=0; LF=0	[49]
SMOTE[i]	Healthy participant (F_1 score): NC=0.72; SE=0.3; TR=0.49; LF=0.82 and stroke survivor (F_1 score): NC=0.83; SE=0.06; TR=0.25; LF=0.01	[49]
SVM SMOTE	Healthy participant (F_1 score): NC=0.66; SE=0.28; TR=0.49; LF=0.73 and stroke survivor (F_1 score): NC=0.8; SE=0.04; TR=0.24; LF=0.05	[49]
Random oversampling	Healthy participant (F_1 score): NC=0.77; SE=0.32; TR=0.51; LF=0.63 and stroke survivor (F_1 score): NC=0.8; SE=0.04; TR=0.23; LF=0.07	[49]
Binary classification	Healthy participant (AUC)—good example: SE=0.94; TR=0.97; LF=0.92; bad example: SE=0.37; TR=0.63; LF=0.52	[50]

[a]SVM: support vector machine.

[b]AUC: area under the curve.

[c]NC: no compensation.

[d]SE: shoulder elevation.

[e]TR: trunk rotation.

[f]LF: lean forward.

[g]*k*-NN: *k*-nearest neighbor.

[h]UE: upper extremity.

[i]SMOTE: synthetic minority oversampling technique.

Statistical Methods

Statistical methods were used to assess compensation from 4 perspectives: real time changes of compensation measurements in body movements, group variance in compensation measurements, effects of an intervention on compensation measurements, and the statistically significant associations of physiological signals with compensation measurements.

A total of 23 studies used mean and SD, canonical correlation analysis, Spearman correlation, step-wise multiple regression, or ANOVA to test the real time changes of compensation measurements in body movements. For instance, Wittmann et al [25] used repeated measures 1-way ANOVAs to test trunk orientation changes during rehabilitation training to assess trunk compensation in real time.

In all, 20 studies tested the differences among groups to assess compensation. The most commonly used statistical methods were ANOVA and Mann–Whitney *U* test. For instance, Kim et al [75] compared all compensation measurements between groups (stroke survivors vs healthy participants) and between hands (within-group factor: more affected hand vs less affected hand in stroke survivors and nondominant hand vs dominant hand in healthy participants) with ANOVA for compensation assessment.

In addition, 10 studies analyzed the differences in compensation measurements before and after the intervention. Wilcoxon signed-rank test, ANOVA, Spearman rank correlation coefficient, paired 1- and 2-tailed *t* test and 1- and 2-tailed Tukey honestly significant difference tests were used in these studies. For instance, Fluet et al [34] used ANOVA to analyze how 2 different training models (traditional vs VR-based training) affect upper limb compensation in the dimensions of peak reaching velocity, finger extension excursion, shoulder excursion, elbow excursion, and trunk excursion.

Overall, 3 studies tested the associations between physiological signals, such as fMRI and sEMG, and compensation parameters [30,39,71]. For instance, Lee et al [71] used Spearman rank correlation coefficient to test the relationship between the brain activation area and shoulder abduction angle. They found that greater activation of the supplementary motor area was required for a larger shoulder abduction angle. Huang et al [39] applied the Pearson correlation test and found a positive correlation between muscle fatigue (measured by sEMG median frequency) and compensation. They concluded that sEMG median frequency was a good indicator of compensation due to muscle fatigue.

AI-Based Methods

A total of 15 studies used AI-based methods to detect compensatory postures, and 9 studies classified 3 common compensatory postures: trunk lean forward, trunk rotation, and shoulder elevation [45,48-50,53,54,74,77,91]. In addition, 4 studies discriminated trunk compensatory movements in the sagittal, transverse, and coronal planes [43,44,55,61], and 2 studies did not mention the type of compensatory posture that

XSL•FO
RenderX

was classified [78,79]. Dolatabadi et al [50] made the compensation data set public for other researchers. A total of 2 studies used this data set to train their ML models to improve the accuracy of compensation detection [48,49]. The remaining studies collected their own data to detect compensation.

Various ML algorithms were applied to train the classification model (Table 6). The most commonly used ML algorithm was the support vector machine (SVM). Cai et al [45] reported the highest average F_1 score (0.99) for recognizing trunk lean forward, trunk rotation, and shoulder elevation based on 5 features extracted from the pressure distribution data. Nordin et al [61] reported the highest accuracy (99.15%) for detecting the 3D trunk compensatory postures.

Notably, 8 studies [44,48,49,54,61,74,77,79] used more than one ML algorithm to compare the classification results for compensatory postures. For example, Zhi et al [48] used both SVM and recurrent neural network classifiers to classify shoulder elevation, trunk rotation, and lean forward. The results demonstrated high accuracy in healthy participants, but low accuracy in stroke survivors. Cai et al [54] applied the *k*-nearest neighbor and SVM algorithms to detect and categorize shoulder elevation, trunk rotation, and lean forward in stroke survivors, and both algorithms yielded high classification accuracies (F_1 score >0.95). Nordin et al [61] used 4 different classification algorithms with 10-fold cross-validation to assess the 3D trunk compensatory movements. The results showed accuracy of 99%, 98%, 96%, and 83% with SVM, *k*-nearest neighbor, random forest, and logistic regression, respectively.

Discussion

To the best of our knowledge, this is the first systematic review of technologies for compensation assessment and detection of UE movements in stroke survivors.

RQ1: What Models Have Been Established to Assess and Detect Compensation?

Notably, 2 types of compensation were categorized. Most (63/72, 88%) studies focused on investigating the awkward use of the affected UE. The reason might be that the awkward pattern is more complicated to be observed than the disuse pattern [81]. The synergy and coupling of body parts are difficult to understand [92], which requires more evidence-based methods to fuse data from more resources across a constant timeline. Sensor technologies offer fine-grained rich data, and together with AI methods, can provide a low-cost solution for continuous monitoring of a person's performance.

The models of the disuse pattern focus on the amount of use of the affected UE. For the awkward pattern, the models focused more on how the unaffected body parts were involved in the motion with the affected UE. The most discussed body parts were the trunk, shoulders, and elbows. Trunk compensation was the most discussed factor, suggesting that it is more common among stroke survivors.

Models were established for different task scenarios. For the disuse pattern, bilateral tasks were the most common. For the awkward pattern, reaching tasks were mostly used. Reaching was the basic movement of the upper limbs that constituted most daily life behaviors [93]. Reaching requires coordination of multiple joints of the arm and is controlled by the central nervous system [93]. Different reaching ranges can result in various compensations for the trunk, shoulder, and elbow [15].

RQ2: What Measurements Are Used to Evaluate Compensatory Movements?

Notably, 2 clinical scales, the Motor Activity Log [16] and the Actual Amount of Use Test [17], have traditionally been used for the evaluation of disuse patterns. However, these are subjective and difficult to replace using technology-based methods. Levin et al [15] proposed the Reaching Performance Scale for awkward pattern evaluation. However, none of the reviewed studies have quantified this scale using technological methods. Moreover, UE functional impairment scales (eg, FMA) were not used to assess compensation.

Quantitative measurements have been proposed for technology-based compensation assessments. For the disuse pattern, measurements such as the movement duration and frequency of use were used to describe the use of the affected UE. For the awkward pattern, linear displacement, angular displacement, acceleration, and sEMG signals of the trunk and upper limb joints were the most common measurements. Furthermore, the trunk compensation measurements, which are the kinematic measurements of the trunk in the 3 anatomical planes, are more uniform. In contrast, shoulder compensation measurements are more diverse and complex. This could be because the shoulder has more freedom of movement, and the configuration of these movements could vary across different experimental tasks [26-28,37,39,45,48-50,53,54,64,66,69,91].

Further studies could be conducted to explore the relations among all these compensation measurements and to develop a set of gold standard quantitative measurements.

RQ3: What Types of Sensor Technology Are Used for Compensation Assessment and Detection?

Marker-based motion capture systems yield accurate and robust real time motion tracking and have been used as ground truth to verify the effectiveness of other sensors for compensation assessment and detection [42,43,45]. In our reviewed studies, marker-based motion capture systems were used to detect various compensations, including arm nonuse [86,87], trunk compensation [26,45], shoulder compensation [66,72] and interlimb coordination [36]. The drawbacks of these systems include but are not limited to the cost of both hardware and software, complicated setup, and the need for professionals to operate the systems [50]. These systems may also require a specific space, such as an area with a clear line of sight for the cameras [44]. The use of cameras in a home environment may raise privacy concerns [44].

Similarly, although with great accuracy, the setup of physiological signal sensing technologies is complex and has been limited to its use in laboratories or other controlled environments. In addition, professionals are required to collect and analyze these physiological signals [27]. The advantage of using this sensor technology is that the recorded sEMG signals of relevant muscles [30,84], brain scalp EEG signals [33,80],

XSL•FO
RenderX

and cortical activation patterns [71] could help researchers gain more insight into compensation from the perspective of muscle activities and brain activities, which in turn would provide more information for compensation detection and correction to improve UE motor performance in stroke survivors.

Body-worn sensors were the most common technology used for compensation assessment and detection in the reviewed studies. They were able to monitor all compensation models [27,28,42-44,70,81-83]. Compared with marker-based motion capture systems, body-worn sensors are more affordable and portable, with a simpler setup [43,44]. More than half (47/72, 65%) of the studies used these sensors in technology-based therapies or home settings, which shows that this sensor technology has great potential for use in less-supervised therapy environments. The main disadvantage of this technology is that it can induce unnatural movements owing to the sensor attachment on the user's body, which may affect the accuracy of compensation assessment [45]. Future research could focus on reducing or avoiding the possible unnatural movements caused by sensor attachment during a compensation assessment process, such as correcting the deviation through algorithms or adopting a more ingenious physical layout of the sensors.

Similar to body-worn sensors, marker-free vision sensors are low-cost and have an easy setup [94]. Owing to their size and portability, they could be an ideal option for home use. Marker-free vision sensors have been used to detect arm use [84]; arm nonuse [86]; and trunk [20,23], shoulder [69,74], and elbow [74] compensation. They were used together with ML algorithms to automatically detect typical compensatory postures (no compensation, shoulder elevation, trunk rotation,

lean forward, etc) [48,49,61,74]. The sensors can capture stroke survivors' motion images in real time for clinicians to determine the compensation adopted during the training process. These images were used to train AI models to automatically detect compensatory postures. Compared with the RGB camera, Kinect was more commonly used. This could be because of the various types of information provided by the Kinect depth sensor, including color images, depth images, and 3D skeleton joint positions of the human body. However, it has been reported that the prediction of joint positions of the shoulder and trunk by Kinect suffers from large errors when sitting with trunk flexion (approximately 100 mm), which is a common compensatory movement after stroke [61]. One of the weaknesses of using marker-free vision sensors is that they can introduce privacy concerns if used in a home and may induce unnatural behaviors owing to the negative feelings caused by surveillance [44].

Relatively few studies have been conducted on sensors embedded in rehabilitation systems and ambient sensors for compensation assessment and detection. When a stroke survivor completes exercises with the assistance of a rehabilitation training system, it is intuitive to use the same system for compensation assessment [95]. However, for less complex rehabilitation robots with a simpler setup, such as end-effector robots, external measures may be required because the data collected by the system are not sufficient to detect compensation [31,42,59]. Ambient sensors are typically simple and unobtrusive [45]. They have great potential for use in compensation assessment and detection in less-supervised therapy environments, especially in home settings. However, only limited compensation can be detected by ambient sensors. Thus, more research could focus on accurately detecting compensatory movements using these sensors.

In summary, all sensor technologies have their own advantages and disadvantages. Both marker-based motion capture systems and physiological sensing technologies are limited by their use of a more controlled environment. Although with great accuracy in compensation detection, the setup is complicated and requires expert experience. Marker-based technology is usually used as the gold standard to test the accuracy of other technologies in compensation detection and measurement. In comparison with the results of marker-based technologies, body-worn sensors [27,28,43], marker-free vision sensors [86], sensors embedded in rehabilitation training systems [42], and ambient sensors [45] have also been proven effective in compensation assessment and detection. Body-worn sensors, marker-free vision sensors, and ambient sensors are low-cost, easy to set up, and can be used in less-controlled environments, such as home settings. However, marker-free vision sensors can increase privacy concerns. Thus, it may cause deployment issues in the home environment. Both wearable sensors and marker-free vision sensors can cause incorrect postures owing to the unnatural movements induced by the sensors. Directly using sensors embedded in rehabilitation training systems to assess and detect compensation could be a simple and convenient method. However, researchers should be aware of (1) whether the sensors in the system can meet the accuracy requirements and (2) whether the sensors in the system can capture all the necessary data for compensation assessment and detection. Finally, it is suggested that a rehabilitation training system be built that integrates training exercises, compensation assessment and detection, and real time compensation feedback for stroke survivors to perform effective rehabilitation with less or even without the supervision of a therapist.

RQ4: Which Statistical or AI Methods Have Been Used for Compensation Assessment and Detection?

Research based on statistical methods provides valuable information about compensation assessment and detection, such as the difference in compensation measurements between healthy people and stroke survivors [20,66,73,84], changes in compensation measurements before and after an intervention [34,62,63,72], and the correlation of physiological signals with compensation measurements [30,39,71]. This information can be processed further in future studies for compensation assessment and detection.

The majority of studies used descriptive statistics, such as mean and SD, for real time compensation detection [23,47,68,69,82,83,86]. Although descriptive statistics are simple to use, the application of this method to detect compensation relies heavily on expert experience. For example, an acceptable range of compensation measurements was set by therapists, and the occurrence of compensation was decided by the therapists based on observation of the stroke survivors' movements if they exceeded the compensation range. Therefore,

XSL•FO
RenderX

this method is subjective and may not be accurate. In future, more research could focus on using other statistical methods, such as logistic regression, for real time detection of compensatory movements.

In contrast to statistical methods, AI methods have been used to automatically detect compensatory postures. They showed great potential for real time compensatory posture detection in less-supervised therapy environments [43-45,48-50,53,54,74,78,79,91]. One limitation of this research area is that there are few public data sets on compensatory movements in stroke survivors. In our review, only one open data set (the Toronto Rehab Stroke Posture data set) was found. Open research data are an originally collected data set that is accessible and can be reused by other researchers to conduct their research [96,97]. It has been gaining attention and growing popularity among researchers and funding agencies [96,98]. As such, future studies should make data accessible and sharable among research communities.

Furthermore, although a variety of ML algorithms have been identified for compensatory posture detection, they can only identify a single compensatory posture at a time, which cannot meet the situation where multiple compensatory postures appear concurrently. Moreover, AI methods have not yet been used to predict the occurrence of compensation. Therefore, more effort is needed to build more heterogeneous AI models, such as multilabel ML models and deep learning models, for multiple compensation detection and prediction.

Strengths and Limitations

Our study had several strengths. This study applied comprehensive searches in both technology and medical fields. This is the first comprehensive systematic review of technology-based compensation assessment and detection in UE rehabilitation for stroke survivors. It is the only systematic review summarizing compensation models and their measurements and has reviewed the use of statistical and AI methods for compensation assessment and detection.

Our study has some limitations. First, the review included only references in English. Second, owing to inconsistencies in compensation assessment criteria across studies, the review did not include comparisons of the effectiveness of different technologies for compensation evaluation.

Conclusions and Future Research

This systematic review focuses on how technologies are used for compensation assessment and detection during UE rehabilitation of stroke survivors. It covers models and measurements to describe the compensation and different types of sensors and statistical and AI methods for compensation assessment and detection. Evidence suggests that technology-based compensation assessment and detection can augment rehabilitation without the constant presence of therapists. Future studies could (1) explore how to develop a set of gold standard quantitative compensation measurements; (2) investigate how to overcome the discussed defects of body-worn sensors, marker-free vision sensors, and system-embedded sensors in compensation evaluation and how to integrate feedback with these sensors so that they can be used in less-supervised or even unsupervised UE rehabilitation environments; (3) focus more on open data as they provide opportunities for reuse in algorithm development for automatic real time compensation assessment and detection; (4) study multilabel classification algorithms and deep learning algorithms for multiple compensation detection; and (5) research more on compensation prediction.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (71771098). This work was performed with close collaboration among researchers affiliated with the University of Toronto and Huazhong University of Science and Technology Collaborative Center for Robotics and Eldercare. The authors also thank the rehabilitation professionals in the Hubei Provincial Hospital of Traditional Chinese Medicine.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 31 KB - jmir_v24i6e34307_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOCX File , 46 KB - jmir_v24i6e34307_app2.docx]

References

1. GBD 2016 Stroke Collaborators. Global, regional, and national burden of stroke, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol 2019 May;18(5):439-458 [FREE Full text] [doi: 10.1016/S1474-4422(19)30034-1] [Medline: 30871944]

XSL•FO
RenderX

2. Pollock A, Farmer SE, Brady MC, Langhorne P, Mead GE, Mehrholz J, et al. Interventions for improving upper limb function after stroke. Cochrane Database Syst Rev 2014 Nov 12;2014(11):CD010820 [FREE Full text] [doi: 10.1002/14651858.CD010820.pub2] [Medline: 25387001]
3. Levin MF, Kleim JA, Wolf SL. What do motor "recovery" and "compensation" mean in patients following stroke? Neurorehabil Neural Repair 2009 May;23(4):313-319. [doi: 10.1177/1545968308328727] [Medline: 19118128]
4. Alaverdashvili M, Foroud A, Lim DH, Whishaw IQ. "Learned baduse" limits recovery of skilled reaching for food after forelimb motor cortex stroke in rats: a new analysis of the effect of gestures on success. Behav Brain Res 2008 Apr 09;188(2):281-290. [doi: 10.1016/j.bbr.2007.11.007] [Medline: 18155782]
5. Allred RP, Cappellini CH, Jones TA. The "good" limb makes the "bad" limb worse: experience-dependent interhemispheric disruption of functional outcome after cortical infarcts in rats. Behav Neurosci 2010 Feb;124(1):124-132 [FREE Full text] [doi: 10.1037/a0018457] [Medline: 20141287]
6. Neuroplasticity. Toronto Stroke Networks. URL: http://strokerecovery.guide/neuroplasticity [accessed 2022-04-12]
7. Kleim JA, Jones TA. Principles of experience-dependent neural plasticity: implications for rehabilitation after brain damage. J Speech Lang Hear Res 2008 Feb;51(1):S225-S239. [doi: 10.1044/1092-4388(2008/018)] [Medline: 18230848]
8. Walker MF, Sunnerhagen KS, Fisher RJ. Evidence-based community stroke rehabilitation. Stroke 2013 Jan;44(1):293-297. [doi: 10.1161/STROKEAHA.111.639914] [Medline: 23093614]
9. Kwakkel G, Lannin NA, Borschmann K, English C, Ali M, Churilov L, et al. Standardized measurement of sensorimotor recovery in stroke trials: consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable. Int J Stroke 2017 Jul;12(5):451-461. [doi: 10.1177/1747493017711813] [Medline: 28697709]
10. Rodríguez-de-Pablo C, Popović M, Savić A, Perry JC, Belloso A, Tomić DT. Post-stroke robotic upper-limb telerehabilitation using serious games to increase patient motivation: first results from ArmAssist system clinical trial. In: Proceedings of the 2nd International Congress on Neurotechnology, Electronics and Informatics. 2016 Presented at: NEUROTECHNIX '14; October 25-26, 2014; Rome, Italy p. 63-78.
11. Chen Y, Abel KT, Janecek JT, Chen Y, Zheng K, Cramer SC. Home-based technologies for stroke rehabilitation: a systematic review. Int J Med Inform 2019 Mar;123:11-22 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.12.001] [Medline: 30654899]
12. Lister MJ. Contemporary management of motor control problems: proceedings of the II STEP Conference. Alexandria, VA, USA: Foundation for Physical Therapy; 1990.
13. Cochrane Handbook Section 3.6. Cochrane Training. URL: https://training.cochrane.org/handbook/current/chapter-04-technical-supplement%20searching-and-selecting-studies [accessed 2022-04-12]
14. CEBM Levels of Evidence Working Group. The Oxford Levels of Evidence 1. Oxford Centre for Evidence-Based Medicine. 2009. URL: https://www.cebm.net/index.aspx?o=5653 [accessed 2022-04-12]
15. Levin MF, Desrosiers J, Beauchemin D, Bergeron N, Rochette A. Development and validation of a scale for rating motor compensations used for reaching in patients with hemiparesis: the reaching performance scale. Phys Ther 2004 Jan;84(1):8-22. [Medline: 14992673]
16. Taub E, Miller NE, Novack TA, Cook 3rd EW, Fleming WC, Nepomuceno CS, et al. Technique to improve chronic motor deficit after stroke. Arch Phys Med Rehabil 1993 Apr;74(4):347-354. [Medline: 8466415]
17. Taub E, Crago JE, Uswatte G. Constraint-induced movement therapy: a new approach to treatment in physical rehabilitation. Rehabil Psychol 1998;43(2):152-170. [doi: 10.1037/0090-5550.43.2.152]
18. Gowland C, Stratford P, Ward M, Moreland J, Torresin W, Van Hullenaar S, et al. Measuring physical impairment and disability with the Chedoke-McMaster Stroke Assessment. Stroke 1993 Jan;24(1):58-63. [doi: 10.1161/01.str.24.1.58] [Medline: 8418551]
19. 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]
20. Garcia AT, da Kelbouscas AL, Silva SA, Guimarães LL, de Oliveria VM. Monitoring and analysis of compensatory trunk movements with RGB-D camera and wireless system for rehabilitation of upper limbs. In: Proceedings of the 2020 Latin American Robotics Symposium (LARS), 2020 Brazilian Symposium on Robotics (SBR) and 2020 Workshop on Robotics in Education (WRE). 2020 Presented at: LARS/SBR/WRE '20; November 9-13, 2020; Natal, Brazil p. 1-6. [doi: 10.1109/LARS/SBR/WRE51543.2020.9307087]
21. Alhwoaimel N, Turk R, Hughes AM, Ferrari F, Burridge J, Wee SK, et al. Instrumented trunk impairment scale (iTIS): a reliable measure of trunk impairment in the stroke population. Top Stroke Rehabil 2021 Sep;28(6):456-463. [doi: 10.1080/10749357.2020.1834273] [Medline: 33070742]
22. Contreras RC, Parnandi A, Coelho BG, Silva C, Schambra H, Nonato LG. NE-motion: visual analysis of stroke patients using motion sensor networks. Sensors (Basel) 2021 Jun 30;21(13):4482 [FREE Full text] [doi: 10.3390/s21134482] [Medline: 34208996]
23. Foreman MH, Engsberg JR. A virtual reality tool for measuring and shaping trunk compensation for persons with stroke: design and initial feasibility testing. J Rehabil Assist Technol Eng 2019 Feb 7;6:2055668318823673 [FREE Full text] [doi: 10.1177/2055668318823673] [Medline: 31245028]

XSL•FO
RenderX

24. Liebermann DG, Levin MF, Berman S, Weingarden HP, Weiss PL. Kinematic features of arm and trunk movements in stroke patients and age-matched healthy controls during reaching in virtual and physical environments. In: Proceedings of the 2009 Virtual Rehabilitation International Conference. 2009 Presented at: ICVR '09; June 29-July 2, 2009; Haifa, Israel p. 179-184. [doi: 10.1109/ICVR.2009.5174228]
25. Wittmann F, Held JP, Lambercy O, Starkey ML, Curt A, Höver R, et al. Self-directed arm therapy at home after stroke with a sensor-based virtual reality training system. J Neuroeng Rehabil 2016 Aug 11;13(1):75 [FREE Full text] [doi: 10.1186/s12984-016-0182-1] [Medline: 27515583]
26. Duff M, Chen Y, Attygalle S, Herman J, Sundaram H, Qian G, et al. An adaptive mixed reality training system for stroke rehabilitation. IEEE Trans Neural Syst Rehabil Eng 2010 Oct;18(5):531-541. [doi: 10.1109/TNSRE.2010.2055061] [Medline: 20934938]
27. Salazar AJ, Silva AS, Silva C, Borges CM, Correia MV, Santos RS, et al. Low-cost wearable data acquisition for stroke rehabilitation: a proof-of-concept study on accelerometry for functional task assessment. Top Stroke Rehabil 2014;21(1):12-22. [doi: 10.1310/tsr2101-12] [Medline: 24521836]
28. Borges CL, Silva C, Salazar AJ, Silva AS, Correia MV, Santos RS, et al. Compensatory movement detection through inertial sensor positioning for post-stroke rehabilitation. In: Proceedings of the International Conference on Bio-inspired Systems and Signal Processing. 2012 Presented at: BIOSIGNALS '12; February 1-4, 2012; Vilamoura, Algarve, Portugal p. 297-302. [doi: 10.5220/0003798102970302]
29. Lum PS, Burgar CG, Shor PC. Use of the MIME robotic system to retrain multijoint reaching in post-stroke hemiparesis: why some movement patterns work better than others. In: Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2003 Presented at: IEMBS '03; September 17-21,2003; Cancun, Mexico p. 1475-1478. [doi: 10.1109/iembs.2003.1279614]
30. Lum PS, Burgar CG, Shor PC. Evidence for improved muscle activation patterns after retraining of reaching movements with the MIME robotic system in subjects with post-stroke hemiparesis. IEEE Trans Neural Syst Rehabil Eng 2004 Jun;12(2):186-194. [doi: 10.1109/TNSRE.2004.827225] [Medline: 15218933]
31. Passon A, Schauer T, Seel T. Inertial-robotic motion tracking in end-effector-based rehabilitation robots. Front Robot AI 2020 Nov 27;7:554639 [FREE Full text] [doi: 10.3389/frobt.2020.554639] [Medline: 33501318]
32. Amano Y, Noma T, Etoh S, Miyata R, Kawamura K, Shimodozono M. Reaching exercise for chronic paretic upper extremity after stroke using a novel rehabilitation robot with arm-weight support and concomitant electrical stimulation and vibration: before-and-after feasibility trial. Biomed Eng Online 2020 May 06;19(1):28 [FREE Full text] [doi: 10.1186/s12938-020-00774-3] [Medline: 32375788]
33. Belfatto A, Scano A, Chiavenna A, Mastropietro A, Mrakic-Sposta S, Pittaccio S, et al. A multiparameter approach to evaluate post-stroke patients: an application on robotic rehabilitation. Appl Sci 2018 Nov 14;8(11):2248. [doi: 10.3390/app8112248]
34. Fluet GG, Merians AS, Qiu Q, Rohafaza M, VanWingerden AM, Adamovich SV. Does training with traditionally presented and virtually simulated tasks elicit differing changes in object interaction kinematics in persons with upper extremity hemiparesis? Top Stroke Rehabil 2015 Jun;22(3):176-184 [FREE Full text] [doi: 10.1179/1074935714Z.0000000008] [Medline: 26084322]
35. Nibras N, Liu C, Mottet D, Wang C, Reinkensmeyer D, Remy-Neris O, et al. Dissociating sensorimotor recovery and compensation during exoskeleton training following stroke. Front Hum Neurosci 2021 Apr 30;15:645021 [FREE Full text] [doi: 10.3389/fnhum.2021.645021] [Medline: 33994981]
36. Reisman DS, Scholz JP. Aspects of joint coordination are preserved during pointing in persons with post-stroke hemiparesis. Brain 2003 Nov;126(Pt 11):2510-2527. [doi: 10.1093/brain/awg246] [Medline: 12958080]
37. Scano A, Molteni F, Molinari Tosatti L. Low-cost tracking systems allow fine biomechanical evaluation of upper-limb daily-life gestures in healthy people and post-stroke patients. Sensors (Basel) 2019 Mar 11;19(5):1224 [FREE Full text] [doi: 10.3390/s19051224] [Medline: 30862049]
38. Wu CY, Yang CL, Chen MD, Lin KC, Wu LL. Unilateral versus bilateral robot-assisted rehabilitation on arm-trunk control and functions post stroke: a randomized controlled trial. J Neuroeng Rehabil 2013 Apr 12;10:35 [FREE Full text] [doi: 10.1186/1743-0003-10-35] [Medline: 23587106]
39. Huang S, Cai S, Li G, Chen Y, Ma K, Xie L. sEMG-based detection of compensation caused by fatigue during rehabilitation therapy: a pilot study. IEEE Access 2019 Aug 5;7:127055-127065. [doi: 10.1109/access.2019.2933287]
40. Mandon L, Boudarham J, Robertson J, Bensmail D, Roche N, Roby-Brami A. Faster reaching in chronic spastic stroke patients comes at the expense of arm-trunk coordination. Neurorehabil Neural Repair 2016 Mar;30(3):209-220. [doi: 10.1177/1545968315591704] [Medline: 26089311]
41. Valdés BA, Glegg SM, Van der Loos HF. Trunk compensation during bimanual reaching at different heights by healthy and hemiparetic adults. J Mot Behav 2017;49(5):580-592. [doi: 10.1080/00222895.2016.1241748] [Medline: 27935472]
42. Bertomeu-Motos A, Morales R, Diez JA, Lledó LD, Badesa FJ, Garcia-Aracil N. Kinematic reconstruction of the upper limb joints in planar robot-aided therapies. In: Proceedings of the 2015 IEEE International Conference on Rehabilitation Robotics. 2015 Presented at: ICORR '15; August 11-14, 2015; Singapore, Singapore p. 888-893. [doi: 10.1109/icorr.2015.7281315]

XSL•FO
RenderX

43. Ranganathan R, Wang R, Dong B, Biswas S. Identifying compensatory movement patterns in the upper extremity using a wearable sensor system. Physiol Meas 2017 Nov 30;38(12):2222-2234. [doi: 10.1088/1361-6579/aa9835] [Medline: 29099724]
44. Ranganathan R, Wang R, Gebara R, Biswas S. Detecting compensatory trunk movements in stroke survivors using a wearable system. In: Proceedings of the 2017 Workshop on Wearable Systems and Applications. 2017 Presented at: WearSys '17; June 19, 2017; Niagara Falls, NY, USA p. 29-32. [doi: 10.1145/3089351.3089353]
45. Cai S, Wei X, Su E, Wu W, Zheng H, Xie L. Online compensation detecting for real-time reduction of compensatory motions during reaching: a pilot study with stroke survivors. J Neuroeng Rehabil 2020 Apr 28;17(1):58 [FREE Full text] [doi: 10.1186/s12984-020-00687-1] [Medline: 32345335]
46. Hsieh YW, Liing RJ, Lin KC, Wu CY, Liou TH, Lin JC, et al. Sequencing bilateral robot-assisted arm therapy and constraint-induced therapy improves reach to press and trunk kinematics in patients with stroke. J Neuroeng Rehabil 2016 Mar 22;13:31 [FREE Full text] [doi: 10.1186/s12984-016-0138-5] [Medline: 27000446]
47. Proffitt RM, Alankus G, Kelleher CL, Engsberg JR. Use of computer games as an intervention for stroke. Top Stroke Rehabil 2011;18(4):417-427. [doi: 10.1310/tsr1804-417] [Medline: 21914607]
48. Zhi YX, Lukasik M, Li MH, Dolatabadi E, Wang RH, Taati B. Automatic detection of compensation during robotic stroke rehabilitation therapy. IEEE J Transl Eng Health Med 2017 Dec 15;6:2100107 [FREE Full text] [doi: 10.1109/JTEHM.2017.2780836] [Medline: 29404226]
49. Uy SR, Abu PA. Analysis of detecting compensation for robotic stroke rehabilitation therapy using imbalanced learning and outlier detection. In: Proceedings of the 2020 International Conference on Artificial Intelligence in Information and Communication. 2020 Presented at: ICAIIC '20; February 19-21, 2020; Fukuoka, Japan p. 432-437. [doi: 10.1109/icaiic48513.2020.9064992]
50. Dolatabadi E, Zhi YX, Ye B, Coahran M, Lupinacci G, Mihailidis A, et al. The toronto rehab stroke pose dataset to detect compensation during stroke rehabilitation therapy. In: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2017 Presented at: PervasiveHealth '17; May 23-26, 2017; Barcelona, Spain p. 375-381. [doi: 10.1145/3154862.3154925]
51. Valdés BA, Van der Loos HF. Biofeedback vs. game scores for reducing trunk compensation after stroke: a randomized crossover trial. Top Stroke Rehabil 2018 Mar;25(2):96-113. [doi: 10.1080/10749357.2017.1394633] [Medline: 29078743]
52. Valdés BA, Schneider AN, Van der Loos HF. Reducing trunk compensation in stroke survivors: a randomized crossover trial comparing visual and force feedback modalities. Arch Phys Med Rehabil 2017 Oct;98(10):1932-1940. [doi: 10.1016/j.apmr.2017.03.034] [Medline: 28526482]
53. Ma K, Chen Y, Zhang X, Zheng H, Yu S, Cai S, et al. sEMG-based trunk compensation detection in rehabilitation training. Front Neurosci 2019 Nov 21;13:1250 [FREE Full text] [doi: 10.3389/fnins.2019.01250] [Medline: 31824250]
54. Cai S, Li G, Zhang X, Huang S, Zheng H, Ma K, et al. Detecting compensatory movements of stroke survivors using pressure distribution data and machine learning algorithms. J Neuroeng Rehabil 2019 Nov 04;16(1):131 [FREE Full text] [doi: 10.1186/s12984-019-0609-6] [Medline: 31684970]
55. Griffith H, Ranganathan R, Biswas S. Towards a first-reflection ultrasonic sensor array for compensatory movement identification in stroke sufferers. In: Proceedings of the IEEE 35th International Performance Computing and Communications Conference. 2016 Presented at: IPCCC '16; December 9-11, 2016; Las Vegas, NV, USA p. 1-2. [doi: 10.1109/pccc.2016.7820611]
56. Thielman G. Rehabilitation of reaching poststroke: a randomized pilot investigation of tactile versus auditory feedback for trunk control. J Neurol Phys Ther 2010 Sep;34(3):138-144. [doi: 10.1097/NPT.0b013e3181efa1e8] [Medline: 20716988]
57. Fan W, Zhang Y, Wang QM, Bai Y, Wu Y. An interactive motion-tracking system for home-based assessing and training reach-to-target tasks in stroke survivors-a preliminary study. Med Biol Eng Comput 2020 Jul;58(7):1529-1547. [doi: 10.1007/s11517-020-02173-1] [Medline: 32405968]
58. Corti M, McGuirk TE, Wu SS, Patten C. Differential effects of power training versus functional task practice on compensation and restoration of arm function after stroke. Neurorehabil Neural Repair 2012 Sep;26(7):842-854. [doi: 10.1177/1545968311433426] [Medline: 22357633]
59. Passon A, Schauer T, Seel T. Hybrid inertial-robotic motion tracking for posture biofeedback in upper limb rehabilitation. In: Proceedings of the 7th IEEE International Conference on Biomedical Robotics and Biomechatronics. 2018 Presented at: Biorob '18; August 26-29, 2018; Enschede, The Netherlands p. 1163-1168. [doi: 10.1109/biorob.2018.8487203]
60. Alankus G. Motion-based video games for stroke rehabilitation with reduced compensatory motions. Washington University in St.Louis. 2011. URL: https://openscholarship.wustl.edu/etd/547/ [accessed 2022-04-12]
61. Nordin N, Xie SQ, Wunsche BC. Simple torso model for upper limb compensatory assessment after stroke. In: Proceedings of the 2016 IEEE International Conference on Advanced Intelligent Mechatronics. 2016 Presented at: AIM '16; July 12-15, 2016; Banff, AB, Canada p. 775-780. [doi: 10.1109/aim.2016.7576862]
62. Wu CY, Yang CL, Chuang LL, Lin KC, Chen HC, Chen MD, et al. Effect of therapist-based versus robot-assisted bilateral arm training on motor control, functional performance, and quality of life after chronic stroke: a clinical trial. Phys Ther 2012 Aug;92(8):1006-1016. [doi: 10.2522/ptj.20110282] [Medline: 22517782]

XSL•FO
RenderX

63. Carpinella I, Lencioni T, Bowman T, Bertoni R, Turolla A, Ferrarin M, et al. Effects of robot therapy on upper body kinematics and arm function in persons post stroke: a pilot randomized controlled trial. J Neuroeng Rehabil 2020 Jan 30;17(1):10 [FREE Full text] [doi: 10.1186/s12984-020-0646-1] [Medline: 32000790]
64. Delbressine F, Timmermans A, Beursgens L, de Jong M, van Dam A, Verweij D, et al. Motivating arm-hand use for stroke patients by serious games. Annu Int Conf IEEE Eng Med Biol Soc 2012;2012:3564-3567. [doi: 10.1109/EMBC.2012.6346736] [Medline: 23366697]
65. Barth J, Klaesner JW, Lang CE. Relationships between accelerometry and general compensatory movements of the upper limb after stroke. J Neuroeng Rehabil 2020 Oct 20;17(1):138 [FREE Full text] [doi: 10.1186/s12984-020-00773-4] [Medline: 33081783]
66. Alt Murphy M, Willén C, Sunnerhagen KS. Kinematic variables quantifying upper-extremity performance after stroke during reaching and drinking from a glass. Neurorehabil Neural Repair 2011 Jan;25(1):71-80. [doi: 10.1177/1545968310370748] [Medline: 20829411]
67. Sucar LE, Luis R, Leder RS, Hernández J, Sánchez I. Gesture therapy: a vision-based system for upper extremity stroke rehabilitation. In: Proceedings of the 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology. 2010 Presented at: IEMBS '10; August 30-September 4, 2010; Buenos Aires, Argentina p. 71-80. [doi: 10.1109/IEMBS.2010.5627458]
68. Nguyen G, Maclean J, Stirling L. Quantification of compensatory torso motion in post-stroke patients using wearable inertial measurement units. IEEE Sensors J 2021 Jul 1;21(13):15349-15360. [doi: 10.1109/jsen.2021.3072010]
69. Lin S, Mann J, Mansfield A, Wang RH, Harris JE, Taati B. Investigating the feasibility and acceptability of real-time visual feedback in reducing compensatory motions during self-administered stroke rehabilitation exercises: a pilot study with chronic stroke survivors. J Rehabil Assist Technol Eng 2019 Mar 18;6:2055668319831631 [FREE Full text] [doi: 10.1177/2055668319831631] [Medline: 31245031]
70. Hennes M, Bollue K, Arenbeck H, Disselhorst-Klug C. A proposal for patient-tailored supervision of movement performance during end-effector-based robot-assisted rehabilitation of the upper extremities. Biomed Tech (Berl) 2015 Jun;60(3):193-197. [doi: 10.1515/bmt-2014-0021] [Medline: 25460278]
71. Lee MY, Park JW, Park RJ, Hong JH, Son SM, Ahn SH, et al. Cortical activation pattern of compensatory movement in stroke patients. NeuroRehabilitation 2009;25(4):255-260. [doi: 10.3233/NRE-2009-0523] [Medline: 20037218]
72. Khizhnikova AE, Klochkov AS, Kotov–Smolenskiy AM, Suponeva NA, Piradov MA. Dynamics of post-stroke hand paresis kinematic pattern during rehabilitation. Bull Russ State Med Univ 2019 Aug 31(4):32-38. [doi: 10.24075/brsmu.2019.056]
73. Cruz VT, Bento V, Ruano L, Ribeiro DD, Fontão L, Mateus C, et al. Motor task performance under vibratory feedback early poststroke: single center, randomized, cross-over, controlled clinical trial. Sci Rep 2014 Jul 11;4:5670 [FREE Full text] [doi: 10.1038/srep05670] [Medline: 25011667]
74. Fu Y, Wang X, Zhu Z, Tan J, Zhao Y, Ding Y, et al. Vision-based automatic detection of compensatory postures of after-stroke patients during upper-extremity robot-assisted rehabilitation: a pilot study in reaching movement. In: Proceedings of the 2020 International Conference on Assistive and Rehabilitation Technologies. 2020 Presented at: iCareTech '20; August 28-29, 2020; Gaza, Palestine p. 62-66. [doi: 10.1109/icaretech49914.2020.00019]
75. Kim Y, Koh K, Yoon B, Kim WS, Shin JJ, Park HS, et al. Examining impairment of adaptive compensation for stabilizing motor repetitions in stroke survivors. Exp Brain Res 2017 Dec;235(12):3543-3552. [doi: 10.1007/s00221-017-5074-5] [Medline: 28879510]
76. Furudate Y, Ohnuki N, Chiba K, Ishida Y, Mikami S. Real-time evaluation of hand motor function recovery in home use finger rehabilitation device using Gaussian process Regression. In: Proceedings of the IEEE 20th International Conference on Bioinformatics and Bioengineering. 2020 Presented at: BIBE '20; October 26-28, 2020; Cincinnati, OH, USA p. 942-945. [doi: 10.1109/bibe50027.2020.00159]
77. Chen Y, Ma K, Yang L, Yu S, Cai S, Xie L. Trunk compensation electromyography features purification and classification model using generative adversarial network. Biomed Signal Process Control 2021 Mar;65:102345. [doi: 10.1016/j.bspc.2020.102345]
78. Giorgino T, Tormene P, Maggioni G, Capozzi D, Quaglini S, Pistarini C. Assessment of sensorized garments as a flexible support to self-administered post-stroke physical rehabilitation. Eur J Phys Rehabil Med 2009 Mar;45(1):75-84 [FREE Full text] [Medline: 19293756]
79. Giorgino T, Lorussi F, De Rossi D, Quaglini S. Posture classification via wearable strain sensors for neurological rehabilitation. Conf Proc IEEE Eng Med Biol Soc 2006;2006:6273-6276. [doi: 10.1109/IEMBS.2006.260620] [Medline: 17946755]
80. Spüler M, Rosenstiel W, Bogdan M. Predicting wrist movement trajectory from Ipsilesional ECoG in chronic stroke survivors. In: Proceedings of the 2nd International Congress on Neurotechnology, Electronics and Informatics. 2014 Presented at: NEUROTECHNIX '14; October 25-26, 2014; Rome, Italy p. 38-45. [doi: 10.5220/0005165200380045]
81. Miller A, Duff S, Quinn L, Bishop L, Youdan G, Ruthrauff H, et al. Development of sensor-based measures of upper extremity interlimb coordination. Annu Int Conf IEEE Eng Med Biol Soc 2018 Jul;2018:3160-3164. [doi: 10.1109/EMBC.2018.8512903] [Medline: 30441065]

XSL•FO
RenderX

82. Thrane G, Emaus N, Askim T, Anke A. Arm use in patients with subacute stroke monitored by accelerometry: association with motor impairment and influence on self-dependence. J Rehabil Med 2011 Mar;43(4):299-304 [FREE Full text] [doi: 10.2340/16501977-0676] [Medline: 21347506]
83. Ballester BR, Lathe A, Duarte E, Duff A, Verschure PF. A wearable bracelet device for promoting arm use in stroke patients. In: Proceedings of the 3rd International Congress on Neurotechnology, Electronics and Informatics. 2015 Presented at: NEUROTECHNIX '15; November 16-17, 2015; Lisbon, Portugal p. 24-31. [doi: 10.5220/0005662300240031]
84. Johnson MJ, Van der Loos HM, Burgar CG, Shor P, Leifer LJ. Experimental results using force-feedback cueing in robot-assisted stroke therapy. IEEE Trans Neural Syst Rehabil Eng 2005 Sep;13(3):335-348. [doi: 10.1109/TNSRE.2005.850428] [Medline: 16200757]
85. Hung JW, Chou CX, Chang YJ, Wu CY, Chang KC, Wu WC, et al. Comparison of Kinect2Scratch game-based training and therapist-based training for the improvement of upper extremity functions of patients with chronic stroke: a randomized controlled single-blinded trial. Eur J Phys Rehabil Med 2019 Oct;55(5):542-550 [FREE Full text] [doi: 10.23736/S1973-9087.19.05598-9] [Medline: 30781936]
86. Bakhti KK, Laffont I, Muthalib M, Froger J, Mottet D. Kinect-based assessment of proximal arm non-use after a stroke. J Neuroeng Rehabil 2018 Nov 14;15(1):104 [FREE Full text] [doi: 10.1186/s12984-018-0451-2] [Medline: 30428896]
87. Johnson MJ, Wisneski KJ, Hermsen A, Smith RO, Walton T, Hingtgen B, et al. Kinematic implications of learned non-use for robotic therapy. In: Proceedings of the 9th International Conference on Rehabilitation Robotics. 2005 Presented at: ICORR '05; June 28-July 1, 2005; Chicago, IL, USA p. 70-73. [doi: 10.1109/icorr.2005.1501054]
88. Johnson M, Paranjape R, Strachota E, Tchekanov G, McGuire J. Quantifying learned non-use after stroke using unilateral and bilateral steering tasks. IEEE Int Conf Rehabil Robot 2011;2011:5975457. [doi: 10.1109/ICORR.2011.5975457] [Medline: 22275655]
89. Barth J, Geed S, Mitchell A, Lum PS, Edwards DF, Dromerick AW. Characterizing upper extremity motor behavior in the first week after stroke. PLoS One 2020 Aug 10;15(8):e0221668 [FREE Full text] [doi: 10.1371/journal.pone.0221668] [Medline: 32776927]
90. Schwarz A, Bhagubai MM, Wolterink G, Held JP, Luft AR, Veltink PH. Assessment of upper limb movement impairments after stroke using wearable inertial sensing. Sensors (Basel) 2020 Aug 24;20(17):4770 [FREE Full text] [doi: 10.3390/s20174770] [Medline: 32846958]
91. Cai S, Li G, Su E, Wei X, Huang S, Ma K, et al. Real-Time Detection of Compensatory Patterns in Patients With Stroke to Reduce Compensation During Robotic Rehabilitation Therapy. IEEE J Biomed Health Inform 2020 Sep;24(9):2630-2638. [doi: 10.1109/JBHI.2019.2963365] [Medline: 31902785]
92. Taati B, Wang R, Huq R, Snoek J, Mihailidis A. Vision-based posture assessment to detect and categorize compensation during robotic rehabilitation therapy. In: Proceedings of the 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics. 2012 Presented at: BioRob '12; June 24-27,2012; Roma, Italy p. 1607-1613. [doi: 10.1109/biorob.2012.6290668]
93. Otaka E, Otaka Y, Kasuga S, Nishimoto A, Yamazaki K, Kawakami M, et al. Clinical usefulness and validity of robotic measures of reaching movement in hemiparetic stroke patients. J Neuroeng Rehabil 2015 Aug 12;12:66 [FREE Full text] [doi: 10.1186/s12984-015-0059-8] [Medline: 26265327]
94. Patrizi A, Pennestrì E, Valentini PP. Comparison between low-cost marker-less and high-end marker-based motion capture systems for the computer-aided assessment of working ergonomics. Ergonomics 2016;59(1):155-162. [doi: 10.1080/00140139.2015.1057238] [Medline: 26043178]
95. Nordin N, Xie SQ, Wünsche B. Assessment of movement quality in robot- assisted upper limb rehabilitation after stroke: a review. J Neuroeng Rehabil 2014 Sep 12;11:137 [FREE Full text] [doi: 10.1186/1743-0003-11-137] [Medline: 25217124]
96. Chauvette A, Schick-Makaroff K, Molzahn AE. Open data in qualitative research. Int J Qual Methods 2019 Jan 22;18:160940691882386. [doi: 10.1177/1609406918823863]
97. What is Open Data? Open Data Handbook. URL: https://opendatahandbook.org/guide/en/what-is-open-data/ [accessed 2022-04-12]
98. Tri-Agency Statement of Principles on Digital Data Management. Government of Canada. 2021. URL: https://www.ic.gc.ca/eic/site/063.nsf/eng/h_83F7624E.html [accessed 2022-04-12]

Abbreviations

AI: artificial intelligence
CEBM: Centre for Evidence-Based Medicine
EEG: electroencephalogram
FMA: Fugl–Meyer Assessment
fMRI: functional magnetic resonance imaging
GRASP: Graded Repetitive Arm Supplementary Program
IMU: inertial measurement unit
LNU: learned nonuse

XSL•FO
RenderX

ML: machine learning
MR: mixed reality
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RQ: research question
sEMG: surface electromyogram
SVM: support vector machine
UE: upper extremity
VR: virtual reality

Edited by A Mavragani; submitted 16.10.21; peer-reviewed by BF Kocyigit, D Parry; comments to author 30.01.22; revised version received 25.03.22; accepted 10.04.22; published 13.06.22.

Please cite as:
Wang X, Fu Y, Ye B, Babineau J, Ding Y, Mihailidis A
Technology-Based Compensation Assessment and Detection of Upper Extremity Activities of Stroke Survivors: Systematic Review
J Med Internet Res 2022;24(6):e34307
URL: https://www.jmir.org/2022/6/e34307
doi:10.2196/34307
PMID:35699982

©Xiaoyi Wang, Yan Fu, Bing Ye, Jessica Babineau, Yong Ding, Alex Mihailidis. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.06.2022. 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

The Effects of Nonclinician Guidance on Effectiveness and Process Outcomes in Digital Mental Health Interventions: Systematic Review and Meta-analysis

Calista Leung[1], BA; Julia Pei[1], BSc; Kristen Hudec[1], PhD; Farhud Shams[1], BSc; Richard Munthali[1], PhD; Daniel Vigo[1], DrPH, MD

Department of Psychiatry, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Daniel Vigo, DrPH, MD
Department of Psychiatry
Faculty of Medicine
University of British Columbia
2255 Wesbrook Mall
Vancouver, BC, V6T2B8
Canada
Phone: 1 6048228048
Email: daniel.vigo@ubc.ca

Abstract

Background: Digital mental health interventions are increasingly prevalent in the current context of rapidly evolving technology, and research indicates that they yield effectiveness outcomes comparable to in-person treatment. Integrating professionals (ie, psychologists and physicians) into digital mental health interventions has become common, and the inclusion of guidance within programs can increase adherence to interventions. However, employing professionals to enhance mental health programs may undermine the scalability of digital interventions. Therefore, delegating guidance tasks to paraprofessionals (peer supporters, technicians, lay counsellors, or other nonclinicians) can help reduce costs and increase accessibility.

Objective: This systematic review and meta-analysis evaluates the effectiveness, adherence, and other process outcomes of nonclinician-guided digital mental health interventions.

Methods: Four databases (MEDLINE, Embase, CINAHL, and PsycINFO) were searched for randomized controlled trials published between 2010 and 2020 examining digital mental health interventions. Three journals that focus on digital intervention were hand searched; gray literature was searched using ProQuest and the Cochrane Central Register of Control Trials (CENTRAL). Two researchers independently assessed risk of bias using the Cochrane risk-of-bias tool version 2. Data were collected on effectiveness, adherence, and other process outcomes, and meta-analyses were conducted for effectiveness and adherence outcomes. Nonclinician-guided interventions were compared with treatment as usual, clinician-guided interventions, and unguided interventions.

Results: Thirteen studies qualified for inclusion. Nonclinician-guided interventions yielded higher posttreatment effectiveness outcomes when compared to conditions involving control programs (eg, online psychoeducation and monitored attention control) or wait-list controls (k=7, Hedges g=–0.73; 95% CI –1.08 to –0.38). There were also significant differences between nonclinician-guided interventions and unguided interventions (k=6, Hedges g=–0.17; 95% CI –0.23 to –0.11). In addition, nonclinician-guided interventions did not differ in effectiveness from clinician-guided interventions (k=3, Hedges g=0.08; 95% CI –0.01 to 0.17). These results suggest that guided digital mental health interventions are helpful to improve mental health outcomes regardless of the qualifications of the individual performing the intervention, and that the presence of a nonclinician guide improves effectiveness outcomes compared to having no guide. Nonclinician-guided interventions did not yield significantly different adherence outcomes when compared with unguided interventions (k=3, odds ratio 1.58; 95% CI 0.51 to 4.92), although a general trend of improved adherence was observed within nonclinician-guided interventions.

Conclusions: Integrating paraprofessionals and nonclinicians appears to improve the outcomes of digital mental health interventions, and may also enhance adherence outcomes (though this trend was nonsignificant). Further research should focus on the specific types of tasks these paraprofessionals can successfully provide (ie, psychosocial support, therapeutic alliance, and technical augmentation) and their associated outcomes.

XSL•FO
RenderX

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42020191226; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=191226

(J Med Internet Res 2022;24(6):e36004) doi:10.2196/36004

KEYWORDS

digital mental health; nonclinician guidance; e-Mental health intervention; internet-based intervention; mental health; task shifting; digital health; digital health intervention; patient outcome

Introduction

The 2017 World Psychiatric Association-Lancet Psychiatry Commission on the Future of Psychiatry highlighted digital psychiatry and the reform of traditionally structured mental health services as key priority areas for the future of the field [1]. Digital mental health interventions (or e-mental health interventions) have become increasingly prevalent in recent years, and research suggests that these interventions have similar effectiveness as in-person mental health treatment [2]. These interventions have been effective in addressing a range of mental health concerns and can reduce the severity of depression [3], anxiety, and stress, reduce eating disorder symptoms, improve social well-being [4], and reduce alcohol consumption [5].

In addition to generating positive health outcomes, offering mental health treatment through digital platforms offers several advantages over brick-and-mortar formats. A digital intervention's inherent scalability enhances social welfare, protects patients from stigma and discrimination, and allows for low- and middle-income countries or geographically inaccessible areas to deploy critical mental health care that would otherwise be impractical due to insufficiencies in service infrastructure [6]. It can also be of use in higher-income countries, where it can provide increased convenience and accessibility for populations wishing to remain anonymous due to mental health stigma, reduce costs, broaden the reach of treatment, and increase the flexibility of treatment [7,8]. Digital interventions can also increase willingness to use mental health services: a study of US soldiers reported that 33% of those unwilling to utilize in-person counselling were willing to utilize a technology-based mental health treatment [9].

Digital mental health interventions have become widely available, but adherence has been poor [10]. Low adherence may subvert the effectiveness of digital mental health tools [11]. Implementing human support for digital interventions may offer a solution by improving adherence and effectiveness outcomes; this improvement may be mediated by the increased accountability that coaches provide through assistance, support, and scheduled contacts [12].

While human support is often provided by clinicians with positive effects [13,14], integrating professional clinicians into digital interventions can be costly and resource intensive. Engaging nonclinicians, such as lay workers and peers, offers a cost-effective way to address the gap in treatment [1]; shifting certain tasks that a professional would normally provide (such as developing a therapeutic alliance, providing weekly reminders for program completion, or general administrative tasks) onto a lesser-trained nonclinician coach can reduce costs and enable scaling up of digital interventions. The literature suggests this strategy can be effective; a meta-analysis of digital interventions for anxiety disorders did not identify significant differences in treatment outcomes between coaches of varying qualifications or levels of training [15]. Further, a systematic review of peer-to-peer interactions in digital interventions reported that peer support yielded positive effects on effectiveness and adherence outcomes alongside increased perceptions of social support for individuals with psychotic disorders [16]. Therefore, it seems intuitive to utilize paraprofessionals or peers to administer certain forms of support.

Despite the abundance of research on clinician-guided digital mental health interventions and studies suggesting the benefits of integrating nonclinicians, the pooled effects of nonclinician-guided digital interventions on a broader range of mental health and substance use issues do not appear to have been formally evaluated. As such, we conducted a systematic literature review and meta-analysis examining the effectiveness, adherence, and other process outcomes of nonclinician-guided digital mental health interventions compared to clinician-guided and unguided digital mental health interventions and to treatment as usual.

Methods

Inclusion and Exclusion Criteria

Randomized controlled trials (RCTs) qualified for inclusion if (1) they evaluated a digital intervention addressing clinical or subthreshold mental health, substance use–related issues, or direct determinants of these issues; (2) the digital intervention targeted the mental health of the individual receiving the intervention (eg, parenting interventions targeting the mental health of the child were excluded); (3) the digital intervention targeted primary mental health outcomes (as opposed to mental health outcomes secondary to physical conditions); (4) the digital intervention was supported by a nonclinician (eg, a peer, research assistant, or other layperson); (5) the control groups were (a) offered an unguided intervention, (b) offered clinician-guided intervention (ie, by a psychiatrist, psychologist, therapist, social worker, graduate student in a mental health–related field, or student completing clinical practicum training), (c) offered an in-person intervention, (d) put on a wait-list for a digital intervention or offered any form of "treatment as usual," or (e) offered an active control intervention (eg, monitored attention control or informative emails); (6) they included subjects between 16 and 64 years old; and (7) they reported effectiveness, adherence, or other process outcomes as primary outcomes. The inclusion criteria were piloted on small samples of studies and refined accordingly. Any disagreements were resolved through discussion or consultation

XSL•FO
RenderX

with a third researcher (DV). Only English-language or English-translated publications were included.

Search Strategy

A systematic search of literature published between 2010 and 2020 was conducted in July 2020. The publication time frame was selected to ensure included technologies were current rather than outdated (eg, video conferencing vs CD-ROM); thus, the findings are applicable to the current landscape of digital intervention research. Four databases (MEDLINE, Embase, CINAHL and PsycINFO) were searched using MeSH terms, keywords, and text (Multimedia Appendix 1).

Three theme-specific journals (Internet Interventions, Lancet Digital Health and the Journal of Medical Internet Research) were also hand searched. ProQuest and the Cochrane Central Register of Controlled Trials (CENTRAL) were searched for gray literature. Forward and backward reference chaining of included studies was performed and relevant reviews found through screening were searched for pertinent papers. Emails were sent to authors of relevant protocols and conference proceedings to ascertain whether an RCT had been conducted. The review protocol was registered on the International Prospective Register of Systematic Reviews (PROSPERO) before data extraction was initiated (CRD42020191226).

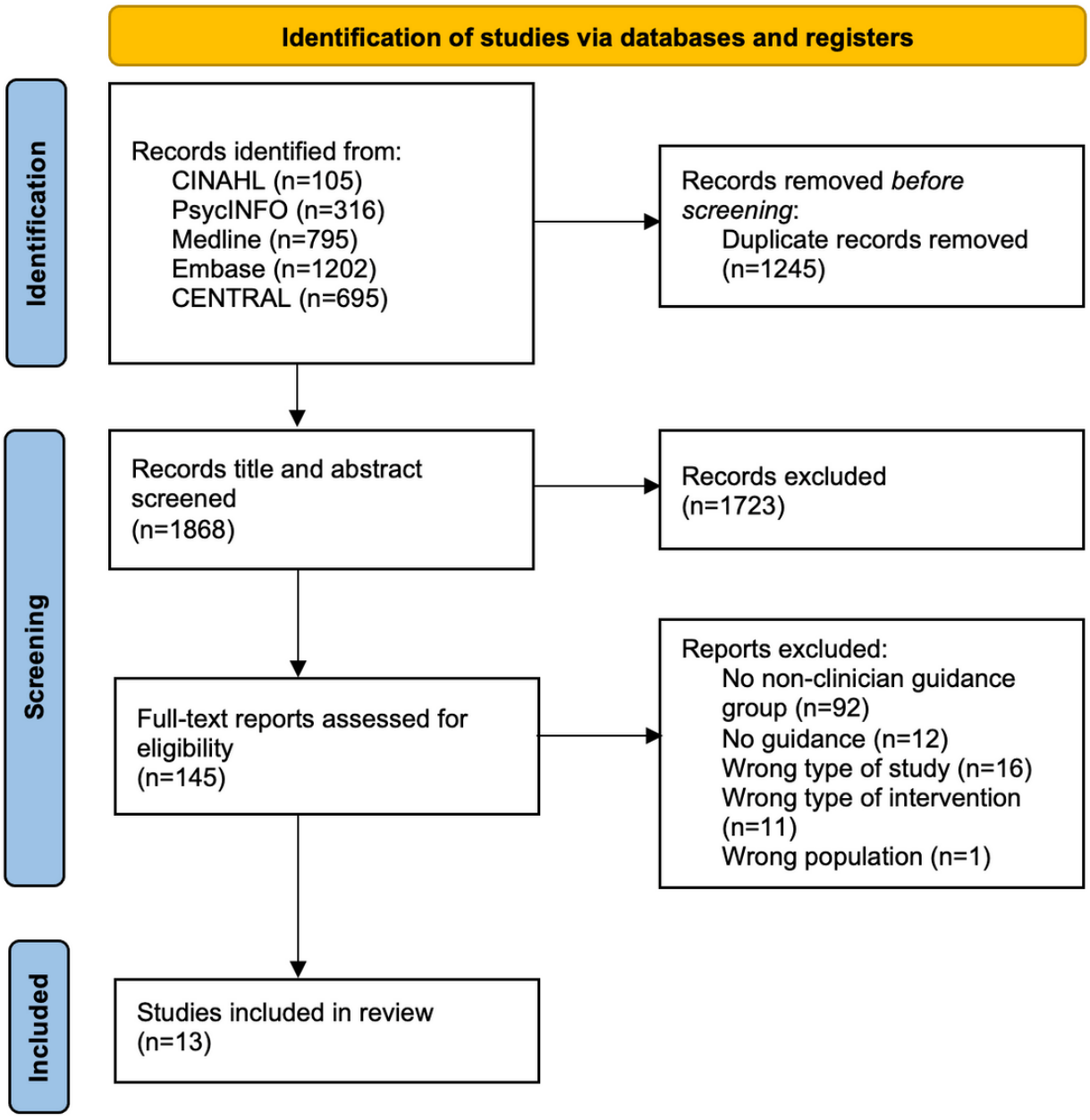
Study Selection

Titles and abstracts were independently screened by 2 researchers (CL and JP), then full text reports were independently evaluated by the same 2 researchers. Conflicts were resolved through discussion or, when needed, consultation with a third researcher (DV). Covidence, a web-based screening tool (Veritas Health Innovation), was used to facilitate collaborative screening [17].

Through searching, 3113 studies were identified. After deduplication, titles and abstracts of 1868 studies and full texts of 145 studies were screened. Thirteen studies qualified for inclusion. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram is presented in Figure 1.

XSL•FO
RenderX

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



Data Extraction and Study Characteristics

CL extracted data from the included articles (N=13), and JP validated the extracted data. Disagreements were resolved through discussion. Extracted data included intervention name, location, duration, design, arms, sample size, targeted mental health problem or disorder, theoretical model, nonclinician guide qualification, effectiveness outcomes, adherence outcomes, process outcomes, and results for these outcomes (results are shown in Table 1; additional information is shown in Multimedia Appendix 2). There were 3227 participants across the 13 studies. Treatment durations ranged from 4 to 12 weeks and sample sizes ranged from 30 to 1405 participants. The majority of interventions targeted mood and anxiety disorders (n=7). Other studies targeted well-being (n=1), stress (n=1), posttraumatic stress disorder (n=1), obsessive compulsive disorder (n=1), bipolar disorder (n=1), and substance use (n=2). Cognitive behavioral therapy was the most common theoretical model underpinning the interventions.

XSL•FO
RenderX

Table 1. Summary of results.

Study, year	Results
An et al, 2013 [18]	
Targeted disorder	Substance use (smoking)
Subjects for each study condition, n	Nonclinician, 456; unguided, 473; control; 476
Effectiveness outcome (30-day smoking abstinence), %	Nonclinician, 14%; unguided, 11%; control, 9%
Arjadi et al, 2018 [19]	
Targeted disorder	Depression
Subjects for each study condition, n	Nonclinician, 159; control, 154
Effectiveness outcome (Patient Health Questionnaire-9 score), mean (SD)	Nonclinician, 8.5 (5.74); control, 10.83 (6.21)
Day et al, 2013 [20]	
Targeted disorder	Depression
Subjects for each study condition, n	Nonclinician, 33; control (delayed access), 33
Effectiveness outcome (Depression, Anxiety and Stress Scale depression score), mean (SD)	Nonclinician, 10.43 (4.49); control, 14.6 (9.51)
Adherence outcome (completion of all modules), %	Nonclinician, 61%; control, N/A[a] (adherence outcomes unreported)
Process outcome	Usefulness
Results	The average usefulness rating of the overall modules was 6.78/10 (ranging from 1, “not useful at all,” to 10, “extremely useful”).
Dirkse et al, 2020 [21]	
Targeted disorder	Depression
Subjects for each study condition, n	Nonclinician, 41; unguided, 42
Effectiveness outcome (Patient Health Questionnaire-9 score), mean (SD)	Nonclinician, 4.83 (2.7); unguided, 5.51 (4.5)
Adherence outcome (completion of all modules), %	Nonclinician, 93%; unguided, 81%
Process outcome	Satisfaction
Results	A total of 85% of unguided and 90% of nonclinician-guided participants were either “satisfied” or “very satisfied” with the course (no significant difference), 93% of unguided and 100% of nonclinician-guided participants were either “satisfied” or “very satisfied” with the quality of the lessons and the materials (no significant difference); nonclinician-guided participants had significantly higher levels of satisfaction with the level of support, though both groups had relatively high satisfaction (96% of participants overall were “satisfied” or “very satisfied”).
Farrer et al, 2011 [22]	
Targeted disorder	Depression
Subjects for each study condition, n	Nonclinician, 41; unguided, 38; control, 35
Effectiveness outcome (Center for Epidemiologic Studies Depression Scale score), mean (SD)	Nonclinician, 21 (12.4); unguided, 24.4 (13.6); control, 35.1 (13.9)
Adherence outcome (minimum dose: 3/5 modules), %	Nonclinician, 37.7%; unguided, 31.6%; control, N/A (received no intervention)
Adherence outcome (completion of all modules), %	Nonclinician, 17.8%; unguided, 15.8%; control, N/A (received no intervention)
Flynn et al, 2020 [23]	
Targeted disorder	Mental well-being
Subjects for each study condition, n	Nonclinician, 30; unguided, 30

XSL•FO
RenderX

Study, year	Results
Effectiveness outcome (Warwick-Edinburgh Mental Wellbeing Scale score), mean (SD)	Nonclinician, 48.43 (12.66); unguided, 42.88 (9.66)
Adherence outcome (completion of all modules), %	Nonclinician, 52%; unguided, 43%
Heber et al, 2016 [24]	
Targeted disorder	Stress
Subjects for each study condition, n	Nonclinician, 132; control (delayed access), 132
Effectiveness outcome (Perceived Stress Scale-10 score), mean (SD)	Nonclinician, 17.88 (6.17); control, 22.96 (6.07)
Adherence outcome (completion of all modules), %	Nonclinician, 70.5%; control, N/A (adherence outcomes unreported)
Process outcome	Satisfaction
Results	A total of 92.2% of participants were “satisfied in an overall, general sense” (ie, either “very satisfied” or “mostly satisfied”).
Kobak et al, 2015 [25]	
Targeted disorder	Obsessive compulsive disorder
Subjects for each study condition, n	Clinician, 31; nonclinician, 28; unguided, 28
Effectiveness outcome (Yale Brown Obsessive Compulsive Scale score), mean (SD)	Clinician, 15.32 (7.04); nonclinician, 15.61 (5.88); unguided, 16.32 (6.97)
Process outcomes	Satisfaction, usability
Results	A total of 98% of participants “agreed” or “strongly agreed” with the statement that “they were satisfied with bt steps.” For usability, the mean total system usability score was 83.5/100 (between “good” and “excellent”).
Possemato et al, 2019[b] [26]	
Targeted disorder	Posttraumatic stress disorder and hazardous drinking
Subjects for each study condition, n	Nonclinician, 15; unguided, 15
Effectiveness outcome (Posttraumatic Stress Disorder Checklist—Military score), mean (SD)	Nonclinician, 41.78 (14.90); unguided, 43.16 (13.42)
Process outcome	Satisfaction
Results	A total of 78% of participants were “very satisfied.”
Proudfoot et al, 2012 [27]	
Targeted disorder	Bipolar disorder (perception of illness)
Subjects for each study condition, n	Nonclinician, 139; unguided, 141; control, 139
Adherence outcome (minimum dose; 4/8 module workbooks)	Nonclinician, 79.9%; unguided, 69.1%; control, N/A (received no intervention)
Adherence outcome (completion of all modules)	38.8% across 3 groups
Robinson et al, 2010 [28]	
Targeted disorder	Generalized anxiety disorder
Subjects for each study condition, n	Clinician, 47; nonclinician, 50; control (delayed access), 48
Effectiveness outcome (General Anxiety Disorder-7 score), mean (SD)	Clinician, 5.55 (4.73); nonclinician, 6.02 (3.43); control, 11.25 (4.70)
Adherence outcome (completion of all modules), %	Clinician, 74%, nonclinician, 80%; control, N/A (received no intervention)
Process outcome	Satisfaction
Results	A total of 87% of participants in the nonclinician-guided and clinician-guided groups were either “very satisfied” or “mostly satisfied” with the overall program (no significant difference).

XSL•FO
RenderX

Study, year	Results
Rosso et al, 2017 [29]	
Targeted disorder	Depression
Subjects for each study condition, n	Nonclinician, 37; control, 40
Effectiveness outcome (Hamilton Depression Rating Scale-17 score), mean (SD)	Nonclinician, 9.17 (6.92), control, 14.05 (5.34)
Adherence outcome (completion of all modules), %	Nonclinician, 92%; control, 75%
Titov et al, 2010 [30]	
Targeted disorder	Depression
Subjects for each study condition, n	Clinician, 46; nonclinician, 41; control, 40
Effectiveness outcome (Beck Depression Inventory-II score), mean (SD)	Clinician, 14.59 (11.12); nonclinician, 15.29 (9.81); control, 26.15 (10.14)
Adherence outcome (completion of all modules), %	Clinician, 80%; nonclinician, 80%; control, N/A (adherence outcomes unreported)
Process outcome	Satisfaction
Results	A total of 87% of participants in the nonclinician-guided or clinician-guided groups were either "very satisfied" or "mostly satisfied" with the overall program (no significant difference).

[a]N/A: not applicable.

[b]Possemato reported a nonclinician intervention retention rate of 93% and unguided intervention retention rate of 73% but did not define "intervention retention."

Quality Assessment

Two researchers (CL and JP) independently assessed risk of bias using Version 2 of the Cochrane risk-of-bias tool version 2 (RoB 2) [31]. Disagreements were resolved through discussion with a third researcher (DV). The RoB 2 evaluates the risk of bias associated with randomization, deviation from the intended intervention, missing outcome data, outcome measurement, and selection of the reported result. Each domain was assigned a judgment of "low risk of bias," "some concerns," or "high risk of bias."

Outcomes

The effectiveness outcomes described changes in mental health symptomology or substance use behaviors. Five studies included primary effectiveness outcomes, so 2 mental health clinicians were consulted in developing a hierarchy of outcomes [32]. When multiple mental health concerns were fully assessed as primary outcomes, the clinical metric that was reported as a primary outcome (eg, obsessive compulsive disorder over stress) among a greater number of studies was selected. When multiple instruments were used, the clinical outcomes were prioritized and reported (eg, Posttraumatic Stress Disorder Checklist-Military, which assesses posttraumatic stress disorder, was selected over the World Health Organization Quality of Life Questionnaire, which assesses quality of life). When multiple clinical instruments were reported, the most thorough instrument was reported (eg, Beck Depression Inventory, a 21-item inventory, was selected over the Patient Health Questionnaire, a 9-item inventory).

Adherence outcomes were defined as the proportion of participants that either fully completed the intervention or completed a defined minimum dose of the intervention; both full completion and minimum dose completion outcomes were included in the adherence meta-analysis due to the small number of studies reporting minimum dose adherence. Process outcomes consisted of participant satisfaction, intervention usefulness, and digital tool usability.

Data Analysis

A random effects model was used to conduct all meta-analyses [33]. This model assumes a distribution of true effect sizes, accounting for the different populations that each publication studied [32]. Outcomes were analyzed using the meta [34], metafor [35], and esc [36] packages in RStudio (version 3.6.2; R Foundation) (Multimedia Appendix 3 includes the full code).

Three meta-analyses of effectiveness outcomes were conducted: nonclinician-guided interventions versus clinician-guided interventions [25,28,30], nonclinician-guided interventions versus unguided interventions [18,21-23,25,26], and nonclinician-guided interventions versus controls (ie, wait-list or monitored attention control) [18-20,22,24,28-30]. Unguided interventions provided the same content as clinician or nonclinician-guided interventions (without the guide component), whereas control programs may not have included an intervention (ie, they used a wait-list) or may have provided different content than was utilized in the intervention arm. This meta-analytic approach avoids conflating active treatment arms with wait-list controls and more clearly elucidates the effects of nonclinician guidance. Meta-analyses of posttreatment effects were conducted for all 3 comparisons, and meta-analyses of follow-up effects were conducted for the nonclinician-guided intervention versus unguided intervention and nonclinician-guided intervention versus control comparisons. Although both posttreatment standardized mean difference (SMD) and pretest-posttest control group (d_{ppc2}) [37] effect

XSL•FO
RenderX

sizes have been utilized in similar meta-analyses [38,39], we determined that the posttreatment SMD effect size was most appropriate due to the lack of pre-post correlation values available from the included studies and the criticisms of pre-post effect size methods [40]. A sensitivity analysis was conducted comparing the 2 methods and resulted in the same pattern of findings. Hedges *g* effect sizes were used alongside their respective 95% CIs to correct for small sample sizes [41] and were interpreted according to recommendations [42] (small effect: <0.20; medium effect: 0.21-0.50; and large effect: 0.51-0.80). When a high level of heterogeneity was observed in the meta-analysis of nonclinician-guided interventions and controls, a meta-regression evaluating the effects of the control group type (wait-list vs control intervention) was conducted to determine whether these effects contributed to the heterogeneity.

Most studies reported reductions in symptoms as negative effects. A minority of papers reported effects that increased with symptom reduction, so these outcomes were reverse coded [18,23]. Proudfoot et al (2012) was excluded from the meta-analyses because only coefficients (rather than group scores) were reported [27], and the authors could not be reached to obtain the necessary data.

One meta-analysis was conducted for adherence outcomes, comparing nonclinician-guided interventions and unguided interventions, as there was insufficient data to conduct additional comparisons. Odds ratios were used as effect sizes [43]. Study selection for this meta-analysis was based on whether the results of the publication described full intervention completion rates in nonclinician-guided groups and unguided groups (3 publications satisfied these criteria; Table 1). No meta-analysis was conducted for other process outcomes (ie, satisfaction, usability, and usefulness) due to the small number of studies reporting these outcomes, but findings have been summarized below.

Results

Quality Assessment

RoB 2 was used to conduct an assessment of the methodological quality of the 13 included papers (Table 2). Separate assessments were conducted for effectiveness, adherence, and process outcomes (Multimedia Appendix 4).

Effectiveness outcomes were assessed for 12 papers: 4 studies scored “high risk,” 5 studies scored “some concerns,” and 4 studies scored “low risk.” High risk was most commonly driven by domain 3 (missing outcome data) and domain 4 (measurement of the outcome). High risk of bias was associated with domain 3 when experimenters did not adequately correct for bias stemming from missing data or did not describe doing so, since participants who completed follow-up measures were more likely to have more favorable efficacy outcomes than dropouts.

Adherence outcomes were assessed for the 12 papers that reported adherence outcomes. Five types of adherence outcomes were evaluated: completion of the whole intervention, completion of a minimum dose, percentage completion of each intervention module, mean number of modules completed, and intervention retention. Papers that reported multiple adherence outcomes were assessed for each adherence outcome reported, though each of these papers scored the same overall risk of bias for each adherence outcome. Two studies scored “high risk,” 7 studies scored “some concerns,” and 3 studies scored “low risk.” High risk was driven by domain 2 (deviations from intended interventions) and domain 5 (selection of the reported result).

Process outcomes (satisfaction, usefulness, and usability) were assessed for 7 papers; 6 studies scored “high risk” and 1 study scored “some concerns.” High risk was most associated with domain 3 (missing outcome data); since many of these outcomes were subjective in nature, participants who completed follow-up measures may have been more likely to rate these outcomes favorably than dropouts.

XSL•FO
RenderX

Table 2. Cochrane risk-of-bias tool version 2 summary.

Study, year	Randomization process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result	Overall
Effectiveness outcome assessments						
An et al, 2013 [18]	Low	Some concerns	High	Some concerns	High	High
Arjadi et al, 2018 [19]	Low	Low	Low	Low	Low	Low
Day et al, 2013 [20]	Low	Low	Low	Low	Some concerns	Some concerns
Dirkse et al, 2020 [21]	Low	Low	Low	Low	Some concerns	Some concerns
Farrer et al, 2011 [22]	Some concerns	Low	Low	Low	High	High
Flynn et al, 2020 [23]	Low	High	High	Low	Some concerns	High
Heber et al, 2016 [24]	Low	Low	Low	Low	Some concerns	Some concerns
Kobak et al, 2015 [25]	Some concerns	High	Low	Low	Some concerns	High
Possemato et al, 2019 [26]	Low	Low	Low	Some concerns	Some concerns	Some concerns
Robinson et al, 2010 [28]	Some concerns	Low	Low	Some concerns	Low	Some concerns
Rosso et al, 2017 [29]	Low	Low	Low	Low	Some concerns	Some concerns
Titov et al, 2010 [30]	Some concerns	Low	High	Low	Some concerns	High
Adherence outcome assessments—completion of whole intervention						
Day et al, 2013 [20]	Low	Low	Low	Low	Low	Low
Dirkse et al, 2020 [21]	Some concerns	Low	Low	Low	Low	Some concerns
Farrer et al, 2011 [22]	Some concerns	Low	Low	Low	Low	Some concerns
Flynn et al, 2020 [23]	Low	High	Low	Low	Low	High
Heber et al, 2016 [24]	Low	Some concerns	Low	Low	Low	Some Concerns
Proudfoot et al, 2012 [27]	Low	Low	Low	Low	High	High
Robinson et al, 2010 [28]	Some concerns	Low	Low	Low	Low	Some concerns
Rosso et al, 2017 [29]	Low	Low	Low	Low	Low	Low
Titov et al, 2010 [30]	Some concerns	Low	Low	Low	Low	Some concerns
Adherence outcome assessments—completion of minimum dose						
An et al, 2013 [18]	Low	Some concerns	Low	Low	Low	Some concerns
Farrer et al, 2011 [22]	Some concerns	Low	Low	Low	Low	Some concerns
Proudfoot et al, 2012 [27]	Low	Low	Low	Low	High	High
Adherence outcome assessments—percentage completion of each intervention module						
Arjadi et al, 2018 [19]	Low	Low	Low	Low	Low	Low
Farrer et al, 2011 [22]	Some concerns	Low	Low	Low	Low	Some concerns
Adherence outcome assessments—mean number of modules completed						
Possemato et al, 2019 [26]	Low	Low	Low	Low	Some concerns	Some concerns
Process outcome assessments						
Day et al, 2013 [20]	Low	Low	High	High	Low	High

XSL•FO
RenderX

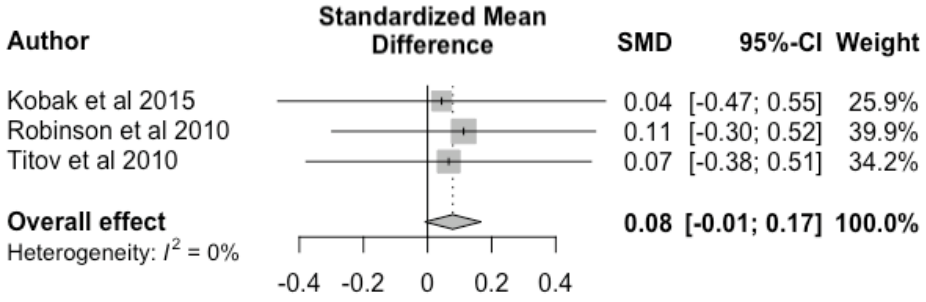
Study, year	Randomization process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result	Overall
Dirkse et al, 2020 [21]	Low	Low	Low	High	Some concerns	High
Heber et al, 2016 [24]	Low	Some concerns	High	Low	Low	High
Kobak et al, 2015 [25]	Some concerns	High	Low	Some concerns	Some concerns	High
Possemato et al, 2019 [26]	Low	Low	High	High	Low	High
Robinson et al, 2010 [28]	Some concerns	Low	High	Low	Low	High
Titov et al, 2010 [30]	Some concerns	Low	High	High	Some concerns	High

Primary Posttreatment Effectiveness Outcomes

Nonclinician Versus Clinician

The overall effect size from 3 studies was 0.08 (95% CI –0.01 to 0.17), indicating nonclinician-guided interventions did not significantly differ from clinician-guided interventions with respect to participant mental health outcomes. The distribution of effect sizes was homogeneous *(P*=.98) and is shown in Figure 2 as a forest plot.

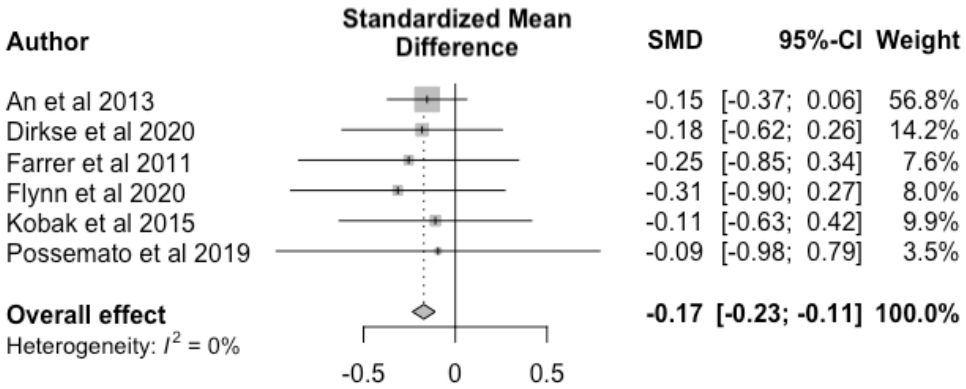
Figure 2. Nonclinician versus clinician, posttreatment. SMD: standardized mean difference.



Nonclinician Versus Unguided

The overall effect size (*k*=6, Hedges *g*=-0.17; 95% CI –0.23 to –0.11) between nonclinician-guided interventions and unguided interventions was significant. This small effect size indicates that digital mental health interventions were more effective when paraprofessionals or nonclinicians were involved in the intervention. The distribution of effect sizes was homogeneous *(P*=.99), ranging from –0.31 to –0.09, and is shown in Figure 3 as a forest plot.

Figure 3. Nonclinician versus unguided, posttreatment. SMD: standardized mean difference.



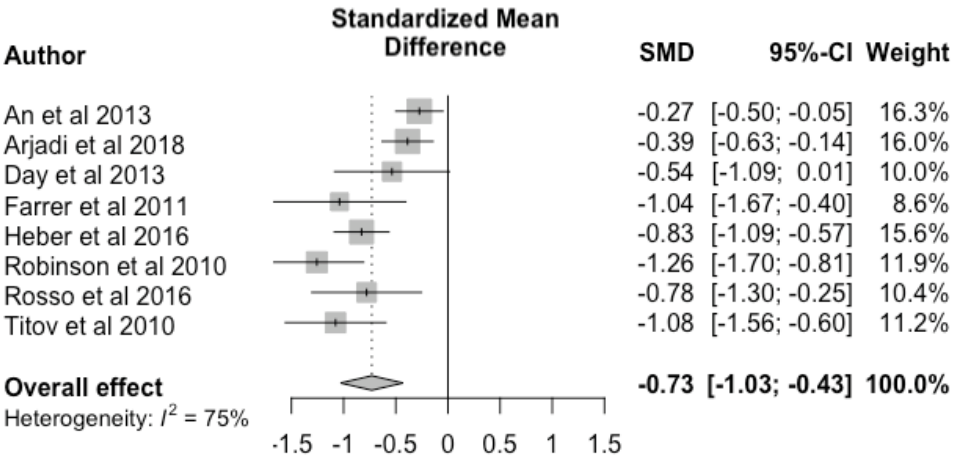
Nonclinician Versus Control

Based on 8 studies, the overall effect size was –0.73 (95% CI –1.08 to –0.38). This significant, large effect size indicates nonclinician-guided interventions yielded higher posttreatment effectiveness outcomes than control programs (eg, online psychoeducation and monitored attention control) or wait-list controls. The distribution of effect sizes was heterogeneous *(P*<.001), ranging from –1.26 to –0.27, and is shown in Figure 4 as a forest plot. The heterogeneity was further examined through a meta-regression using type of control (wait-list control, *k*=5 [20,22,24,28,30] or control intervention program,

XSL•FO
RenderX

k=3 [18,19,29]). Results from the meta-regression indicate that variability in the observed effect sizes can be explained by whether the study implemented a wait-list control or control intervention program (*k*=8, R^2=94.32%, *P*=.23).

Figure 4. Nonclinician versus control, posttreatment. SMD: standardized mean difference.

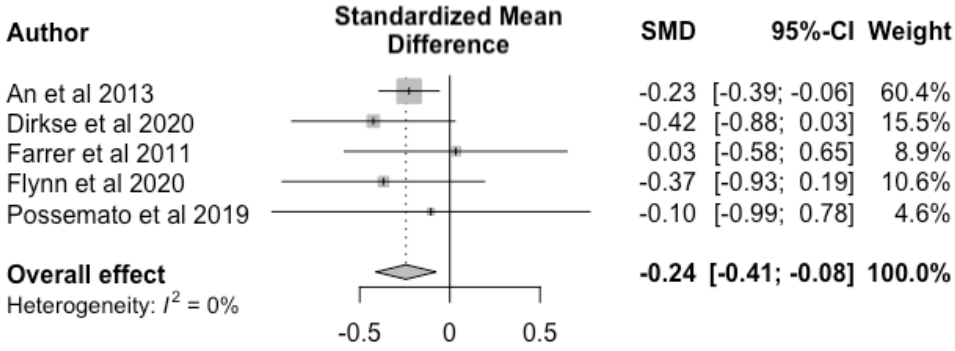


Follow-up Outcomes

Nonclinician Versus Unguided

Nonclinician-guided interventions yielded higher effectiveness outcomes than unguided interventions at follow-up, with a medium effect size (*k*=5, Hedges *g*=-0.24; 95% CI –0.41 to –0.08). The distribution of effect sizes was homogeneous (*P*=.79); results are shown in Figure 5 as a forest plot.

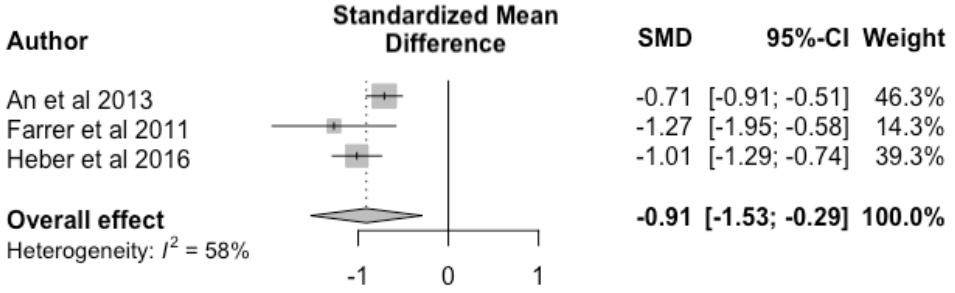
Figure 5. Nonclinician versus unguided, follow up. SMD: standardized mean difference.



Nonclinician Versus Control

Overall, nonclinician-guided interventions exhibited sustained improvement in effectiveness outcomes when compared to conditions involving wait-list controls and monitored attention controls at follow-up assessments; a large effect size was obtained (*k*=3, Hedges *g*=-0.91; 95% CI –1.53 to –0.29). Results are shown in Figure 6 as a forest plot. As a high level of heterogeneity was obtained, a meta-regression was conducted as a sensitivity analysis. The dependent variable was the effect size obtained from each study, and the explanatory variable was the type of control (wait-list control or control intervention program). The results indicated that all heterogeneity was accounted for by the type of control (*k*=3, R^2=100%, *P*=.50), but this result could have been influenced by the minimal number of studies.

Figure 6. Nonclinician versus control, follow up.



XSL•FO
RenderX

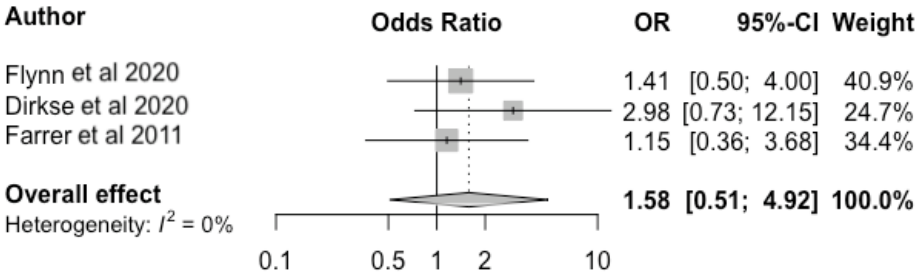
Adherence Outcomes

Of the 13 studies, 9 reported the percentage of participants who completed the intervention [18,20,21,23,24,27-30]; 2 studies reported the percentage of participants who completed a defined minimum dose [22,27]; 2 studies reported the percentage of participants who completed each module [21,24], and 1 study reported the "intervention retention" percentage [26]. There was wide variation in adherence rates between the studies for both minimum dose and full completion measures. Minimum dose completion rates ranged from 31.6% to 79.9% and intervention completion rates ranged from 15.8% to 93%.

Full completion adherence rates in nonclinician-guided and unguided groups were compared in 3 studies [21-23] and were therefore pooled in a meta-analysis; all 3 studies reported higher adherence rates in the nonclinician-guided groups. The meta-analysis indicated no significant effects on adherence outcomes in nonclinician versus unguided interventions (k=3, odds ratio 1.58 (95% CI 0.51 to 4.92)), although there was a general trend toward improved adherence outcomes when a nonclinician was involved (Figure 7).

Of the 2 studies that compared adherence rates in nonclinician-guided and clinician-guided groups, Robinson et al [28] reported higher adherence rates in the nonclinician-guided group, and Titov et al [30] reported the same adherence rates in each group. One study compared adherence rates in nonclinician-guided and monitored-attention control groups and reported higher adherence rates in the nonclinician-guided group [29]. Only 1 study reported the significance of the between-group difference and found that the nonclinician-guided group had significantly higher rates of intervention completion than the unguided group [28].

Figure 7. Adherence outcomes. OR: odds ratio.



Process Outcomes: Satisfaction, Usefulness, and Usability Results

Of the 13 studies, 6 evaluated participant satisfaction, 1 measured usefulness ratings of modules, and 1 measured system usability (Table 1). To measure participant satisfaction, 2 studies used questionnaires based on the Credibility/Expectancy Questionnaire [28,30], 2 studies used the Client Satisfaction Questionnaire [24,26], and 2 studies appeared to generate their own satisfaction measures [21,25]. The usefulness rating did not appear to be based on a preexisting scale [20]; system usability was measured using the System Usability Scale [25].

All studies reported that at least 78% of participants were satisfied with the intervention. Three studies compared satisfaction between groups. Two studies [28,30] found no significant difference in satisfaction between nonclinician-guided and clinician-guided groups. Dirkse et al [21] found no significant difference in intervention satisfaction between unguided and nonclinician-guided groups but reported that nonclinician-guided participants had significantly higher levels of satisfaction with the level of support. Day et al [20] reported a mean usefulness rating of 6.78/10 across guided and unguided groups. Kobak et al [25] reported a mean total system usability score of 83.5/100 across guided and unguided groups, which was between "good" and "excellent."

Discussion

Main Results

Guided Versus Unguided and Control Interventions

Our meta-analysis indicates that guided digital mental health interventions significantly improve effectiveness outcomes compared to both control (intervention programs and wait-list) and unguided interventions. These results align with a seminal systematic review of guided digital mental health interventions by Baumeister et al, which reported that guided interventions were more favorable than unguided interventions [13]. Two previous meta-analyses also concluded that significant improvements in effectiveness were associated with guide involvement [44,45]. It is interesting and noteworthy that our results align with these previous meta-analyses [13,44,45], as these studies examined digital mental health intervention research published from 2002 to 2013—a period of time when the technological landscape was vastly different from today. Collectively, these findings suggest that the beneficial effects of guidance in digital mental health interventions have been sustained through large shifts in both use of and attention to technology and come at a time when digital mental health interventions are critical to meet increasing need [14,46]. As additional digital interventions are designed and deployed, administrators, developers, and user groups (such as patients) must be aware of the potential contributions of guides and consider these benefits when attempting to optimize mental health intervention outcomes.

XSL•FO
RenderX

Nonclinician- Versus Clinician-Guided Interventions

Nonclinician-guided interventions were associated with greater effectiveness compared to unguided interventions, yet there was no significant difference between nonclinician and clinician guidance. Despite the scarcity of longer follow-up data, it also appears that the positive effects of nonclinician-guided interventions persist beyond the intervention period. Taken together, our findings suggest that the use of nonclinicians is a promising way of incorporating cost-effective guidance into digital mental health interventions; their involvement can improve mental health outcomes to a degree on par with that achieved by professional mental health guidance. Interventions with guidance have improved outcomes compared to those without guidance, and have lasting effects.

There is often an assumption that clinical intervention requires highly trained professionals to optimize outcomes, despite research suggesting that the presence of human support alone increases adherence to digital mental health interventions, thereby yielding improved efficacy and outcomes [47]. In line with our findings, the presence of a guide—clinician or nonclinician—is beneficial for evoking positive changes. These results align with a review by Baumeister et al [13] that reported that changes in symptom severity did not differ significantly in groups supported by guides with differing levels of qualifications (n=4). Although Baumeister et al [13] considered clinical psychology students and psychologists without specialized postgraduate training as guides with lower qualifications, our study limited the designation of lower-qualified (nonclinician) guides to true nonclinicians, meaning graduate students in a mental health field were excluded. Still, our results indicate that Baumeister's [13] findings (ie, that levels of effectiveness were comparable across levels of guide qualification) remain true with "lay" guides as well. This is particularly pertinent considering the push to increase the accessibility of mental health interventions (ie, through digital mental health platforms), given that these tools are likely to be more beneficial when supported by a guide. These results, therefore, show the possibility that larger-scale digital mental health interventions supported by personnel with lower levels of qualifications are feasible.

Adherence and Other Process Outcomes

With respect to adherence outcomes, the meta-analysis of the 3 studies we were able to pool showed no significant differences, although there appeared to be a trend toward higher adherence in the nonclinician-guided group relative to the unguided groups. The adherence results excluded in the meta-analysis were consistent with this trend. This is relatively unsurprising, given research suggesting that human support increases adherence by providing accountability [48]. However, there are limited studies reporting this metric, so additional information is needed.

With respect to other process outcomes, participants in all 6 studies that evaluated satisfaction reported high satisfaction across unguided, nonclinician-guided, and clinician-guided groups, though it is difficult to draw conclusions, as only 3 studies reported satisfaction in multiple groups. Furthermore, satisfaction measures included a heterogeneous landscape of satisfaction and usability scales, with many generated only for a specific study, and were prone to selection bias, as participants who are less satisfied with an intervention are more likely to drop out of the study. A more systematic understanding of how users perceive digital mental health interventions and which measures affect adherence would be gained if more studies reported standardized scales for process outcomes.

Limitations of the Literature and Future Directions

Some included studies lacked a robust description of the roles and qualifications of the nonclinician guides. Within and across studies, nonclinician guides may have received a wide range of training and undertaken a variety of roles. Therefore, overall conclusions will not capture the likely heterogeneous effects of varying types of nonclinician support. Notably, 1 paper [26] included guides who utilized a psychosocial support approach through divulging anecdotes and reflecting upon their own recovery story to personally connect with participants. All other studies included in our analyses appeared to employ a supportive accountability model in which guides established participant accountability by creating, revising, and monitoring adherence goals and progress [48]. The inconsistencies and lack of detailed descriptions of the tasks performed by guides challenged our evaluation of which nonclinician roles were most effective, but this limitation likely reflects the infancy of this line of research. As nonclinician guidance appears beneficial in this context, future examination of the support type and the nonclinician guides' training will be especially valuable in understanding how best to offer support that is effective and feasible within the digital mental health intervention format.

A similar issue (heterogeneity in definitions and measures) hampers the evaluation of adherence and other process outcomes. Intervention adherence differs from study attrition or dropout, as it refers to intervention uptake rather than study completion (eg, follow up). Study completion rates may not reflect the actual use of the intervention (eg, Christensen H et al [49]). The focus on adherence to digital interventions is related to whether the user engages with the tool, rather than whether they complete a follow-up assessment; this can be related more to study incentives than to intervention uptake. For this analysis, to minimize the risk of conflating intervention adherence with study completion, we defined adherence as the percentage of participants that completed all modules of the intervention, as it was the most-reported measure across the included studies. Two studies reported the percentage of participants who completed a "minimum dose" defined by the authors, and it has been posited that defining such minimum intended use may improve our understanding of intervention adherence [50]. Future studies should provide measures of intervention adherence that can be readily understood and differentiated from other variables, such as study completion.

Limitations of the Current Study

Our results should be interpreted with caution due to several limitations. First, our search identified only 13 studies that assessed the effects of nonclinician guidance in digital mental health interventions through RCTs. Though we attempted to minimize the risk of missing studies by using a wide range of terms, we may have missed relevant studies given the lack of consensus around the terminology of this emerging category of

XSL•FO
RenderX

nonclinician support. Further, our search was limited to English-language studies, which may have excluded studies that would have otherwise qualified for inclusion.

Another limitation was the inconsistency in methodology and the poor quality of many studies, which may hamper interpretation of results. Most studies were flagged as having "some concerns," which aligns with the findings of other digital mental health systematic reviews [4,51]. To mitigate this limitation, we provided a structured assessment of bias as a general picture of the quality of the included studies. The duration between posttreatment and follow-up assessments also varied, and a wide range of sample sizes was found in our search. Notably, some papers included upwards of 100 participants in each trial arm [18,19,24], while other papers included approximately 30 participants in each trial arm. We evaluated our meta-analytic results to ensure study size did not unduly influence or skew the overall findings, but future evaluations of digital mental health interventions should aim to include more participants, in addition to standardizing follow-up assessments, to accurately capture lasting effects of the intervention.

Finally, while there was wide variation in heterogeneity across meta-analyses, the nonclinician versus control meta-analyses at both posttreatment and follow-up time points had moderate to high heterogeneity values (I^2=75% and 58%, respectively) [52]. It appears the variation in control groups contributed to the heterogeneity; the type of control implemented by the study accounted for a large portion (94%) of heterogeneity. It is also important to note that we did not evaluate all possible explanatory variables through meta-regression, given our small number of eligible studies. Future meta-regressions should evaluate the effects of other factors, such as setting and population. Despite these limitations, our study provides valuable results in terms of next steps for this field of research, as well as allowing for a promising preliminary assessment of nonclinician guidance of digital interventions.

Conclusion

Digital mental health interventions have emerged as a promising means of providing more accessible mental health care. This review demonstrates that nonclinician guidance yields more improvement in effectiveness outcomes than unguided or control interventions, and that nonclinician guidance can generate effectiveness outcomes comparable to those of clinician guidance in the context of digital mental health interventions. These results are encouraging, as integrating nonclinician guidance can increase the scalability and cost efficiency of digital interventions to meet the current demand for support. In particular, nonclinicians such as peers or technicians are much more readily available than clinicians and may be perceived as more relatable (eg, through having lived experience with mental health difficulties) and approachable (eg, it may be less stigmatizing to talk with a peer than a professional) by individuals seeking support. Incorporating nonclinician guides may be an advantageous way in which to facilitate access to effective support, since health system administrators and funding agencies may be more responsive to interventions that are likely to optimize benefits (ie, improved individual and community health and reduction in use of other services) but require relatively minimal resource demands. Further studies investigating the effects of guide qualification on digital health intervention effectiveness and process outcomes are needed and should clearly describe the specific roles of the guides, compare different levels of nonclinician support (eg, technician guidance vs psychosocial support), investigate the contributing mechanisms, and examine implementation feasibility for different types of guides.

Acknowledgments

This project was funded by Health Canada. We would also like to thank University of British Columbia reference librarian Charlotte Beck for her guidance in developing our search strategy.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Medline search strategy.
[DOCX File , 14 KB - jmir_v24i6e36004_app1.docx]

Multimedia Appendix 2
Summary of accepted papers.
[DOCX File , 20 KB - jmir_v24i6e36004_app2.docx]

Multimedia Appendix 3
Meta-analysis R code.
[DOCX File , 17 KB - jmir_v24i6e36004_app3.docx]

Multimedia Appendix 4

XSL•FO
RenderX

Cochrane Risk of Bias 2 Graphical Representation.
[DOCX File , 2155 KB - jmir_v24i6e36004_app4.docx]

References

1. Bhugra D, Tasman A, Pathare S, Priebe S, Smith S, Torous J, et al. The WPA-Lancet Psychiatry Commission on the Future of Psychiatry. Lancet Psychiatry 2017 Oct;4(10):775-818. [doi: 10.1016/S2215-0366(17)30333-4] [Medline: 28946952]
2. van't Hof E, Cuijpers P, Stein DJ. Self-help and Internet-guided interventions in depression and anxiety disorders: a systematic review of meta-analyses. CNS Spectr 2009 Feb;14(2 Suppl 3):34-40. [doi: 10.1017/s1092852900027279] [Medline: 19238128]
3. Josephine K, Josefine L, Philipp D, David E, Harald B. Internet- and mobile-based depression interventions for people with diagnosed depression: A systematic review and meta-analysis. J Affect Disord 2017 Dec 01;223:28-40. [doi: 10.1016/j.jad.2017.07.021] [Medline: 28715726]
4. Harrer M, Adam SH, Baumeister H, Cuijpers P, Karyotaki E, Auerbach RP, et al. Internet interventions for mental health in university students: A systematic review and meta-analysis. Int J Methods Psychiatr Res 2019 Jun;28(2):e1759 [FREE Full text] [doi: 10.1002/mpr.1759] [Medline: 30585363]
5. Riper H, Blankers M, Hadiwijaya H, Cunningham J, Clarke S, Wiers R, et al. Effectiveness of guided and unguided low-intensity internet interventions for adult alcohol misuse: a meta-analysis. PLoS One 2014 Jun 17;9(6):e99912 [FREE Full text] [doi: 10.1371/journal.pone.0099912] [Medline: 24937483]
6. Carter H, Araya R, Anjur K, Deng D, Naslund JA. The emergence of digital mental health in low-income and middle-income countries: A review of recent advances and implications for the treatment and prevention of mental disorders. J Psychiatr Res 2021 Jan;133:223-246 [FREE Full text] [doi: 10.1016/j.jpsychires.2020.12.016] [Medline: 33360867]
7. Lal S, Adair CE. E-mental health: a rapid review of the literature. Psychiatr Serv 2014 Jan 01;65(1):24-32. [doi: 10.1176/appi.ps.201300009] [Medline: 24081188]
8. Price M, Yuen EK, Goetter EM, Herbert JD, Forman EM, Acierno R, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014;21(5):427-436 [FREE Full text] [doi: 10.1002/cpp.1855] [Medline: 23918764]
9. Wilson JAB, Onorati K, Mishkind M, Reger MA, Gahm GA. Soldier attitudes about technology-based approaches to mental health care. Cyberpsychol Behav 2008 Dec;11(6):767-769. [doi: 10.1089/cpb.2008.0071] [Medline: 18991533]
10. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
11. 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]
12. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
13. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — A systematic review. Internet Interv 2014 Oct;1(4):205-215 [FREE Full text] [doi: 10.1016/j.invent.2014.08.003]
14. Lakhtakia T, Torous J. Current directions in digital interventions for mood and anxiety disorders. Curr Opin Psychiatry 2022 Mar 01;35(2):130-135. [doi: 10.1097/YCO.0000000000000772] [Medline: 34966117]
15. Domhardt M, Geßlein H, von Rezori RE, Baumeister H. Internet- and mobile-based interventions for anxiety disorders: A meta-analytic review of intervention components. Depress Anxiety 2019 Mar;36(3):213-224. [doi: 10.1002/da.22860] [Medline: 30450811]
16. Biagianti B, Quraishi SH, Schlosser DA. Potential Benefits of Incorporating Peer-to-Peer Interactions Into Digital Interventions for Psychotic Disorders: A Systematic Review. Psychiatr Serv 2018 Apr 01;69(4):377-388 [FREE Full text] [doi: 10.1176/appi.ps.201700283] [Medline: 29241435]
17. Babineau J. Product Review: Covidence (Systematic Review Software). J Can Health Libr Assoc 2014 Aug 01;35(2):68. [doi: 10.5596/c14-016]
18. An LC, Demers MRS, Kirch MA, Considine-Dunn S, Nair V, Dasgupta K, et al. A randomized trial of an avatar-hosted multiple behavior change intervention for young adult smokers. J Natl Cancer Inst Monogr 2013 Dec;2013(47):209-215 [FREE Full text] [doi: 10.1093/jncimonographs/lgt021] [Medline: 24395994]
19. Arjadi R, Nauta MH, Scholte WF, Hollon SD, Chowdhary N, Suryani AO, et al. Internet-based behavioural activation with lay counsellor support versus online minimal psychoeducation without support for treatment of depression: a randomised controlled trial in Indonesia. Lancet Psychiatry 2018 Sep;5(9):707-716. [doi: 10.1016/S2215-0366(18)30223-2] [Medline: 30006262]
20. Day V, McGrath PJ, Wojtowicz M. Internet-based guided self-help for university students with anxiety, depression and stress: a randomized controlled clinical trial. Behav Res Ther 2013 Jul;51(7):344-351. [doi: 10.1016/j.brat.2013.03.003] [Medline: 23639300]

XSL•FO
RenderX

21. Dirkse D, Hadjistavropoulos H, Alberts NA, Karin E, Schneider L, Titov N, et al. Making Internet-delivered cognitive behaviour therapy scalable for cancer survivors: a randomized non-inferiority trial of self-guided and technician-guided therapy. J Cancer Surviv 2020 Apr 18;14(2):211-225. [doi: 10.1007/s11764-019-00810-9] [Medline: 31853727]
22. Farrer L, Christensen H, Griffiths KM, Mackinnon A. Internet-based CBT for depression with and without telephone tracking in a national helpline: randomised controlled trial. PLoS One 2011;6(11):e28099 [FREE Full text] [doi: 10.1371/journal.pone.0028099] [Medline: 22140514]
23. Flynn S, Hastings RP, Burke C, Howes S, Lunsky Y, Weiss JA, et al. Online Mindfulness Stress Intervention for Family Carers of Children and Adults with Intellectual Disabilities: Feasibility Randomized Controlled Trial. Mindfulness 2020 Jun 20;11(9):2161-2175 [FREE Full text] [doi: 10.1007/s12671-020-01436-0]
24. Heber E, Lehr D, Ebert D, Berking M, Riper H. Web-Based and Mobile Stress Management Intervention for Employees: A Randomized Controlled Trial. J Med Internet Res 2016 Jan 27;18(1):e21 [FREE Full text] [doi: 10.2196/jmir.5112] [Medline: 26818683]
25. Kobak KA, Greist R, Jacobi DM, Levy-Mack H, Greist JH. Computer-assisted cognitive behavior therapy for obsessive-compulsive disorder: a randomized trial on the impact of lay vs. professional coaching. Ann Gen Psychiatry 2015;14:10 [FREE Full text] [doi: 10.1186/s12991-015-0048-0] [Medline: 25722737]
26. Possemato K, Johnson EM, Emery JB, Wade M, Acosta MC, Marsch LA, et al. A pilot study comparing peer supported web-based CBT to self-managed web CBT for primary care veterans with PTSD and hazardous alcohol use. Psychiatr Rehabil J 2019 Sep;42(3):305-313 [FREE Full text] [doi: 10.1037/prj0000334] [Medline: 30489140]
27. Proudfoot J, Parker G, Manicavasagar V, Hadzi-Pavlovic D, Whitton A, Nicholas J, et al. Effects of adjunctive peer support on perceptions of illness control and understanding in an online psychoeducation program for bipolar disorder: a randomised controlled trial. J Affect Disord 2012 Dec 15;142(1-3):98-105. [doi: 10.1016/j.jad.2012.04.007] [Medline: 22858215]
28. Robinson E, Titov N, Andrews G, McIntyre K, Schwencke G, Solley K. Internet treatment for generalized anxiety disorder: a randomized controlled trial comparing clinician vs. technician assistance. PLoS One 2010 Jun 03;5(6):e10942 [FREE Full text] [doi: 10.1371/journal.pone.0010942] [Medline: 20532167]
29. Rosso IM, Killgore WDS, Olson EA, Webb CA, Fukunaga R, Auerbach RP, et al. Internet-based cognitive behavior therapy for major depressive disorder: A randomized controlled trial. Depress Anxiety 2017 Mar;34(3):236-245 [FREE Full text] [doi: 10.1002/da.22590] [Medline: 28009467]
30. Titov N, Andrews G, Davies M, McIntyre K, Robinson E, Solley K. Internet treatment for depression: a randomized controlled trial comparing clinician vs. technician assistance. PLoS One 2010 Jun 08;5(6):e10939 [FREE Full text] [doi: 10.1371/journal.pone.0010939] [Medline: 20544030]
31. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343(oct18 2):d5928-d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
32. Harrer M, Cuijpers P, Furukawa TA, Ebert DD. Doing Meta-Analysis with R: A Hands-On Guide. Boca Raton, FL: Chapmann & Hall/CRC Press; 2021.
33. Cuijpers P. Meta-Analyses in Mental Health Research: A Practical Guide. ResearchGate. 2016. URL: https://www.researchgate.net/profile/Pim-Cuijpers/publication/301815425_Meta-analyses_in_mental_health_research_A_practical_guide/links/5729b16708aef5d48d2cff0a/Meta-analyses-in-mental-health-research-A-practical-guide.pdf?origin=publication_detail [accessed 2021-11-15]
34. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. Evid Based Ment Health 2019 Nov;22(4):153-160. [doi: 10.1136/ebmental-2019-300117] [Medline: 31563865]
35. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Soft 2010;36(3):1-48. [doi: 10.18637/jss.v036.i03]
36. Lüdecke D. esc: Effect Size Computation for Meta Analysis. Zenodo. URL: https://zenodo.org/record/1249218 [accessed 2021-11-15]
37. Morris SB. Estimating Effect Sizes From Pretest-Posttest-Control Group Designs. Organ Res Methods 2007 Jul 23;11(2):364-386. [doi: 10.1177/1094428106291059]
38. Sommers-Spijkerman M, Austin J, Bohlmeijer E, Pots W. New Evidence in the Booming Field of Online Mindfulness: An Updated Meta-analysis of Randomized Controlled Trials. JMIR Ment Health 2021 Jul 19;8(7):e28168 [FREE Full text] [doi: 10.2196/28168] [Medline: 34279240]
39. Simblett S, Birch J, Matcham F, Yaguez L, Morris R. A Systematic Review and Meta-Analysis of e-Mental Health Interventions to Treat Symptoms of Posttraumatic Stress. JMIR Ment Health 2017 May 17;4(2):e14 [FREE Full text] [doi: 10.2196/mental.5558] [Medline: 28526672]
40. Cuijpers P, Weitz E, Cristea IA, Twisk J. Pre-post effect sizes should be avoided in meta-analyses. Epidemiol Psychiatr Sci 2017 Aug;26(4):364-368 [FREE Full text] [doi: 10.1017/S2045796016000809] [Medline: 27790968]
41. Hedges LV, Olkin I. Statistical methods for meta-analysis. San Diego, CA: Academic Press; 1985:A.
42. Cohen J. Statistical power analysis for the behavioural sciences (2nd ed.). Hillside, NJ: Lawrence Erlbaum Associates; 1988.

XSL•FO
RenderX

43. Chang B, Hoaglin D. Meta-Analysis of Odds Ratios: Current Good Practices. Med Care 2017 Apr;55(4):328-335 [FREE Full text] [doi: 10.1097/MLR.0000000000000696] [Medline: 28169977]
44. Johansson R, Andersson G. Internet-based psychological treatments for depression. Expert Rev Neurother 2012 Jul;12(7):861-9; quiz 870. [doi: 10.1586/ern.12.63] [Medline: 22853793]
45. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
46. Mohr DC, Azocar F, Bertagnolli A, Choudhury T, Chrisp P, Frank R, Banbury Forum on Digital Mental Health. Banbury Forum Consensus Statement on the Path Forward for Digital Mental Health Treatment. Psychiatr Serv 2021 Jun;72(6):677-683. [doi: 10.1176/appi.ps.202000561] [Medline: 33467872]
47. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct 09;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
48. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
49. Christensen H, Griffiths KM, Jorm AF. Delivering interventions for depression by using the internet: randomised controlled trial. BMJ 2004 Jan 31;328(7434):265 [FREE Full text] [doi: 10.1136/bmj.37945.566632.EE] [Medline: 14742346]
50. 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]
51. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital Mental Health Interventions for Depression, Anxiety, and Enhancement of Psychological Well-Being Among College Students: Systematic Review. J Med Internet Res 2019 Jul 22;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]
52. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 06;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]

Abbreviations

RCT: randomized controlled trial
RoB 2: Cochrane risk-of-bias tool version 2
SMD: standardized mean difference

Edited by T Leung; submitted 31.12.21; peer-reviewed by E Lattie, S Adam; comments to author 21.01.22; revised version received 15.03.22; accepted 25.04.22; published 15.06.22.

Please cite as:
Leung C, Pei J, Hudec K, Shams F, Munthali R, Vigo D
The Effects of Nonclinician Guidance on Effectiveness and Process Outcomes in Digital Mental Health Interventions: Systematic Review and Meta-analysis
J Med Internet Res 2022;24(6):e36004
URL: https://www.jmir.org/2022/6/e36004
doi:10.2196/36004
PMID:35511463

©Calista Leung, Julia Pei, Kristen Hudec, Farhud Shams, Richard Munthali, Daniel Vigo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 15.06.2022. 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

Effectiveness of Using Virtual Reality–Supported Exercise Therapy for Upper Extremity Motor Rehabilitation in Patients With Stroke: Systematic Review and Meta-analysis of Randomized Controlled Trials

Jiayin Chen[1], MSc; Calvin Kalun Or[1], PhD; Tianrong Chen[1], BSc

Department of Industrial and Manufacturing Systems Engineering, The University of Hong Kong, Hong Kong, China (Hong Kong)

Corresponding Author:
Calvin Kalun Or, PhD
Department of Industrial and Manufacturing Systems Engineering
The University of Hong Kong
Room 8-7, 8/F., Haking Wong Building
Pokfulam
Hong Kong
China (Hong Kong)
Phone: 852 39172587
Email: klor@hku.hk

Abstract

Background: In recent years, efforts have been made to implement virtual reality (VR) to support the delivery of poststroke upper extremity motor rehabilitation exercises. Therefore, it is important to review and analyze the existing research evidence of its effectiveness.

Objective: Through a systematic review and meta-analysis of randomized controlled trials, this study examined the effectiveness of using VR-supported exercise therapy for upper extremity motor rehabilitation in patients with stroke.

Methods: This study followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. The CINAHL Plus, MEDLINE, Web of Science, Embase, and Cochrane Library databases were searched on December 31, 2021. Changes in outcomes related to impairments in upper extremity functions and structures, activity limitations, and participation restrictions in life situations from baseline to after intervention, after intervention to follow-up assessment, and baseline to follow-up assessment were examined. Standardized mean differences (SMDs) were calculated using a random-effects model. Subgroup analyses were performed to determine whether the differences in treatment outcomes depended on age, stroke recovery stage, VR program type, therapy delivery format, similarities in intervention duration between study groups, intervention duration in VR groups, and trial length.

Results: A total of 42 publications representing 43 trials (aggregated sample size=1893) were analyzed. Compared with the control groups that used either conventional therapy or no therapy, the intervention groups that used VR to support exercise therapy showed significant improvements in upper extremity motor function (Fugl-Meyer Assessment-Upper Extremity; SMD 0.45, 95% CI 0.21-0.68; *P*<.001), range of motion (goniometer; SMD 1.01, 95% CI 0.50-1.52; *P*<.001), muscle strength (Manual Muscle Testing; SMD 0.79, 95% CI 0.28-1.30; *P*=.002), and independence in day-to-day activities (Functional Independence Measure; SMD 0.23, 95% CI 0.06-0.40; *P*=.01, and modified Rankin Scale; SMD 0.57, 95% CI 0.01-1.12; *P*=.046). Significant subgroup differences were observed in hand dexterity (Box and Block Test), spasticity (Ashworth Scale or modified Ashworth Scale), arm and hand motor ability (Wolf Motor Function Test and Manual Function Test), hand motor ability (Jebsen Hand Function Test), and quality of life (Stroke Impact Scale). There was no evidence that the benefits of VR-supported exercise therapy were maintained after the intervention ended.

Conclusions: VR-supported upper extremity exercise therapy can be effective in improving motor rehabilitation results. Our review showed that of the 12 rehabilitation outcomes examined during the course of VR-based therapy, significant improvements were detected in 2 (upper extremity motor function and range of motion), and both significant and nonsignificant improvements were observed in another 2 (muscle strength and independence in day-to-day activities), depending on the measurement tools or methods used.

XSL•FO
RenderX

Trial Registration: PROSPERO CRD42021256826; https://tinyurl.com/2uarftbh

(J Med Internet Res 2022;24(6):e24111) doi:10.2196/24111

KEYWORDS

virtual reality; stroke; rehabilitation; upper extremity; meta-analysis

Introduction

Upper extremity motor impairment after stroke significantly impedes the performance of daily activities and affects patients' quality of life [1-6]. A major health goal for these patients is to recover their motor function and regain independence. Upper extremity therapeutic exercises are the main approach used to achieve this goal [7].

The physical therapist–led, face-to-face approach to delivering therapeutic exercises has been a common practice, but it can be costly and inconvenient owing to professional and institutional resource requirements. Therefore, alternative delivery protocols that leverage technology have been developed. In particular, the application of virtual reality (VR) technology in poststroke therapeutic exercise delivery has received considerable attention in recent years [8-11].

Although previous studies have reported the application of VR to deliver therapeutic exercise, a greater understanding of its effectiveness in poststroke functioning and health improvement is also required. Such knowledge can be acquired by reviewing the existing literature. Despite some reviews that have examined the effectiveness of using VR for upper extremity motor rehabilitation [12-17], there have been several new studies published in recent years; therefore, an updated review of the existing evidence is warranted. Moreover, previous reviews [12,16,17] have categorized study outcomes into three levels: (1) impairments in body functions (ie, problems with the physiological function of body systems) and structures (eg, extremities), (2) activity limitations (ie, difficulties in executing activities), and (3) restrictions on participation in life situations (ie, difficulties in involvement in life situations), according to the International Classification of Functioning, Disability, and Health Framework [18]. However, some study outcomes that have previously been grouped at the same level may not actually measure the same construct. For example, hand dexterity (as measured by the Box and Block Test [BBT]), and independence in day-to-day activities (as measured by the Functional Independence Measure [FIM]) have both been categorized as activity limitations, but are, in fact, 2 different types of outcomes. Therefore, it may not be appropriate to group the 2 measures together. Moreover, several recent reviews have mainly analyzed a small number of common outcomes [19-21], such as upper extremity motor function (as measured by the Fugl-Meyer Assessment-Upper Extremity [FMA-UE]) and hand dexterity (BBT), whereas relatively less attention has been paid to other outcomes (eg, range of motion [ROM] and muscle strength as measured by Manual Muscle Testing [MMT]), which may also be important for evaluating the effects of VR-supported exercise therapy on upper extremity motor recovery. Furthermore, previous reviews [15,16] performed subgroup analyses to demonstrate the effects of several moderating factors (eg, the stage of stroke recovery, the type of VR program, and the intervention duration) on the association between VR-supported exercise therapy and relevant study outcomes. However, similar to the aforementioned issues, the moderating effects on individual outcomes could not be accurately determined because outcomes that were actually related to different aspects were inappropriately grouped into the same category (eg, grouping grip strength and ROM into one category).

In view of the aforementioned limitations of previous reviews, we conducted this systematic review and meta-analysis to provide more evidence for the effectiveness of VR-supported exercise therapy for upper extremity motor rehabilitation in patients with stroke, particularly relating to outcomes in impairment of upper extremity functions and structures, activity limitations, and participation restrictions in life situations. In addition, we attempted to examine additional factors (eg, therapy delivery format) for their moderating effects on these 3 outcome categories.

Methods

This review was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement and its associated checklist (Multimedia Appendix 1) [22] and was registered with PROSPERO (CRD42021256826).

Search Strategy

A literature search was performed on December 31, 2021, using the following databases: CINAHL Plus via EBSCO (from 1937 to present), MEDLINE via Ovid (from 1946 to present), Web of Science (from 1956 to present), Embase via Ovid (from 1974 to present), and the Cochrane Library (no date restriction). Medical Subject Headings and free-text search terms related to stroke, VR, upper extremity, and rehabilitation were used. Details of the search are presented in Multimedia Appendix 2.

Inclusion and Exclusion Criteria

Studies were included if (1) they were randomized controlled trials examining the effectiveness of VR-supported exercise therapy for upper extremity motor rehabilitation; (2) the intervention groups used either VR-supported exercise therapy alone or in combination with conventional therapy and the control groups used either conventional therapy alone or no therapy; (3) they examined adult patients with stroke (aged >18 years); (4) they assessed outcomes related to impairments in upper extremity functions or structures, activity limitations, and participation restrictions in life situations; and (5) they were written in English and published in peer-reviewed journals. Studies were excluded if (1) they did not focus on motor rehabilitation only for the upper extremities, as the independent

XSL•FO
RenderX

effects of VR-supported exercise therapy on the upper extremities may be difficult to identify in combined studies; (2) they did not report mean and SD values for the changes in outcomes for effect size calculations; (3) the data could not be imputed based on the information available in the publication; (4) the data could not be obtained within 1 month of contacting the corresponding authors; or (5) they were review studies, case reports, or abstracts.

Study Selection

After removing duplicate publications from the search results, 2 authors (JC and TC) independently screened the titles and abstracts of the remaining publications and excluded those that were deemed irrelevant. The full texts of the potentially relevant publications were further reviewed to determine their eligibility for inclusion. The reference lists of the included articles and relevant review articles were manually searched to identify additional studies. Agreement between the authors on inclusion and exclusion decisions was assessed using the κ statistic, with κ values from 0.40 to 0.59, 0.60 to 0.74, and ≥0.75 considered as fair, good, and excellent agreement, respectively [23]. Any disagreements were resolved through discussions between the authors until a consensus was reached.

Data Extraction

JC and TC used a standardized form to independently extract data related to the characteristics of the trial, the attributes of the participants, the details of the intervention and control conditions, the outcomes examined in each trial, and the mean and SD values for changes in outcomes (ie, changes from baseline to after intervention, changes from after intervention to follow-up assessment, and changes from baseline to follow-up assessment). Data from the final follow-up assessment were used for the trials with multiple follow-up assessments. Any disagreements regarding data extraction were resolved through discussion between the authors until a consensus was reached.

Assessment of Risk of Bias

The risk of bias in the included trials was independently assessed by JC and TC using the Cochrane Collaboration tool [24]. The following aspects were assessed: random sequence generation; allocation concealment; blinding of participants and health care providers; blinding of outcome assessors; incomplete outcome data; selective reporting; and other sources of bias, including significant differences between study groups at baseline and different intervention durations between study groups.

Data Analysis

Outcomes were included in the meta-analysis if they were reported in at least 2 trials. For data from follow-up assessments, outcomes were included in the meta-analysis if they were reported in at least 2 follow-up assessments. We pooled the data across trials using random-effects models and calculated the standardized mean difference (SMD) for each outcome. Positive (or negative) SMDs indicated that the results favored the intervention (or control) condition. Unreported SDs were imputed according to the guidelines provided in the Cochrane Handbook for Systematic Reviews of Interventions [24]. Outliers in the meta-analysis were identified using studentized residuals (>3 in absolute value) and leave-one-out sensitivity analyses [25]. Heterogeneity across trials was assessed using Cochran Q test and I^2 statistics (25%, 50%, and 75% were considered low, moderate, and high levels of heterogeneity, respectively) [26]. Egger regression test was used to measure the possibility of publication bias, with 2-tailed P values of <.05 indicating potential publication bias [27]. Comprehensive Meta-Analysis (version 3.0) was used to perform the meta-analysis.

Subgroup analysis was performed to investigate the factors that may moderate the effects of at least 1 trial in each subgroup. The following moderating factors were examined: age (below the median value of the participants' ages vs equal to or above the median value of the participants' ages), stage of recovery (subacute vs chronic stroke) [28], type of VR program (specialized programs designed for rehabilitation vs commercial games) [7], therapy delivery format (VR-supported exercise therapy alone compared with a control condition vs VR-supported exercise therapy+conventional therapy compared with a control condition), similarity of the intervention duration between the study groups (same intervention duration in both VR and control groups vs longer intervention duration in VR groups), intervention duration in VR groups (≤15 hours vs >15 hours) [15], and length of the trial (≤1 month vs >1 month and ≤2 months vs >2 months).

Assessment of Quality of Evidence

The quality of evidence for each outcome was assessed using the Grading of Recommendations Assessment, Development, and Evaluation approach [29]. For each outcome, the quality of evidence was downgraded from high by one level for each serious issue found in the domains of risk of bias, inconsistency, indirectness, imprecision, and publication bias.

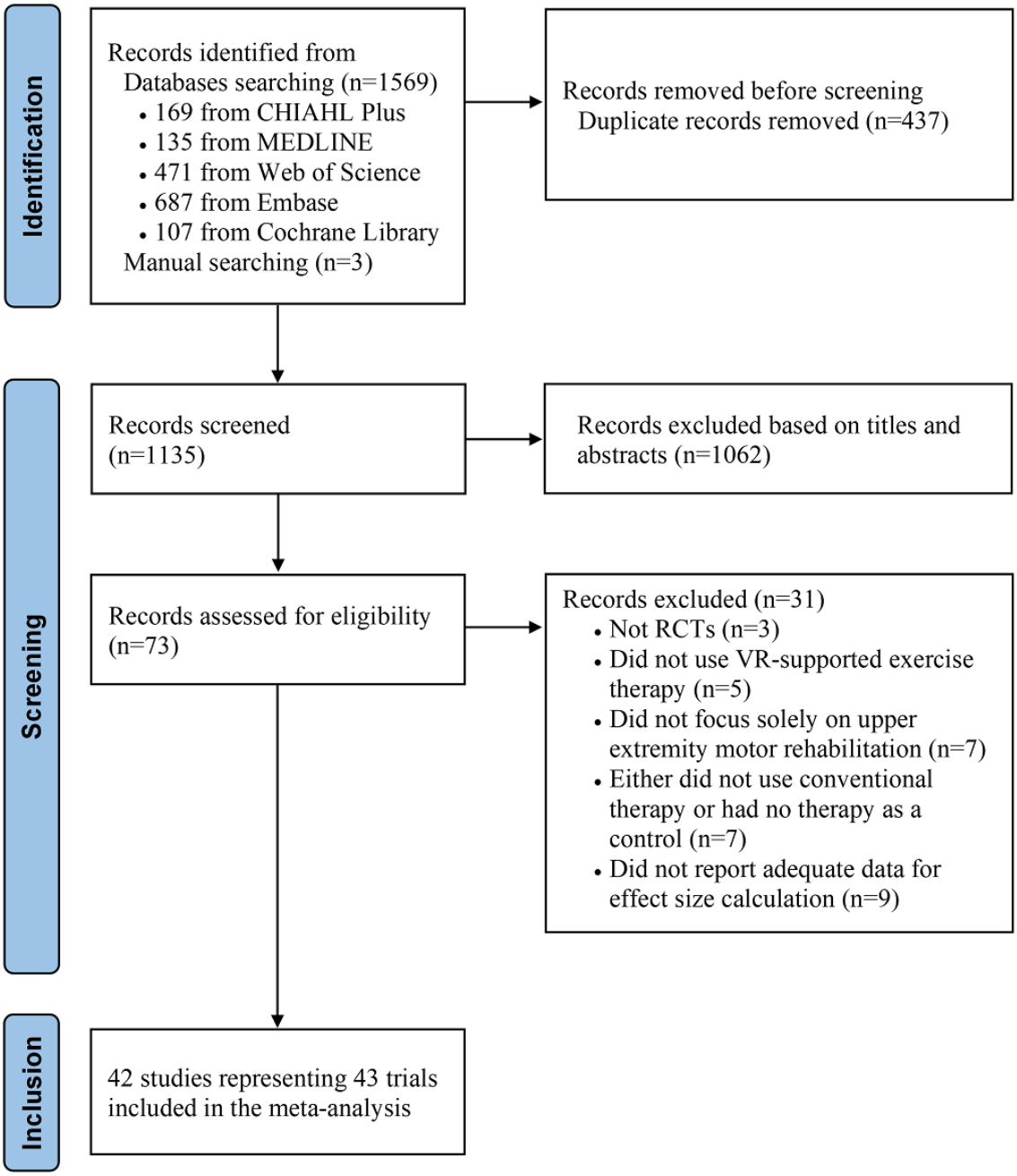
Results

Study Selection Process

Figure 1 illustrates the study selection process. A total of 42 studies were identified as being eligible [8-10,30-68]. A study [52] had 2 groups of participants: individuals with subacute stroke and individuals with chronic stroke. Therefore, the study was divided into 2 trials (ie, Miclaus et al (1) [52] and Miclaus et al (2) [52]) for analysis. Altogether, 42 studies representing 43 trials (aggregated sample size=961 [intervention groups] and 932 [control groups]) were included in the final analysis. The agreement between the 2 authors on the inclusion and exclusion decisions was good at both the title and abstract screening (κ=0.64) and full-text reading steps (κ=0.61).

XSL•FO
RenderX

Figure 1. Study selection process. RCT: randomized controlled trial; VR: virtual reality.



Characteristics of the Included Trials

Table 1 summarizes the characteristics of the 43 included trials. Multimedia Appendix 3 [8-10,30-68] presents the characteristics of the participants and the study groups in each trial. Multimedia Appendix 4 [8-10,30-68] describes the outcome of each trial.

XSL•FO
RenderX

Table 1. Summary of trial characteristics (N=43).

Characteristics	Values
Publication year	
2011 and before, n (%)	3 (7) [43,55,56]
2012-2016, n (%)	17 (40) [10,37,38,42,44-48,50,51,57-60,64,66]
2017-2021, n (%)	23 (53) [8,9,30-36,39-41,49,52-54,61-63,65,67,68]
Value, median (IQR)	2017 (2014-2019)
Trial location, n (%)	
Asia	25 (58) [31,32,34-37,39-41,44,46-50,54,57-60,62,63,65,66,68]
Europe	11 (26) [9,30,38,42,43,45,52,55,56,61]
North America	2 (5) [53,64]
Oceania	1 (2) [51]
Africa	1 (2) [67]
South America	1 (2) [33]
Multiple locations	2 (5) [8,10]
Sample size, median (range)	33 (11-235)
Participant age (years), median (range)	60.36 (49.64-74.07)[a]
Males (%), median (range)	61.04 (36.36-86.00)[b]
Ischemic stroke (%), median (range)	70.83 (38.46-100)[c]
Stroke recovery stage, n (%)	
Subacute stroke (≤6 months)	22 (51) [8-10,30,31,34,36,37,41-44,46,49,52,57,59,61-63,65,68]
Chronic stroke (>6 months)	20 (47) [32,33,35,38-40,45,47,48,50-56,58,60,64,67]
No adequate information was provided	1 (2) [66]
Type of VR[d] program, n (%)	
Specialized program designed for rehabilitation	27 (63) [8,9,34,35,38,40-43,45,46,48,50,52-58,61,63-66,68]
Commercial game	16 (37) [10,30-33,36,37,39,44,47,49,51,59,60,62,67]
Therapy delivery format, n (%)	
VR-supported exercise therapy alone compared with no therapy	2 (5) [33,61]
VR-supported exercise therapy alone compared with conventional therapy	13 (30) [8,30,34,37,38,45,48,50,51,55,56,59,64]
VR-supported exercise therapy+conventional therapy compared with conventional therapy	28 (65) [9,10,31,32,35,36,39-44,46,47,49,52-54,57,58,60,62,63,65-68]
VR-supported exercise therapy delivery frequency, n (%)	
2 to 3 times per week	11 (25) [38-40,45,47-49,53,59,60,67]
>3 times per week	27 (63) [8,9,30-37,41-44,46,51,52,54-56,61-63,65,66,68]
No adequate information was provided	5 (12) [10,50,57,58,64]
Duration of each VR-supported exercise therapy session, n (%)	
20 to 45 minutes per session	23 (54) [30-32,34,37-41,45-50,53,54,57,58,60,63,67,68]
>45 and ≤75 minutes per session	16 (37) [9,10,33,35,42-44,51,52,55,56,59,62,65,66]
No adequate information was provided	4 (9) [8,36,61,64]
Intervention duration for VR groups, n (%)	
≤15 hours	15 (35) [8,33,34,37,38,41,45,48,50-52,57,63,64]
>15 hours	23 (53) [9,30-32,35,39,42-44,46,47,49,54-56,58-62,65,66,68]
No adequate information was provided	5 (12) [10,36,40,53,67]

XSL•FO
RenderX

Characteristics	Values
Trial length, n (%)	
2 weeks to 1 month	31 (72) [8-10,31,33-38,41-44,46,50-58,62-66,68]
>1 and ≤2 months	10 (23) [30,32,39,45,47-49,60,61,67]
>2 and ≤3 months	2 (5) [40,59]
Time point of the final follow-up assessment after the end of intervention, n (%)	
1 month	8 (19) [10,41,45,50,53-55,58]
1.5 months	1 (2) [38]
3 months	3 (7) [8,40,44]
6 months	2 (5) [30,51]
No follow-up assessment	29 (67) [9,31-37,39,42,43,46-49,52,56,57,59-68]

[a]Anjum et al [34], Miclaus et al (1) [52], and Miclaus et al (2) [52] did not report the participants' mean age.

[b]Anjum et al [34] did not report the number or ratio of male participants in their study.

[c]Ain et al [32], Anjum et al [34], Crosbie et al [38], Ersoy and Iyigun [39], Jo et al [66], Levin et al [50], Mokhtar et al [67], Park et al [65], Shin et al [57], Standen et al [61], Xie et al [63], and Zondervan et al [64] did not report the participants' stroke types.

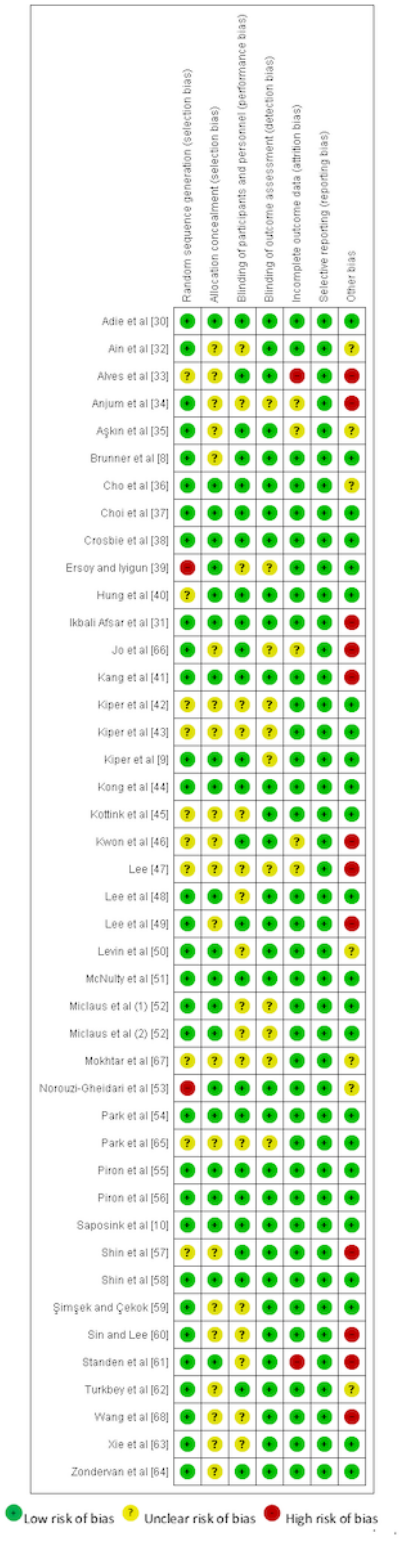
[d]VR: virtual reality.

Risk of Bias

Figure 2 [8-10,30-68] shows the results of the risk of bias assessment for all 43 trials. Random sequence generation was assessed as adequate in 72% (31/43) of the trials. Allocation concealment was assessed as adequate in 51% (22/43) of the trials. Blinding of the participants or health care providers was reported in 58% (25/43) of the trials, and blinding of the outcome assessors was reported in 74% (32/43) of the trials. We assessed 84% (36/43) of the trials as free of bias in terms of incomplete outcome data. All the trials were assessed as having a low risk of reporting bias. Of the trials, 56% (24/43) had a low risk of bias in terms of significant differences between study groups at baseline or different intervention durations between study groups.

XSL•FO
RenderX

Figure 2. Risk of bias summary for the included trials [8-10,30-68].



Meta-analysis of the Effects of VR-Supported Exercise Therapy

Table 2 presents the results of the meta-analyses and the assessments of heterogeneity, publication bias, and quality of evidence. Forest plots for each outcome are presented in Multimedia Appendix 5 (Figures S1-S20 [8-10,30-68]).

XSL•FO
RenderX

Table 2. Meta-analyses and assessments of heterogeneity, publication bias, and quality of evidence.

Outcomes	Tools or methods used to assess the outcomes	Number of trials analyzed and number of participants involved	Standardized mean difference (95% CI)	Between-group difference, *P* value	Heterogeneity			Egger test, *P* value	Quality of evidence
					Cochrane *Q* test	*P* value	I^2 (%)		
Impairments in upper extremity functions and structures									
(1) Upper extremity motor function	Fugl-Meyer Assessment-Upper Extremity	28 [9,31-33,35-37,40-46,48,50-58,60,63,65]; $N_{VR\ group}$[a]=526, $N_{control\ group}$=509	0.45 (0.21 to 0.68)	<.001	83.72	<.001	68	.33	Moderate[b]
(1) Upper extremity motor function (after one outlier removed[c])	Fugl-Meyer Assessment-Upper Extremity	27 [9,31-33,35-37,40-46,48,50-58,60,63,65]; $N_{VR\ group}$=502, $N_{control\ group}$=487	0.35 (0.19 to 0.50)	<.001	35.23	.11	26	.84	High
(2) Grip strength	Dynamometer	6 [10,36,37,41,65,67]; $N_{VR\ group}$=157, $N_{control\ group}$=155	−0.002 (−0.30 to 0.30)	.99	7.41	.19	32	.23	Moderate[d]
(3) Spasticity	Ashworth Scale or modified Ashworth Scale	6 [35,43,47,52,55]; $N_{VR\ group}$=109, $N_{control\ group}$=111	0.09 (−0.28 to 0.47)	.63	8.68	.12	42	.35	Moderate[d]
(4) Range of motion	Goniometer	4 [52,54,60]; $N_{VR\ group}$=56, $N_{control\ group}$=56	1.01 (0.50 to 1.52)	<.001	4.65	.20	35	.99	Low[d]
(5) Stroke recovery stage	Brunnstrom stages of stroke recovery for upper extremity	2 [35,62]; $N_{VR\ group}$=28, $N_{control\ group}$=29	0.26 (−0.26 to 0.79)	.32	0.27	.61	0	N/A[e]	Low[d]
(6) Muscle strength	Manual Muscle Testing	3 [47,52]; $N_{VR\ group}$=33, $N_{control\ group}$=33	0.79 (0.28 to 1.30)	.002	2.03	.36	1	.73	Low[d]
(6) Muscle strength	Motricity Index	2 [35,38]; $N_{VR\ group}$=27, $N_{control\ group}$=29	0.09 (−0.43 to 0.62)	.73	0.88	.35	0	N/A	Low[d]
Activity limitations									
(7) Independence in day-to-day activities	Functional Independence Measure	13 [8-10,31,42-44,47,52,56,59,62]; $N_{VR\ group}$=406, $N_{control\ group}$=395	0.23 (0.06 to 0.40)	.01	16.01	.19	25	.03	High
(7) Independence in day-to-day activities	Barthel Index or modified Barthel Index	11 [10,34,36,37,41,46,48,54,57,65,67]; $N_{VR\ group}$=224, $N_{control\ group}$=221	0.20 (−0.16 to 0.55)	.28	30.54	.001	67	.59	Moderate[b]
(7) Independence in day-to-day activities	Modified Rankin Scale	2 [52]; $N_{VR\ group}$=26, $N_{control\ group}$=26	0.57 (0.01 to 1.12)	.046	0.55	.46	0	N/A	Low[d]
(8) Hand dexterity	Box and Block Test	13 [8,10,31,32,35,37,41,48,51,53,60,62,64]; $N_{VR\ group}$=297, $N_{control\ group}$=286	0.26 (−0.08 to 0.60)	.13	42.63	<.001	72	.33	Moderate[b]

XSL•FO
RenderX

Outcomes	Tools or methods used to assess the outcomes	Number of trials analyzed and number of participants involved	Standardized mean difference (95% CI)	Between-group difference, *P* value	Heterogeneity			Egger test, *P* value	Quality of evidence
					Cochrane *Q* test	*P* value	I^2 (%)		
(9) Arm and hand motor ability	Action Research Arm Test	6 [8,30,38,44,45,64]; $N_{VR\ group}$=238, $N_{control\ group}$=238	−0.03 (−0.21 to 0.15)	.76	2.08	.84	0	.22	High
(9) Arm and hand motor ability	Wolf Motor Function Test task completion time	9 [10,40,50,51,54,61,62,66,68]; $N_{VR\ group}$=174, $N_{control\ group}$=170	0.15 (−0.06 to 0.37)	.16	7.19	.52	0	.28	Moderate[d]
(9) Arm and hand motor ability	Wolf Motor Function Test task performance score	7 [39,40,50,54,62,66,68]; $N_{VR\ group}$=93, $N_{control\ group}$=91	0.36 (−0.07 to 0.79)	.10	11.97	.06	50	.28	Low[d,f]
(9) Arm and hand motor ability	Manual Function Test	4 [37,46,48,49]; $N_{VR\ group}$=51, $N_{control\ group}$=51	0.20 (−0.37 to 0.78)	.49	6.28	.10	52	.70	Low[d,f]
(10) Hand motor ability	Jebsen Hand Function Test	4 [36,41,58,65]; $N_{VR\ group}$=70, $N_{control\ group}$=67	0.90 (−0.42 to 2.22)	.18	36.25	<.001	92	.65	Very low[b,d]
Participation restrictions in life situations									
(11) Quality of life	Stroke Impact Scale total score	3 [30,53,54]; $N_{VR\ group}$=138, $N_{control\ group}$=140	0.13 (−0.41 to 0.66)	.65	4.26	.12	53	.12	Very low[d,f]
(11) Quality of life	Stroke Impact Scale hand function score	2 [10,44]; $N_{VR\ group}$=104, $N_{control\ group}$=105	−0.04 (−0.31 to 0.23)	.78	0.89	.35	0	N/A	Low[d]
(12) Upper extremity use in daily life	Motor Activity Log quality of movement score	6 [40,50,51,53,61,64]; $N_{VR\ group}$=71, $N_{control\ group}$=68	0.50 (−0.05 to 1.05)	.08	11.78	.04	58	.31	Low[d,f]
(12) Upper extremity use in daily life	Motor Activity Log amount of use score	5 [40,50,53,61,64]; $N_{VR\ group}$=50, $N_{control\ group}$=48	0.27 (−0.13 to 0.67)	.18	3.36	.50	0	.91	Moderate[d]

[a]VR: virtual reality.

[b]Downgraded owing to a high level of heterogeneity.

[c]Shin et al [58] was removed.

[d]Downgraded owing to an inadequate sample size.

[e]N/A: not applicable.

[f]Downgraded owing to a moderate level of heterogeneity.

Effects on Outcomes Related to Impairments in Upper Extremity Functions and Structures

Compared with the control condition, the use of VR-supported exercise therapy was associated with significant improvements in upper extremity motor function (FMA-UE; SMD 0.45, 95% CI 0.21-0.68; *P*<.001 or SMD 0.35, 95% CI 0.19-0.50; *P*<.001 after outlier [58] removal), upper extremity ROM (goniometer; SMD 1.01, 95% CI 0.50-1.52; *P*<.001), and upper extremity muscle strength (MMT; SMD 0.79, 95% CI 0.28-1.30; *P*=.002). No significant improvements were observed in grip strength (dynamometer), spasticity (ie, involuntary muscle contraction, stiffening, and tightening upon the movement of body parts; Ashworth Scale [AS] or Modified AS [mAS]), upper extremity stroke recovery stage (Brunnstrom Stages of Stroke Recovery for Upper Extremity), and upper extremity muscle strength (Motricity Index).

XSL•FO
RenderX

Effects on Outcomes Related to Activity Limitation

Compared with the control condition, the use of VR-supported exercise therapy was associated with significant improvements in independence in day-to-day activities (FIM; SMD 0.23, 95% CI 0.06-0.40; *P*=.01 and modified Rankin Scale scores; SMD 0.57, 95% CI 0.01-1.12; *P*=.046). However, no significant association was observed with the Barthel Index or modified Barthel Index.

No significant improvements were detected in hand dexterity (BBT), arm and hand motor ability (Action Research Arm Test [ARAT], Wolf Motor Function Test [WMFT], and Manual Function Test [MFT]), and hand motor ability (Jebsen Hand Function Test [JHFT]).

Effects on Outcomes Related to Participation Restrictions in Life Situations

No significant improvements were detected in quality of life (Stroke Impact Scale [SIS]) or upper extremity use in daily life (Motor Activity Log).

Subgroup Analyses

Overview

The subgroup analyses for outcomes examined in at least 10 trials are presented in this paper (Tables 3-6). For outcomes that were examined in <10 trials, the subgroup analyses are presented in Multimedia Appendix 6 (Tables S1-S16 [8-10,30-68]).

Significant subgroup differences were observed in the following outcomes: hand dexterity (BBT), spasticity (AS or mAS), arm and hand motor ability (WMFT task performance score and MFT), hand motor ability (JHFT), and quality of life (SIS total score). The details of this process are presented in the following sections.

XSL•FO
RenderX

Table 3. Subgroup analyses of upper extremity motor function as assessed by the Fugl-Meyer Assessment-Upper Extremity.

Moderating factors	Number of trials analyzed and number of participants involved	Standardized mean difference (95% CI)	Between-group difference, *P* value	Subgroup difference, *P* value
Age (years)				
Younger (<60.36)	14 [32,33,35,40,41,44,46,50,51,53,54,57,58,63]; N_{VR}[a] $_{group}$=222, $N_{control\ group}$=213	0.54 (0.09 to 1.00)	.02	.43
Older (≥60.36)	12 [9,31,36,37,42,43,45,48,55,56,60,65]; $N_{VR\ group}$=278, $N_{control\ group}$=270	0.35 (0.18 to 0.52)	<.001	.43
Stroke recovery stage				
Subacute stroke	13 [9,31,36,37,41-44,46,52,57,63,65]; $N_{VR\ group}$=273, $N_{control\ group}$=266	0.27 (0.04 to 0.50)	.02	.16
Chronic stroke	15 [32,33,35,40,45,48,50-56,58,60]; $N_{VR\ group}$=253, $N_{control\ group}$=243	0.60 (0.21 to 1.00)	.003	.16
Type of VR program used				
Specialized program designed for rehabilitation	20 [9,35,40-43,45,46,48,50,52-58,63,65]; $N_{VR\ group}$=371, $N_{control\ group}$=364	0.44 (0.15 to 0.74)	.003	.90
Commercial game	8 [31-33,36,37,44,51,60]; $N_{VR\ group}$=155, $N_{control\ group}$=145	0.47 (0.10 to 0.85)	.01	.90
Therapy delivery format				
VR-supported exercise therapy alone compared with no therapy	1 [33]; $N_{VR\ group}$=17, $N_{control\ group}$=10	1.10 (0.27 to 1.94)	.01	.12
VR-supported exercise therapy alone compared with conventional therapy	7 [37,45,48,50,51,55,56]; $N_{VR\ group}$=103, $N_{control\ group}$=100	0.25 (−0.03 to 0.53)	.08	.12
VR-supported exercise therapy+conventional therapy compared with conventional therapy	20 [9,31,32,35,36,40-44,46,52-54,57,58,60,63,65]; $N_{VR\ group}$=406, $N_{control\ group}$=399	0.50 (0.20 to 0.81)	.001	.12
Similarity of intervention duration between groups				
Same intervention duration in both VR and control groups	21 [9,32,37,40-46,48,50-52,54-56,58,63,65]; $N_{VR\ group}$=424, $N_{control\ group}$=418	0.44 (0.16 to 0.73)	.002	.14
Longer intervention duration in VR groups	4 [31,33,57,60]; $N_{VR\ group}$=63, $N_{control\ group}$=50	0.81 (0.42 to 1.20)	<.001	.14
Intervention duration in VR groups (hours)				
≤15	11 [33,37,41,45,48,50-52,57,63]; $N_{VR\ group}$=128, $N_{control\ group}$=119	0.37 (0.05 to 0.69)	.02	.43
>15	14 [9,31,32,35,42-44,46,54-56,58,60,65]; $N_{VR\ group}$=360, $N_{control\ group}$=353	0.56 (0.21 to 0.91)	.002	.43
Trial length				
2 weeks to 1 month	23 [9,31,33,35-37,41-44,46,50-58,63,65]; $N_{VR\ group}$=445, $N_{control\ group}$=428	0.43 (0.16 to 0.69)	.002	.47
>1 and ≤2 months	4 [32,45,48,60]; $N_{VR\ group}$=64, $N_{control\ group}$=65	0.68 (0.18 to 1.18)	.01	.47
>2 and ≤3 months	1 [40]; $N_{VR\ group}$=17, $N_{control\ group}$=16	0.16 (−0.53 to 0.84)	.65	.47

[a]VR: virtual reality.

XSL•FO
RenderX

Table 4. Subgroup analyses of hand dexterity as assessed by the Box and Block Test.

Moderating factors	Number of trials analyzed and number of participants involved	Standardized mean difference (95% CI)	Between-group difference, *P* value	Subgroup difference, *P* value
Age (years)				
Younger (<60.36)	6 [32,35,41,51,53,64]; N_{VR}[a] $_{group}$=94, $N_{control\ group}$=93	0.12 (−0.38 to 0.62)	.64	.47
Older (≥60.36)	7 [8,10,31,37,48,60,62]; $N_{VR\ group}$=203, $N_{control\ group}$=193	0.38 (−0.11 to 0.87)	.13	.47
Stroke recovery stage				
Subacute stroke	6 [8,10,31,37,41,62]; $N_{VR\ group}$=184, $N_{control\ group}$=174	0.11 (−0.27 to 0.48)	.58	.44
Chronic stroke	7 [32,35,48,51,53,60,64]; $N_{VR\ group}$=113, $N_{control\ group}$=112	0.38 (−0.20 to 0.95)	.20	.44
Type of VR program				
Specialized program designed for rehabilitation	6 [8,35,41,48,53,64]; $N_{VR\ group}$=123, $N_{control\ group}$=119	−0.03 (−0.28 to 0.22)	.81	.09
Commercial game	7 [10,31,32,37,51,60,62]; $N_{VR\ group}$=174, $N_{control\ group}$=167	0.54 (−0.06 to 1.14)	.08	.09
Therapy delivery format				
VR-supported exercise therapy alone compared with conventional therapy	5 [8,37,48,51,64]; $N_{VR\ group}$=115, $N_{control\ group}$=109	−0.08 (−0.34 to 0.18)	.56	.046
VR-supported exercise therapy+conventional therapy compared with conventional therapy	8 [10,31,32,35,41,53,60,62]; $N_{VR\ group}$=182, $N_{control\ group}$=177	0.52 (−0.01 to 1.05)	.052	.046
Similarity of intervention duration between groups				
Same intervention duration in both VR and control groups	9 [8,10,32,37,41,48,51,62,64]; $N_{VR\ group}$=233, $N_{control\ group}$=224	0.07 (−0.25 to 0.40)	.66	.002
Longer intervention duration in VR groups	2 [31,60]; $N_{VR\ group}$=37, $N_{control\ group}$=33	1.34 (0.61 to 2.07)	<.001	.002
Intervention duration in VR groups (hours)				
≤15	6 [8,37,41,48,51,64]; $N_{VR\ group}$=127, $N_{control\ group}$=120	−0.10 (−0.35 to 0.15)	.45	<.001
>15	5 [31,32,35,60,62]; $N_{VR\ group}$=90, $N_{control\ group}$=87	0.92 (0.35 to 1.49)	.002	<.001
Trial length				
2 weeks to 1 month	10 [8,10,31,35,37,41,51,53,62,64]; $N_{VR\ group}$=241, $N_{control\ group}$=231	0.02 (−0.22 to 0.26)	.84	.049
>1 and ≤2 months	3 [32,48,60]; $N_{VR\ group}$=56, $N_{control\ group}$=55	0.97 (0.06 to 1.89)	.04	.049

[a]VR: virtual reality.

XSL•FO
RenderX

Table 5. Subgroup analyses of independence in day-to-day activities as assessed by the Functional Independence Measure.

Moderating factors	Number of trials analyzed and number of participants involved	Standardized mean difference (95% CI)	Between-group difference, *P* value	Subgroup difference, *P* value
Age (years)				
Younger (<60.36)	2 [44,59]; N_{VR}[a] $_{group}$=53, $N_{control\ group}$=57	0.36 (−0.54 to 1.26)	.44	.70
Older (≥60.36)	9 [8-10,31,42,43,47,56,62]; $N_{VR\ group}$=327, $N_{control\ group}$=312	0.18 (0.02 to 0.33)	.03	.70
Stroke recovery stage				
Subacute stroke	10 [8-10,31,42-44,52,59,62]; $N_{VR\ group}$=352, $N_{control\ group}$=344	0.26 (0.05 to 0.47)	.02	.79
Chronic stroke	3 [47,52,56]; $N_{VR\ group}$=54, $N_{control\ group}$=51	0.20 (−0.19 to 0.58)	.31	.79
Type of VR program used				
Specialized program designed for rehabilitation	7 [8,9,42,43,52,56]; $N_{VR\ group}$=246, $N_{control\ group}$=236	0.28 (0.06 to 0.51)	.02	.58
Commercial game	6 [10,31,44,47,59,62]; $N_{VR\ group}$=160, $N_{control\ group}$=159	0.18 (−0.10 to 0.46)	.21	.58
Therapy delivery format				
VR-supported exercise therapy alone compared with conventional therapy	3 [8,56,59]; $N_{VR\ group}$=109, $N_{control\ group}$=103	0.27 (−0.18 to 0.73)	.23	.86
VR-supported exercise therapy+conventional therapy compared with conventional therapy	10 [9,10,31,42-44,47,52,62]; $N_{VR\ group}$=297, $N_{control\ group}$=292	0.23 (0.04 to 0.42)	.02	.86
Similarity of intervention duration between groups				
Same intervention duration in both VR and control groups	11 [8-10,42-44,52,56,59,62]; $N_{VR\ group}$=380, $N_{control\ group}$=372	0.25 (0.05 to 0.44)	.01	.96
Longer intervention duration in VR groups	2 [31,47]; $N_{VR\ group}$=26, $N_{control\ group}$=23	0.23 (−0.34 to 0.79)	.43	.96
Intervention duration in VR groups (hours)				
≤15	3 [8,52]; $N_{VR\ group}$=88, $N_{control\ group}$=84	0.47 (−0.24 to 1.17)	.20	.62
>15	9 [9,31,42-44,47,56,59,62]; $N_{VR\ group}$=247, $N_{control\ group}$=241	0.28 (0.10 to 0.46)	.002	.62
Trial length				
2 weeks to 1 month	11 [8-10,31,42-44,52,56,62]; $N_{VR\ group}$=379, $N_{control\ group}$=366	0.19 (0.02 to 0.35)	.03	.14
>1 and ≤2 months	1 [47]; $N_{VR\ group}$=7, $N_{control\ group}$=7	0.32 (−0.74 to 1.37)	.56	.14
>2 and ≤3 months	1 [59]; $N_{VR\ group}$=20, $N_{control\ group}$=22	0.84 (0.21 to 1.47)	.01	.14

[a]VR: virtual reality.

XSL•FO
RenderX

Table 6. Subgroup analyses of independence in day-to-day activities as assessed by the Barthel Index or modified Barthel Index.

Moderating factors	Number of trials analyzed and number of participants involved	Standardized mean difference (95% CI)	Between-group difference, *P* value	Subgroup difference, *P* value
Age (years)				
Younger (<60.36)	6 [34,41,46,54,57,67]; N_{VR}[a] $_{group}$=96, $N_{control\ group}$=94	0.38 (−0.24 to 1.00)	.23	.33
Older (≥60.36)	5 [10,36,37,48,65]; $N_{VR\ group}$=128, $N_{control\ group}$=127	0.04 (−0.25 to 0.33)	.80	.33
Stroke recovery stage				
Subacute stroke	8 [10,34,36,37,41,46,57,65]; $N_{VR\ group}$=169, $N_{control\ group}$=165	0.19 (−0.22 to 0.60)	.36	.97
Chronic stroke	3 [48,54,67]; $N_{VR\ group}$=55, $N_{control\ group}$=56	0.17 (−0.73 to 1.08)	.71	.97
Type of VR program				
Specialized program designed for rehabilitation	7 [34,41,46,48,54,57,65]; $N_{VR\ group}$=101, $N_{control\ group}$=99	0.18 (−0.36 to 0.72)	.52	.93
Commercial game	4 [10,36,37,67]; $N_{VR\ group}$=123, $N_{control\ group}$=122	0.21 (−0.31 to 0.74)	.43	.93
Therapy delivery format				
VR-supported exercise therapy alone compared with conventional therapy	3 [34,37,48]; $N_{VR\ group}$=43, $N_{control\ group}$=43	0.23 (−1.04 to 1.49)	.73	.96
VR-supported exercise therapy+conventional therapy compared with conventional therapy	8 [10,36,41,46,54,57,65,67]; $N_{VR\ group}$=181, $N_{control\ group}$=178	0.19 (−0.15 to 0.52)	.27	.96
Similarity of intervention duration between groups				
Same intervention duration in both VR and control groups	7 [10,34,37,41,48,54,65]; $N_{VR\ group}$=160, $N_{control\ group}$=159	0.10 (−0.38 to 0.58)	.69	.72
Longer intervention duration in VR groups	2 [46,57]; $N_{VR\ group}$=22, $N_{control\ group}$=20	0.24 (−0.37 to 0.85)	.45	.72
Intervention duration in VR groups (hours)				
≤15	5 [34,37,41,48,57]; $N_{VR\ group}$=64, $N_{control\ group}$=61	0.06 (−0.80 to 0.91)	.90	.69
>15	3 [46,54,65]; $N_{VR\ group}$=47, $N_{control\ group}$=48	0.25 (−0.16 to 0.66)	.23	.69
Trial length				
2 weeks to 1 month	9 [10,34,36,37,41,46,54,57,65]; $N_{VR\ group}$=181, $N_{control\ group}$=178	0.17 (−0.20 to 0.54)	.37	.91
>1 and ≤2 months	2 [48,67]; $N_{VR\ group}$=43, $N_{control\ group}$=43	0.26 (−1.15 to 1.66)	.72	.91

[a]VR: virtual reality.

Age

Older patients (SMD 0.47, 95% CI 0.01-0.92; *P*=.05) showed greater improvements in arm and hand motor ability (MFT) than younger patients (SMD −0.52, 95% CI −1.30 to 0.26; *P*=.19); the difference between the groups was significant (*P*=.03; Table S10 in Multimedia Appendix 6).

Moreover, younger patients (SMD 0.49, 95% CI −0.11 to 1.10; *P*=.11) showed greater improvements in quality of life (SIS total score) than older patients (SMD −0.20, 95% CI −0.46 to 0.06; *P*=.13), and the difference between the groups was significant (*P*=.04; Table S13 in Multimedia Appendix 6).

Stroke Recovery Stage

Patients with subacute stroke (SMD 1.13, 95% CI 0.50-1.76; *P*<.001) showed greater improvements in arm and hand motor ability (WMFT task performance score) than those with chronic stroke (SMD −0.07, 95% CI −0.44 to 0.31; *P*=.72), and the difference between the groups was significant (*P*=.001; Table S9 in Multimedia Appendix 6).

In addition, patients with chronic stroke (SMD 3.12, 95% CI 2.26-3.98; *P*<.001) showed greater improvements in hand motor ability (JHFT) than patients with subacute stroke (SMD 0.25, 95% CI −0.16 to 0.67; *P*=.24); the difference between the groups was significant (*P*<.001; Table S11 in Multimedia Appendix 6).

XSL•FO
RenderX

Moreover, patients with chronic stroke (SMD 0.49, 95% CI −0.11 to 1.10; *P*=.11) showed greater improvements in quality of life (SIS total score) than patients with subacute stroke (SMD −0.20, 95% CI −0.46 to 0.06; *P*=.13), and the difference between the groups was significant (*P*=.04; Table S13 in Multimedia Appendix 6).

Type of VR Program Used

The use of specialized programs designed for rehabilitation (SMD 0.49, 95% CI −0.11 to 1.10; *P*=.11) showed greater improvements in quality of life (SIS total score) than those using commercial games (SMD −0.20, 95% CI −0.46 to 0.06; *P*=.13); the difference between the groups was significant (*P*=.04; Table S13 in Multimedia Appendix 6).

Therapy Delivery Format

The use of a combination of VR-supported exercise therapy and conventional therapy (SMD 0.52, 95% CI −0.01 to 1.05; *P*=.052) was associated with greater improvements in hand dexterity (BBT) than the use of VR-supported exercise therapy alone (SMD −0.08, 95% CI −0.34 to 0.18; *P*=.56); the subgroup difference was significant (*P*=.046; Table 4).

Moreover, those using a combination of VR-supported exercise therapy and conventional therapy (SMD 0.49, 95% CI −0.11 to 1.10; *P*=.11) showed greater improvements in quality of life (SIS total score) than those using VR-supported exercise therapy alone (SMD −0.20, 95% CI −0.46 to 0.06; *P*=.13), and the difference between the groups was significant (*P*=.04; Table S13 in Multimedia Appendix 6).

Similarity of Intervention Duration Between Groups

Longer intervention durations for the VR groups (SMD 1.34, 95% CI 0.61-2.07; *P*<.001) were associated with greater improvements in hand dexterity (BBT) than equal intervention durations between the groups (SMD 0.07, 95% CI −0.25 to 0.40; *P*=.66); the subgroup difference was significant (*P*=.002; Table 4).

In addition, longer intervention durations for the VR groups (SMD 0.96, 95% CI 0.36-1.57; *P*=.002) resulted in greater improvements in arm and hand motor ability (WMFT task performance score) than equal intervention durations between the groups (SMD 0.06, 95% CI −0.29 to 0.41; *P*=.72), and the subgroup difference was significant (*P*=.01; Table S9 in Multimedia Appendix 6).

Intervention Duration in VR Groups

The results revealed that receiving >15 hours of VR intervention (SMD 0.92, 95% CI 0.35-1.49; *P*=.002) was associated with significant improvements in hand dexterity (BBT) compared with receiving ≤15 hours of VR intervention (SMD −0.10, 95% CI −0.35 to 0.15; *P*=.45); a significant subgroup difference was observed (*P*<.001; Table 4).

Moreover, receiving >15 hours of VR intervention (SMD 0.33, 95% CI 0.02-0.63; *P*=.04) was associated with a significant decrease in spasticity (AS or mAS) compared with receiving ≤15 hours of VR intervention (SMD −0.50, 95% CI −1.14 to 0.14; *P*=.13); the subgroup difference was significant (*P*=.02; Table S2 in Multimedia Appendix 6).

Trial Length

Receiving VR-supported exercise therapy for >1 month (SMD 0.97, 95% CI 0.06-1.89; *P*=.04) was associated with greater improvements in hand dexterity (BBT) than receiving VR-supported exercise therapy for <1 month (SMD 0.02, 95% CI −0.22 to 0.26; *P*=.84); the difference between the groups was significant (*P*=.049; Table 4).

Furthermore, those who experienced trial lengths of 2 weeks to 1 month (SMD 0.49, 95% CI −0.11 to 1.10; *P*=.11) showed greater improvements in quality of life (SIS total score) than those for whom trial lengths were >1 month (SMD −0.20, 95% CI −0.46 to 0.06; *P*=.13), and the difference between the groups was significant (*P*=.04; Table S13 in Multimedia Appendix 6).

Meta-analysis of the Effects of VR-Supported Exercise Therapy in the Follow-up Assessments

The results of the meta-analyses of outcomes that were examined in the follow-up assessment are presented in Table 7 (from after intervention to follow-up assessment) and Table 8 (from baseline to follow-up assessment). Multimedia Appendix 5 (Figures S21-S44) shows the associated forest plots. Significant improvements (SMD 0.26, 95% CI 0.00-0.51; *P*=.049) in arm and hand motor ability (WMFT task completion time) from baseline to follow-up assessments were observed (Table 8). No statistically significant heterogeneity was observed across trials. No publication bias was observed in the analysis.

XSL•FO
RenderX

Table 7. Meta-analyses of outcomes examined in the follow-up assessments (from after intervention to follow-up assessments).

Outcomes	Tools or methods used to assess the outcomes	Number of trials analyzed and number of participants involved	Standardized mean difference (95% CI)	Between-group difference, *P* value	Heterogeneity			Egger test, *P* value
					Cochran *Q* test	*P* value	I^2 (%)	
Impairments in upper extremity functions and structures								
Upper extremity motor function	Fugl-Meyer Assessment-Upper Extremity	10 [40,41,44,45,50,51,53-55,58]; N_{VR}[a] $_{group}$=160, $N_{control\ group}$=160	0.00 (−0.22 to 0.22)	.99	2.66	.98	0	.91
Activity limitations								
Independence in day-to-day activities	Functional Independence Measure	3 [8,10,44]; $N_{VR\ group}$=166, $N_{control\ group}$=163	−0.05 (−0.27 to 0.17)	.64	0.04	.98	0	.43
Independence in day-to-day activities	Barthel Index or modified Barthel Index	3 [10,41,54]; $N_{VR\ group}$=95, $N_{control\ group}$=94	−0.02 (−0.31 to 0.26)	.87	1.87	.39	0	.32
Arm and hand motor ability	Action Research Arm Test	5 [8,30,38,44,45]; $N_{VR\ group}$=229, $N_{control\ group}$=230	0.10 (−0.08 to 0.29)	.27	1.87	.76	0	.56
Arm and hand motor ability	Wolf Motor Function Test task completion time	5 [10,40,50,51,54]; $N_{VR\ group}$=127, $N_{control\ group}$=125	0.01 (−0.30 to 0.32)	.95	5.07	.28	21	.10
Arm and hand motor ability	Wolf Motor Function Test task performance score	2 [50,54]; $N_{VR\ group}$=18, $N_{control\ group}$=19	−0.24 (−0.88 to 0.41)	.47	0.17	.68	0	N/A[b]
Hand dexterity	Box and Block Test	5 [8,10,41,51,53]; $N_{VR\ group}$=175, $N_{control\ group}$=168	0.13 (−0.09 to 0.34)	.25	1.70	.79	0	.54
Hand motor ability	Jebsen Hand Function Test	2 [41,58]; $N_{VR\ group}$=36, $N_{control\ group}$=33	0.17 (−0.30 to 0.65)	.48	0.38	.54	0	N/A
Participation restrictions in life situations								
Quality of life	Stroke Impact Scale total score	3 [30,53,54]; $N_{VR\ group}$=138, $N_{control\ group}$=140	0.14 (−0.10 to 0.37)	.25	1.48	.48	0	.31
Quality of life	Stroke Impact Scale hand function score	2 [10,44]; $N_{VR\ group}$=104, $N_{control\ group}$=105	−0.19 (−0.46 to 0.08)	.17	0.17	.68	0	N/A
Upper extremity use in daily life	Motor Activity Log quality of movement score	4 [40,50,51,53]; $N_{VR\ group}$=53, $N_{control\ group}$=51	−0.17 (−0.64 to 0.30)	.48	4.11	.25	27	.53
Upper extremity use in daily life	Motor Activity Log amount of use score	3 [40,50,53]; $N_{VR\ group}$=32, $N_{control\ group}$=31	0.06 (−0.43 to 0.56)	.80	0.42	.81	0	.47

[a]VR: virtual reality.

[b]N/A: not applicable.

XSL•FO
RenderX

Table 8. Meta-analyses of outcomes examined in the follow-up assessments (from baseline to follow-up assessments).

Outcomes	Tools or methods used to assess the outcomes	Number of trials analyzed and number of participants involved	Standardized mean difference (95% CI)	Between-group difference, *P* value	Heterogeneity			Egger test, *P* value
					Cochran *Q* test	*P* value	I^2 (%)	
Impairments in upper extremity functions and structures								
Upper extremity motor function	Fugl-Meyer Assessment-Upper Extremity	10 [40,41,44,45,50,51,53-55,58]; N_{VR}[a] $_{group}$=160, $N_{control\ group}$=160	0.48 (−0.15 to 1.11)	.13	62.37	<.001	86	.35
Activity limitations								
In Independence in day-to-day activities	Functional Independence Measure	3 [8,10,44]; $N_{VR\ group}$=166, $N_{control\ group}$=163	−0.05 (−0.27 to 0.16)	.63	0.39	.82	0	.56
Independence in day-to-day activities	Barthel Index or modified Barthel Index	3 [10,41,54]; $N_{VR\ group}$=95, $N_{control\ group}$=94	−0.02 (−0.30 to 0.27)	.90	0.53	.77	0	.54
Arm and hand motor ability	Action Research Arm Test	5 [8,30,38,44,45]; $N_{VR\ group}$=229, $N_{control\ group}$=230	0.03 (−0.15 to 0.22)	.71	1.16	.89	0	.62
Arm and hand motor ability	Wolf Motor Function Test task completion time	5 [10,40,50,51,54]; $N_{VR\ group}$=127, $N_{control\ group}$=125	0.26 (0.00 to 0.51)	.049	4.10	.39	2	.74
Arm and hand motor ability	Wolf Motor Function Test task performance score	2 [50,54]; $N_{VR\ group}$=18, $N_{control\ group}$=19	−0.32 (−0.98 to 0.34)	.34	1.01	.32	1	N/A[b]
Hand dexterity	Box and Block Test	5 [8,10,41,51,53]; $N_{VR\ group}$=175, $N_{control\ group}$=168	0.05 (−0.16 to 0.26)	.66	2.34	.67	0	.62
Hand motor ability	Jebsen Hand Function Test	2 [41,58]; $N_{VR\ group}$=36, $N_{control\ group}$=33	1.81 (−0.85 to 4.46)	.18	19.57	<.001	95	N/A
Participation restrictions in life situations								
Quality of life	Stroke Impact Scale total score	3 [30,53,54]; $N_{VR\ group}$=138, $N_{control\ group}$=140	0.05 (−0.19 to 0.28)	.70	0.88	.64	0	.09
Quality of life	Stroke Impact Scale hand function score	2 [10,44]; $N_{VR\ group}$=104, $N_{control\ group}$=105	−0.25 (−0.52 to 0.02)	.07	0.18	.67	0	N/A
Upper extremity use in daily life	Motor Activity Log quality of movement score	4 [40,50,51,53]; $N_{VR\ group}$=53, $N_{control\ group}$=51	0.17 (−0.21 to 0.56)	.38	1.11	.77	0	.85
Upper extremity use in daily life	Motor Activity Log amount of use score	3 [40,50,53]; $N_{VR\ group}$=32, $N_{control\ group}$=31	0.12 (−0.38 to 0.61)	.64	0.54	.76	0	.83

[a]VR: virtual reality.

[b]N/A: not applicable.

Discussion

Principal Findings

This study included meta-analysis of 43 eligible trials to assess the effects of VR-supported exercise therapy on upper extremity motor rehabilitation in patients following stroke. A total of 12 outcomes regarding impairments in upper extremity functions and structures, activity limitations, and participation restrictions in life situations were examined using 17 tools or methods, with several outcomes examined using different measurement tools or methods. Overall, compared with the use of either conventional therapy or no therapy (ie, control), the use of VR-supported exercise therapy alone or in combination with conventional therapy (ie, intervention) significantly improved 2 outcomes—upper extremity motor function (FMA-UE) and

XSL•FO
RenderX

upper extremity ROM (goniometer). Both significant and nonsignificant improvements were observed in another 2 outcomes, depending on the methods used to measure them: muscle strength (significant when measured by MMT) and independence in day-to-day activities (significant when measured by FIM and modified Rankin Scale). However, as for the other 8 outcomes, the use of VR-supported exercise therapy did not significantly reduce spasticity (AS or mAS) or improve grip strength (dynamometer), upper extremity stroke recovery (Brunnstrom Stages of Stroke Recovery for Upper Extremity), hand dexterity (BBT), arm and hand motor ability (ARAT, WMFT, and MFT), hand motor ability (JHFT), quality of life (SIS), and upper extremity use in daily life (Motor Activity Log).

High-quality evidence was available only for upper extremity motor function (FMA-UE), arm and hand motor ability (ARAT), and independence in day-to-day activities (FIM). In the following sections, we discuss possible explanations for these findings using high-quality evidence. For findings with very low to moderate quality of evidence, further investigation is required before generalizations can be made.

Effects on Upper Extremity Motor Function (FMA-UE)

Our findings contribute further evidence to the literature, showing that VR-supported exercise therapy is effective in improving motor function, especially gross motor function. One possible explanation for our findings is that VR promotes motor learning. First, VR can promote access to therapeutic exercises; it can be used to simulate real-life environments, which allows for real-time interactions and provides a means for individuals to practice therapeutic tasks that may not be feasible to perform in the real world because of resource limitations or safety concerns [69]. Second, virtual environments can provide visual, auditory, or haptic feedback that can facilitate motor skill learning. Such feedback can inform individuals of their success or failure in performing therapeutic tasks [7,69]. Individuals can then make adjustments during tasks. Linking positive feedback to improved or successful therapeutic task performance can also motivate and encourage individuals to engage in rehabilitation therapy [69,70]. Third, VR allows repetitive and intensive therapeutic exercises. Intensive practice can facilitate contraction of muscles involved in exercise and promote muscle coordination [47,71]. At the nervous system level, a large amount of practice can strengthen the connections among neurons and induce reorganization in regions of the cerebral cortex corresponding to the affected extremity, thus improving motor function [69]. Fourth, various types of gaming features can be incorporated in VR-supported exercise therapy protocols, which can be useful for increasing individuals' motivation to perform therapeutic tasks [8,72-75]. For instance, games can set rewards (eg, credits), the pursuit and experience of which motivates users to perform specific behaviors [72]. As another example, games can have different levels of difficulties to meet the needs of different users. Providing appropriate levels of challenges to users can help them avoid boredom or frustration with therapy. Enhanced motivation has been associated with better concentration on therapeutic tasks, higher training intensity, and adherence to therapy [37,69,76].

Effects on Arm and Hand Motor Ability (ARAT)

Our study showed that VR-supported exercise therapy did not have any positive impact on fine motor function improvement (ARAT). The possible explanation for our finding is as follows. In VR-supported exercise therapy, there is a need for interaction with virtual objects, which requires the use of input devices. In most of the reviewed VR-supported exercise therapies, the input devices used were handheld controllers, which required individuals to apply only gross motor skills to hold and move the controllers for interaction (eg, [30,44]). Consequently, fine movements could hardly be involved, and training in them could hardly be achieved. Thus, no significant improvement in fine motor function was observed. This finding suggests that VR systems that use input mechanisms that would facilitate fine motor movements, such as Leap Motion or gloves with sensors [41,63], may be more suitable for supporting fine movement exercises.

Effects on Independence in Day-to-day Activities (FIM)

FIM measures independence in self-care, sphincter control, transfer, locomotion, communication, and social cognition in daily life [77]. Our findings suggest that VR-supported exercise therapy can improve independence in performing such day-to-day activities, which require good upper extremity function. For example, self-care activities, such as eating, bathing, and dressing, usually involve the use of both sides of the upper extremities. Another example is that changing positions from lying down to sitting up may involve the use of the affected upper extremity to support the upper body. As mentioned in the previous section, VR-supported exercise therapy can help improve upper extremity motor function (FMA-UE), enabling patients to participate more actively in the abovementioned day-to-day activities and requiring less assistance from health care providers or caregivers after receiving VR-supported exercise therapy.

Subgroup Analysis of the Effects of VR-Supported Exercise Therapy

We found that the use of VR-supported exercise therapy in combination with conventional therapy, longer VR-supported exercise therapy interventions (ie, >15 hours), and longer trial lengths of VR-supported exercise therapy (ie, >1 and ≤2 months) could improve hand dexterity (BBT), possibly because VR-supported exercise therapy offers longer durations of therapy. Increasing the duration of therapy has been shown to be associated with better motor recovery outcomes [4,7,69,78,79]. It should be noted, however, that motor recovery outcomes are not only determined by the duration of therapy but also by other factors, such as the number of repetitions of the therapeutic tasks, the duration of each training session, the number of sessions, and the frequency of training [7]. More information regarding the details of VR-supported exercise therapy is needed for further analysis before proposing recommendations for the best levels of practice.

Except for the subgroup analyses of hand dexterity (BBT), the number of trials (<10) and participants included in the subgroup analyses for the other outcomes was quite small, implying that these analyses were less likely to produce confirmatory

XSL•FO
RenderX

conclusions [24,80]. Further clinical trials are needed to examine the impact of these moderating factors on the effectiveness of VR-supported exercise therapy.

Effects of VR-Supported Exercise Therapy During Follow-up Assessments

The benefits of VR-supported exercise therapy were not maintained after withdrawing from the technology. However, because we did not have any details on the rehabilitation therapy or exercises that the participants received during the follow-up periods in any of the trials, we could not explore the factors that may have influenced the long-term effects of VR-supported exercise therapy on these outcomes.

Implications for Research

The conclusions of this review have several implications for future studies. First, several trials had small sample sizes (10 trials examined <20 participants) and likely had insufficient statistical power to detect significant changes in the outcomes. Studies with small sample sizes also bear the risk of being less likely to be published [81-83]. Therefore, larger sample sizes are suggested to reduce the risk of failing to detect significant changes and face publication bias. Second, the positive effects of VR-supported exercise therapy were not maintained after withdrawing the technology. However, poststroke rehabilitation and recovery is a long-term, even lifelong, process, and more research is required to determine how best to maintain the long-term effects of VR-supported exercise therapy. Third, most of the VR systems used in the included trials were nonimmersive (eg, Nintendo Wii); the effectiveness of immersive VR-based (eg, head-mounted display) interventions remains relatively less known and should be further examined, as the degree of immersion may influence user experience and the effectiveness of VR-based interventions [84-87].

Implications for Practice

Our review has several practical implications. First, VR-based interventions can be incorporated into therapeutic exercises for motor function training and day-to-day activity training in patients following stroke. Commercial games (eg, Nintendo Wii Sports) appear to be a good option because of their high availability in the market and relatively low prices [62]. Using commercially available games would enable researchers to avoid the costs (eg, time and resources) of designing and developing new games. However, it should be noted that commercial games are typically intended to be played by healthy users and therefore may not meet the heterogeneous needs of patients with impairments [7,88]. For example, commercial games may provide exercises for the overall arm but not for specific joints. To better fulfill the heterogeneous needs of patients and meet specific therapeutic goals, specialized VR programs that allow therapists to customize therapeutic aspects, such as feedback type and difficulty level, based on each patient's condition must be designed [7,69]. Second, patients with stroke are commonly older people [89] who may face difficulties in learning new technologies owing to age-related declines in physical or cognitive functions and other psychological factors (eg, technology anxiety) [90-93]. Therefore, the usability of VR-based interventions must be assessed and improved to provide a user-friendly interface, match the patients' abilities and preferences, and ultimately promote patients' experiences with and acceptance of VR-based interventions, because the acceptance of technology is an essential prerequisite for the successful implementation of technology-based health care interventions [94-107]. Third, as older patients may have limited experience with VR technology [91], the provision of appropriate assistance and guidance is necessary to support patients in learning to use VR input devices and interact with virtual environments.

Limitations

This review has some limitations. First, several study outcomes displayed only a small degree of responsiveness [108-111]; thus, changes in such outcomes may have gone undetected. Second, the risk assessment indicated a low quality of evidence for several outcomes (eg, upper extremity ROM). Therefore, the results related to these study outcomes should be interpreted with caution. Third, the number of trials and participants examined was quite small for several subgroup analyses (eg, Tables S4 and S5 in Multimedia Appendix 6), implying that the findings need to be interpreted cautiously. Fourth, moderate to high levels of heterogeneity were observed in the meta-analysis, which could not be explained by the moderating factors examined and indicated the presence of other moderating factors that require further investigation. Fifth, detection of publication bias suggests that the findings should be interpreted with caution.

Conclusions

This systematic review and meta-analysis provided evidence for the effects of VR-supported exercise therapy on outcomes related to impairments in upper extremity functions and structures, activity limitations, and participation restrictions in life situations. A total of 12 outcomes were examined, some of which were measured using various tools or methods. Of the 12 outcomes, significant improvements were detected in 2, and both significant and nonsignificant improvements were observed in another 2, depending on the measurement tools or methods used. The findings with high-quality evidence suggest that, compared with the use of either conventional therapy or no therapy, VR-supported exercise therapy could effectively improve upper extremity gross motor function (FMA-UE) and independence in daily life (FIM), at least during therapy, but it did not improve fine motor function (ARAT). For findings with low-quality evidence, more research is required before drawing confirmatory conclusions. Future studies should examine how the benefits of VR-supported exercise therapy can be maintained over time.

Acknowledgments

The authors are grateful for the support from Dr Zhenzhen Xie and Dr Yong Zhou for their suggestions and clinical advice.

XSL•FO
RenderX

This review was supported by the Seed Fund for Basic Research (project number 201811159012) at the University of Hong Kong. The funder had no role in the study design, data extraction process, data analysis, result interpretation, or manuscript preparation.

Authors' Contributions

JC and CKO designed the study. JC and TC conducted data screening and data extraction. JC performed data analysis and drafted the manuscript. CKO and TC reviewed and significantly revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 27 KB - jmir_v24i6e24111_app1.docx]

Multimedia Appendix 2
Details of the search.
[DOCX File , 29 KB - jmir_v24i6e24111_app2.docx]

Multimedia Appendix 3
Details of the characteristics of the included trials.
[DOCX File , 48 KB - jmir_v24i6e24111_app3.docx]

Multimedia Appendix 4
Outcomes examined in the included trials.
[XLSX File (Microsoft Excel File), 23 KB - jmir_v24i6e24111_app4.xlsx]

Multimedia Appendix 5
Forest plots of the meta-analyses.
[DOCX File , 2263 KB - jmir_v24i6e24111_app5.docx]

Multimedia Appendix 6
Results of subgroup analyses.
[DOCX File , 492 KB - jmir_v24i6e24111_app6.docx]

References

1. Johnson W, Onuma O, Owolabi M, Sachdev S. Stroke: a global response is needed. Bull World Health Organ 2016 Sep 01;94(9):634-63A [FREE Full text] [doi: 10.2471/BLT.16.181636] [Medline: 27708464]
2. Campbell BC, Khatri P. Stroke. Lancet 2020 Jul 11;396(10244):129-142. [doi: 10.1016/S0140-6736(20)31179-X] [Medline: 32653056]
3. GBD 2016 Lifetime Risk of Stroke Collaborators, Feigin VL, Nguyen G, Cercy K, Johnson CO, Alam T, et al. Global, regional, and country-specific lifetime risks of stroke, 1990 and 2016. N Engl J Med 2018 Dec 20;379(25):2429-2437 [FREE Full text] [doi: 10.1056/NEJMoa1804492] [Medline: 30575491]
4. Langhorne P, Bernhardt J, Kwakkel G. Stroke rehabilitation. Lancet 2011 May 14;377(9778):1693-1702. [doi: 10.1016/S0140-6736(11)60325-5] [Medline: 21571152]
5. Langhorne P, Coupar F, Pollock A. Motor recovery after stroke: a systematic review. Lancet Neurol 2009 Aug;8(8):741-754. [doi: 10.1016/S1474-4422(09)70150-4] [Medline: 19608100]
6. Mayo NE, Wood-Dauphinee S, Côté R, Durcan L, Carlton J. Activity, participation, and quality of life 6 months poststroke. Arch Phys Med Rehabil 2002 Aug;83(8):1035-1042. [doi: 10.1053/apmr.2002.33984] [Medline: 12161823]
7. Levin MF, Demers M. Motor learning in neurological rehabilitation. Disabil Rehabil 2021 Nov 20;43(24):3445-3453. [doi: 10.1080/09638288.2020.1752317] [Medline: 32320305]
8. Brunner I, Skouen JS, Hofstad H, Aßmus J, Becker F, Sanders AM, et al. Virtual reality training for upper extremity in subacute stroke (VIRTUES): a multicenter RCT. Neurology 2017 Dec 12;89(24):2413-2421. [doi: 10.1212/WNL.0000000000004744] [Medline: 29142090]
9. Kiper P, Szczudlik A, Agostini M, Opara J, Nowobilski R, Ventura L, et al. Virtual reality for upper limb rehabilitation in subacute and chronic stroke: a randomized controlled trial. Arch Phys Med Rehabil 2018 May;99(5):834-42.e4. [doi: 10.1016/j.apmr.2018.01.023] [Medline: 29453980]

XSL•FO
RenderX

10. Saposnik G, Cohen LG, Mamdani M, Pooyania S, Ploughman M, Cheung D, Stroke Outcomes Research Canada. Efficacy and safety of non-immersive virtual reality exercising in stroke rehabilitation (EVREST): a randomised, multicentre, single-blind, controlled trial. Lancet Neurol 2016 Sep;15(10):1019-1027 [FREE Full text] [doi: 10.1016/S1474-4422(16)30121-1] [Medline: 27365261]
11. Lee HY, Kim YL, Lee SM. Effects of virtual reality-based training and task-oriented training on balance performance in stroke patients. J Phys Ther Sci 2015 Jun;27(6):1883-1888 [FREE Full text] [doi: 10.1589/jpts.27.1883] [Medline: 26180341]
12. Lohse KR, Hilderman CG, Cheung KL, Tatla S, Van der Loos HF. Virtual reality therapy for adults post-stroke: a systematic review and meta-analysis exploring virtual environments and commercial games in therapy. PLoS One 2014 Mar 28;9(3):e93318 [FREE Full text] [doi: 10.1371/journal.pone.0093318] [Medline: 24681826]
13. Saposnik G, Levin M, Outcome Research Canada (SORCan) Working Group. Virtual reality in stroke rehabilitation: a meta-analysis and implications for clinicians. Stroke 2011 May;42(5):1380-1386. [doi: 10.1161/STROKEAHA.110.605451] [Medline: 21474804]
14. Henderson A, Korner-Bitensky N, Levin M. Virtual reality in stroke rehabilitation: a systematic review of its effectiveness for upper limb motor recovery. Top Stroke Rehabil 2007;14(2):52-61. [doi: 10.1310/tsr1402-52] [Medline: 17517575]
15. Laver KE, Lange B, George S, Deutsch JE, Saposnik G, Crotty M. Virtual reality for stroke rehabilitation. Cochrane Database Syst Rev 2017 Nov 20;11(11):CD008349 [FREE Full text] [doi: 10.1002/14651858.CD008349.pub4] [Medline: 29156493]
16. Aminov A, Rogers JM, Middleton S, Caeyenberghs K, Wilson PH. What do randomized controlled trials say about virtual rehabilitation in stroke? A systematic literature review and meta-analysis of upper-limb and cognitive outcomes. J Neuroeng Rehabil 2018 Mar 27;15:29 [FREE Full text] [doi: 10.1186/s12984-018-0370-2] [Medline: 29587853]
17. Palma GC, Freitas TB, Bonuzzi GM, Soares MA, Leite PH, Mazzini NA, et al. Effects of virtual reality for stroke individuals based on the International Classification of Functioning and Health: a systematic review. Top Stroke Rehabil 2017 May;24(4):269-278. [doi: 10.1080/10749357.2016.1250373] [Medline: 27796177]
18. ICF beginner's guide: towards a common language for functioning, disability and health. World Health Organization. 2002 Jan 1. URL: https://www.who.int/publications/m/item/icf-beginner-s-guide-towards-a-common-language-for-functioning-disability-and-health [accessed 2022-01-03]
19. Domínguez-Téllez P, Moral-Muñoz JA, Salazar A, Casado-Fernández E, Lucena-Antón D. Game-based virtual reality interventions to improve upper limb motor function and quality of life after stroke: systematic review and meta-analysis. Games Health J 2020 Mar;9(1):1-10. [doi: 10.1089/g4h.2019.0043] [Medline: 32027185]
20. Karamians R, Proffitt R, Kline D, Gauthier LV. Effectiveness of virtual reality- and gaming-based interventions for upper extremity rehabilitation poststroke: a meta-analysis. Arch Phys Med Rehabil 2020 May;101(5):885-896. [doi: 10.1016/j.apmr.2019.10.195] [Medline: 31821799]
21. Mekbib DB, Han J, Zhang L, Fang S, Jiang H, Zhu J, et al. Virtual reality therapy for upper limb rehabilitation in patients with stroke: a meta-analysis of randomized clinical trials. Brain Inj 2020 Mar 20;34(4):456-465. [doi: 10.1080/02699052.2020.1725126] [Medline: 32064964]
22. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
23. Orwin R. Evaluating coding decisions. In: Cooper H, Hedges LV, editors. The Handbook of Research Synthesis. New York, NY, USA: Russell Sage Foundation; 1994:139-162.
24. Higgins JP, 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). London, UK: Cochrane; 2021.
25. Viechtbauer W, Cheung MW. Outlier and influence diagnostics for meta-analysis. Res Synth Methods 2010 Apr;1(2):112-125. [doi: 10.1002/jrsm.11] [Medline: 26061377]
26. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 06;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
27. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
28. Bernhardt J, Hayward KS, Kwakkel G, Ward NS, Wolf SL, Borschmann K, et al. Agreed definitions and a shared vision for new standards in stroke recovery research: the Stroke Recovery and Rehabilitation Roundtable taskforce. Int J Stroke 2017 Jul;12(5):444-450. [doi: 10.1177/1747493017711816] [Medline: 28697708]
29. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008 Apr 26;336(7650):924-926 [FREE Full text] [doi: 10.1136/bmj.39489.470347.AD] [Medline: 18436948]
30. Adie K, Schofield C, Berrow M, Wingham J, Humfryes J, Pritchard C, et al. Does the use of Nintendo Wii Sports TM improve arm function? Trial of Wii TM in stroke: a randomized controlled trial and economics analysis. Clin Rehabil 2017 Mar;31(2):173-185. [doi: 10.1177/0269215516637893] [Medline: 26975313]

XSL•FO
RenderX

31. Ikbali Afsar S, Mirzayev I, Umit Yemisci O, Cosar Saracgil SN. Virtual reality in upper extremity rehabilitation of stroke patients: a randomized controlled trial. J Stroke Cerebrovasc Dis 2018 Dec;27(12):3473-3478. [doi: 10.1016/j.jstrokecerebrovasdis.2018.08.007] [Medline: 30193810]
32. Ain QU, Khan S, Ilyas S, Yaseen A, Tariq I, Liu T, et al. Additional effects of Xbox Kinect training on upper limb function in chronic stroke patients: a randomized control trial. Healthcare 2021 Feb 24;9(3):242 [FREE Full text] [doi: 10.3390/healthcare9030242] [Medline: 33668355]
33. Alves SS, Ocamoto GN, de Camargo PS, Santos AT, Terra AM. Effects of virtual reality and motor imagery techniques using Fugl Meyer Assessment scale in post-stroke patients. Int J Ther Rehabil 2018 Nov 02;25(11):587-596. [doi: 10.12968/ijtr.2018.25.11.587]
34. Anjum AF, Jawwad G, Khokhar A, Sadiq N, Masud R, Khalid AM. Effect of "Wii-habilitation" and constraint induced movement therapy on improving quality of life in stroke survivors. RMJ 2021;46(1):220-223.
35. Aşkın A, Atar E, Koçyiğit H, Tosun A. Effects of Kinect-based virtual reality game training on upper extremity motor recovery in chronic stroke. Somatosens Mot Res 2018 Mar;35(1):25-32. [doi: 10.1080/08990220.2018.1444599] [Medline: 29529919]
36. Cho HY, Song E, Moon JH, Hahm SC. Effects of virtual reality based therapeutic exercise on the upper extremity function and activities of daily living in patients with acute stroke: a pilot randomized controlled trial. Med Leg Update 2021 Mar 19;21(2):676-682. [doi: 10.37506/mlu.v21i2.2761]
37. Choi JH, Han EY, Kim BR, Kim SM, Im SH, Lee SY, et al. Effectiveness of commercial gaming-based virtual reality movement therapy on functional recovery of upper extremity in subacute stroke patients. Ann Rehabil Med 2014 Aug;38(4):485-493 [FREE Full text] [doi: 10.5535/arm.2014.38.4.485] [Medline: 25229027]
38. Crosbie JH, Lennon S, McGoldrick MC, McNeill MD, McDonough SM. Virtual reality in the rehabilitation of the arm after hemiplegic stroke: a randomized controlled pilot study. Clin Rehabil 2012 Sep;26(9):798-806. [doi: 10.1177/0269215511434575] [Medline: 22275463]
39. Ersoy C, Iyigun G. Boxing training in patients with stroke causes improvement of upper extremity, balance, and cognitive functions but should it be applied as virtual or real? Top Stroke Rehabil 2021 Mar;28(2):112-126. [doi: 10.1080/10749357.2020.1783918] [Medline: 32574096]
40. Hung JW, Chou CX, Chang YJ, Wu CY, Chang KC, Wu WC, et al. Comparison of Kinect2Scratch game-based training and therapist-based training for the improvement of upper extremity functions of patients with chronic stroke: a randomized controlled single-blinded trial. Eur J Phys Rehabil Med 2019 Oct;55(5):542-550 [FREE Full text] [doi: 10.23736/S1973-9087.19.05598-9] [Medline: 30781936]
41. Kang MG, Yun SJ, Lee SY, Oh BM, Lee HH, Lee SU, et al. Effects of upper-extremity rehabilitation using smart glove in patients with subacute stroke: results of a prematurely terminated multicenter randomized controlled trial. Front Neurol 2020 Nov 9;11:580393 [FREE Full text] [doi: 10.3389/fneur.2020.580393] [Medline: 33240205]
42. Kiper P, Agostini M, Luque-Moreno C, Tonin P, Turolla A. Reinforced feedback in virtual environment for rehabilitation of upper extremity dysfunction after stroke: preliminary data from a randomized controlled trial. Biomed Res Int 2014;2014:752128 [FREE Full text] [doi: 10.1155/2014/752128] [Medline: 24745024]
43. Kiper P, Piron L, Turolla A, Stożek J, Tonin P. The effectiveness of reinforced feedback in virtual environment in the first 12 months after stroke. Neurol Neurochir Pol 2011;45(5):436-444. [doi: 10.1016/s0028-3843(14)60311-x] [Medline: 22127938]
44. Kong KH, Loh YJ, Thia E, Chai A, Ng CY, Soh YM, et al. Efficacy of a virtual reality commercial gaming device in upper limb recovery after stroke: a randomized, controlled study. Top Stroke Rehabil 2016 Oct;23(5):333-340. [doi: 10.1080/10749357.2016.1139796] [Medline: 27098818]
45. Kottink AI, Prange GB, Krabben T, Rietman JS, Buurke JH. Gaming and conventional exercises for improvement of arm function after stroke: a randomized controlled pilot study. Games Health J 2014 Jun;3(3):184-191. [doi: 10.1089/g4h.2014.0026] [Medline: 26196178]
46. Kwon JS, Park MJ, Yoon IJ, Park SH. Effects of virtual reality on upper extremity function and activities of daily living performance in acute stroke: a double-blind randomized clinical trial. NeuroRehabilitation 2012;31(4):379-385. [doi: 10.3233/NRE-2012-00807] [Medline: 23232161]
47. Lee G. Effects of training using video games on the muscle strength, muscle tone, and activities of daily living of chronic stroke patients. J Phys Ther Sci 2013 May;25(5):595-597 [FREE Full text] [doi: 10.1589/jpts.25.595] [Medline: 24259810]
48. Lee M, Son J, Kim J, Pyun SB, Eun SD, Yoon BC. Comparison of individualized virtual reality- and group-based rehabilitation in older adults with chronic stroke in community settings: a pilot randomized controlled trial. Eur J Integr Med 2016 Oct;8(5):738-746. [doi: 10.1016/j.eujim.2016.08.166]
49. Lee MM, Lee KJ, Song CH. Game-based virtual reality canoe paddling training to improve postural balance and upper extremity function: a preliminary randomized controlled study of 30 patients with subacute stroke. Med Sci Monit 2018 Apr 27;24:2590-2598 [FREE Full text] [doi: 10.12659/MSM.906451] [Medline: 29702630]
50. Levin MF, Snir O, Liebermann DG, Weingarden H, Weiss PL. Virtual reality versus conventional treatment of reaching ability in chronic stroke: clinical feasibility study. Neurol Ther 2012 Aug 24;1:3 [FREE Full text] [doi: 10.1007/s40120-012-0003-9] [Medline: 26000209]

XSL•FO
RenderX

51. McNulty PA, Thompson-Butel AG, Faux SG, Lin G, Katrak PH, Harris LR, et al. The efficacy of Wii-based Movement Therapy for upper limb rehabilitation in the chronic poststroke period: a randomized controlled trial. Int J Stroke 2015 Dec;10(8):1253-1260. [doi: 10.1111/ijs.12594] [Medline: 26332338]
52. Miclaus R, Roman N, Caloian S, Mitoiu B, Suciu O, Onofrei RR, et al. Non-immersive virtual reality for post-stroke upper extremity rehabilitation: a small cohort randomized trial. Brain Sci 2020 Sep 21;10(9):655 [FREE Full text] [doi: 10.3390/brainsci10090655] [Medline: 32967160]
53. Norouzi-Gheidari N, Hernandez A, Archambault PS, Higgins J, Poissant L, Kairy D. Feasibility, safety and efficacy of a virtual reality exergame system to supplement upper extremity rehabilitation post-stroke: a pilot randomized clinical trial and proof of principle. Int J Environ Res Public Health 2019 Dec 23;17(1):113 [FREE Full text] [doi: 10.3390/ijerph17010113] [Medline: 31877910]
54. Park M, Ko MH, Oh SW, Lee JY, Ham Y, Yi H, et al. Effects of virtual reality-based planar motion exercises on upper extremity function, range of motion, and health-related quality of life: a multicenter, single-blinded, randomized, controlled pilot study. J Neuroeng Rehabil 2019 Oct 24;16:122 [FREE Full text] [doi: 10.1186/s12984-019-0595-8] [Medline: 31651335]
55. Piron L, Turolla A, Agostini M, Zucconi C, Cortese F, Zampolini M, et al. Exercises for paretic upper limb after stroke: a combined virtual-reality and telemedicine approach. J Rehabil Med 2009 Nov;41(12):1016-1020 [FREE Full text] [doi: 10.2340/16501977-0459] [Medline: 19841835]
56. Piron L, Turolla A, Agostini M, Zucconi CS, Ventura L, Tonin P, et al. Motor learning principles for rehabilitation: a pilot randomized controlled study in poststroke patients. Neurorehabil Neural Repair 2010;24(6):501-508. [doi: 10.1177/1545968310362672] [Medline: 20581337]
57. Shin JH, Ryu H, Jang SH. A task-specific interactive game-based virtual reality rehabilitation system for patients with stroke: a usability test and two clinical experiments. J Neuroeng Rehabil 2014 Mar 06;11:32 [FREE Full text] [doi: 10.1186/1743-0003-11-32] [Medline: 24597650]
58. Shin JH, Kim MY, Lee JY, Jeon YJ, Kim S, Lee S, et al. Effects of virtual reality-based rehabilitation on distal upper extremity function and health-related quality of life: a single-blinded, randomized controlled trial. J Neuroeng Rehabil 2016 Feb 24;13:17 [FREE Full text] [doi: 10.1186/s12984-016-0125-x] [Medline: 26911438]
59. Şimşek TT, Çekok K. The effects of Nintendo Wii(TM)-based balance and upper extremity training on activities of daily living and quality of life in patients with sub-acute stroke: a randomized controlled study. Int J Neurosci 2016 Dec;126(12):1061-1070. [doi: 10.3109/00207454.2015.1115993] [Medline: 26626539]
60. Sin H, Lee G. Additional virtual reality training using Xbox Kinect in stroke survivors with hemiplegia. Am J Phys Med Rehabil 2013 Oct;92(10):871-880. [doi: 10.1097/PHM.0b013e3182a38e40] [Medline: 24051993]
61. Standen PJ, Threapleton K, Richardson A, Connell L, Brown DJ, Battersby S, et al. A low cost virtual reality system for home based rehabilitation of the arm following stroke: a randomised controlled feasibility trial. Clin Rehabil 2017 Mar;31(3):340-350 [FREE Full text] [doi: 10.1177/0269215516640320] [Medline: 27029939]
62. Türkbey TA, Kutlay S, Gök H. Clinical feasibility of Xbox KinectTM training for stroke rehabilitation: a single-blind randomized controlled pilot study. J Rehabil Med 2017 Jan 19;49(1):22-29 [FREE Full text] [doi: 10.2340/16501977-2183] [Medline: 27973678]
63. Xie H, Zhang H, Liang H, Fan H, Zhou J, Ambrose Lo WL, et al. A novel glasses-free virtual reality rehabilitation system on improving upper limb motor function among patients with stroke: a feasibility pilot study. Med Novel Technol Devices 2021 Sep;11:100069. [doi: 10.1016/j.medntd.2021.100069]
64. Zondervan DK, Friedman N, Chang E, Zhao X, Augsburger R, Reinkensmeyer DJ, et al. Home-based hand rehabilitation after chronic stroke: randomized, controlled single-blind trial comparing the MusicGlove with a conventional exercise program. J Rehabil Res Dev 2016;53(4):457-472 [FREE Full text] [doi: 10.1682/JRRD.2015.04.0057] [Medline: 27532880]
65. Park YS, An CS, Lim CG. Effects of a rehabilitation program using a wearable device on the upper limb function, performance of activities of daily living, and rehabilitation participation in patients with acute stroke. Int J Environ Res Public Health 2021 May 21;18(11):5524 [FREE Full text] [doi: 10.3390/ijerph18115524] [Medline: 34063970]
66. Jo K, Yu J, Jung J. Effects of virtual reality-based rehabilitation on upper extremity function and visual perception in stroke patients: a randomized control trial. J Phys Ther Sci 2012;24(11):1205-1208. [doi: 10.1589/jpts.24.1205]
67. Mokhtar MM, Atteya M, Rawash M. Virtual reality Xbox 360 Kinect training for stroke patients with hemiplegia. Biosci Res 2019;16(1):672-676.
68. Wang ZR, Wang P, Xing L, Mei LP, Zhao J, Zhang T. Leap Motion-based virtual reality training for improving motor functional recovery of upper limbs and neural reorganization in subacute stroke patients. Neural Regen Res 2017 Nov;12(11):1823-1831 [FREE Full text] [doi: 10.4103/1673-5374.219043] [Medline: 29239328]
69. Levac DE, Sveistrup H. Motor learning and virtual reality. In: Weiss PL, Keshner EA, Levin MF, editors. Virtual Reality for Physical and Motor Rehabilitation. New York, NY, USA: Springer; 2014:25-46.
70. Holden MK. Virtual environments for motor rehabilitation: review. Cyberpsychol Behav 2005 Jun;8(3):187-211. [doi: 10.1089/cpb.2005.8.187] [Medline: 15971970]

XSL•FO
RenderX

71. Carr JH, Shepherd RB. Chapter 1 - Brain reorganization, the rehabilitation environment, measuring outcomes. In: Carr JH, Shepherd RB, editors. Stroke Rehabilitation: Guidelines for Exercise and Training to Optimize Motor Skill. Oxford, UK: Butterworth-Heinemann; 2003:3-31.
72. Lohse K, Shirzad N, Verster A, Hodges N, Van der Loos HF. Video games and rehabilitation: using design principles to enhance engagement in physical therapy. J Neurol Phys Ther 2013 Dec;37(4):166-175. [doi: 10.1097/NPT.0000000000000017] [Medline: 24232363]
73. Lewis GN, Woods C, Rosie JA, McPherson KM. Virtual reality games for rehabilitation of people with stroke: perspectives from the users. Disabil Rehabil Assist Technol 2011;6(5):453-463. [doi: 10.3109/17483107.2011.574310] [Medline: 21495917]
74. Cheung KL, Tunik E, Adamovich SV, Boyd LA. Neuroplasticity and virtual reality. In: Weiss PL, Keshner EA, Levin MF, editors. Virtual Reality for Physical and Motor Rehabilitation. New York, NY, USA: Springer; 2014:5-24.
75. Wong RS, Yu EY, Wong TW, Fung CS, Choi CS, Or CK, et al. Development and pilot evaluation of a mobile app on parent-child exercises to improve physical activity and psychosocial outcomes of Hong Kong Chinese children. BMC Public Health 2020 Oct 14;20(1):1544 [FREE Full text] [doi: 10.1186/s12889-020-09655-9] [Medline: 33054753]
76. Rohrbach N, Chicklis E, Levac DE. What is the impact of user affect on motor learning in virtual environments after stroke? A scoping review. J Neuroeng Rehabil 2019 Jun 27;16(1):79 [FREE Full text] [doi: 10.1186/s12984-019-0546-4] [Medline: 31248439]
77. Kidd D, Stewart G, Baldry J, Johnson J, Rossiter D, Petruckevitch A, et al. The Functional Independence Measure: a comparative validity and reliability study. Disabil Rehabil 1995 Jan;17(1):10-14. [doi: 10.3109/09638289509166622] [Medline: 7858276]
78. Veerbeek JM, van Wegen E, van Peppen R, van der Wees PJ, Hendriks E, Rietberg M, et al. What is the evidence for physical therapy poststroke? A systematic review and meta-analysis. PLoS One 2014 Feb 4;9(2):e87987 [FREE Full text] [doi: 10.1371/journal.pone.0087987] [Medline: 24505342]
79. Kwakkel G, van Peppen R, Wagenaar RC, Wood Dauphinee S, Richards C, Ashburn A, et al. Effects of augmented exercise therapy time after stroke: a meta-analysis. Stroke 2004 Nov;35(11):2529-2539. [doi: 10.1161/01.STR.0000143153.76460.7d] [Medline: 15472114]
80. Richardson M, Garner P, Donegan S. Interpretation of subgroup analyses in systematic reviews: a tutorial. Clin Epidemiol Glob Health 2019 Jun 01;7(2):192-198. [doi: 10.1016/j.cegh.2018.05.005]
81. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. JAMA 1994 Jul 13;272(2):122-124. [Medline: 8015121]
82. Lin L. Bias caused by sampling error in meta-analysis with small sample sizes. PLoS One 2018 Sep 13;13(9):e0204056 [FREE Full text] [doi: 10.1371/journal.pone.0204056] [Medline: 30212588]
83. Macaskill P, Walter SD, Irwig L. A comparison of methods to detect publication bias in meta-analysis. Stat Med 2001 Mar 28;20(4):641-654. [doi: 10.1002/sim.698] [Medline: 11223905]
84. Gorini A, Capideville CS, De Leo G, Mantovani F, Riva G. The role of immersion and narrative in mediated presence: the virtual hospital experience. Cyberpsychol Behav Soc Netw 2011 Mar;14(3):99-105. [doi: 10.1089/cyber.2010.0100] [Medline: 20649451]
85. Shahrbanian S, Ma X, Aghaei N, Korner-Bitensky N, Moshiri K, Simmonds MJ. Use of virtual reality (immersive vs. non immersive) for pain management in children and adults: a systematic review of evidence from randomized controlled trials. Eur J Exp Biol 2012;2(5):1408-1422 [FREE Full text]
86. Chen J, Xie Z, Kalun Or C. Head-mounted display virtual reality in disease treatment: a systematic review and meta-analysis. Proc Hum Factors Ergon Soc Annu Meet 2020;64(1):1388-1389. [doi: 10.1177/1071181320641331]
87. Chen J, Xie Z, Or C. Effectiveness of immersive virtual reality-supported interventions for patients with disorders or impairments: a systematic review and meta-analysis. Health Technol 2021 Jul 16;11:811-833. [doi: 10.1007/s12553-021-00561-7]
88. Barrett N, Swain I, Gatzidis C, Mecheraoui C. The use and effect of video game design theory in the creation of game-based systems for upper limb stroke rehabilitation. J Rehabil Assist Technol Eng 2016 May 9;3:1-16 [FREE Full text] [doi: 10.1177/2055668316643644] [Medline: 31186903]
89. Hall MJ, Levant S, DeFrances CJ. Hospitalization for stroke in U.S. hospitals, 1989-2009. NCHS Data Brief 2012 May(95):1-8 [FREE Full text] [Medline: 22617404]
90. Deng Z, Mo X, Liu S. Comparison of the middle-aged and older users' adoption of mobile health services in China. Int J Med Inform 2014 Mar;83(3):210-224. [doi: 10.1016/j.ijmedinf.2013.12.002] [Medline: 24388129]
91. Chen J, Or C. Assessing the use of immersive virtual reality, mouse and touchscreen in pointing and dragging-and-dropping tasks among young, middle-aged and older adults. Appl Ergon 2017 Nov;65:437-448. [doi: 10.1016/j.apergo.2017.03.013] [Medline: 28395855]
92. Ng HC, Tao D, Or CK. Age differences in computer input device use: a comparison of touchscreen, trackball, and mouse. In: Proceedings of the 2013 World Conference on Information Systems and Technologies. 2013 Presented at: WorldCIST '13; March 27-30, 2013; Algarve, Portugal p. 1015-1024. [doi: 10.1007/978-3-642-36981-0_96]

XSL•FO
RenderX

93. Or CK, Valdez RS, Casper GR, Carayon P, Burke LJ, Brennan PF, et al. Human factors and ergonomics in home care: current concerns and future considerations for health information technology. Work 2009;33(2):201-209 [FREE Full text] [doi: 10.3233/WOR-2009-0867] [Medline: 19713630]
94. Chen T, Or CK, Chen J. Effects of technology-supported exercise programs on the knee pain, physical function, and quality of life of individuals with knee osteoarthritis and/or chronic knee pain: a systematic review and meta-analysis of randomized controlled trials. J Am Med Inform Assoc 2021 Feb 15;28(2):414-423 [FREE Full text] [doi: 10.1093/jamia/ocaa282] [Medline: 33236109]
95. Or C, Tao D. Usability study of a computer-based self-management system for older adults with chronic diseases. JMIR Res Protoc 2012 Nov 08;1(2):e13 [FREE Full text] [doi: 10.2196/resprot.2184] [Medline: 23612015]
96. Yan M, Or C. A 12-week pilot study of acceptance of a computer-based chronic disease self-monitoring system among patients with type 2 diabetes mellitus and/or hypertension. Health Informatics J 2019 Sep;25(3):828-843 [FREE Full text] [doi: 10.1177/1460458217724580] [Medline: 28820007]
97. Xie Z, Liu K, Or C, Chen J, Yan M, Wang H. An examination of the socio-demographic correlates of patient adherence to self-management behaviors and the mediating roles of health attitudes and self-efficacy among patients with coexisting type 2 diabetes and hypertension. BMC Public Health 2020 Aug 12;20:1227 [FREE Full text] [doi: 10.1186/s12889-020-09274-4] [Medline: 32787809]
98. Cheung DS, Or CK, So MK, Tiwari A. Usability testing of a smartphone application for delivering Qigong training. J Med Syst 2018 Sep 05;42(10):191. [doi: 10.1007/s10916-018-1048-9] [Medline: 30187139]
99. Liu K, Or CK, So M, Cheung B, Chan B, Tiwari A, et al. A longitudinal examination of tablet self-management technology acceptance by patients with chronic diseases: integrating perceived hand function, perceived visual function, and perceived home space adequacy with the TAM and TPB. Appl Ergon 2022 Apr;100:103667. [doi: 10.1016/j.apergo.2021.103667] [Medline: 34920356]
100. Or CK, Karsh BT, Severtson DJ, Burke LJ, Brown RL, Brennan PF. Factors affecting home care patients' acceptance of a Web-based interactive self-management technology. J Am Med Inform Assoc 2011;18(1):51-59 [FREE Full text] [doi: 10.1136/jamia.2010.007336] [Medline: 21131605]
101. Chen T, Or CK. Development and pilot test of a machine learning-based knee exercise system with video demonstration, real-time feedback, and exercise performance score. Proc Hum Factors Ergon Soc Annu Meet 2021 Nov 12;65(1):1519-1523. [doi: 10.1177/1071181321651109]
102. Xie Z, Kalun Or C. Acceptance of mHealth by elderly adults: a path analysis. Proc Hum Factors Ergon Soc Annu Meet 2021 Feb 09;64(1):755-759. [doi: 10.1177/1071181320641174]
103. Chen T, Kalun Or C, Chen J. A systematic review and meta-analysis of randomized controlled trials to evaluate technology-supported exercise programs for knee health. Proc Hum Factors Ergon Soc Annu Meet 2021 Feb 09;64(1):639-640. [doi: 10.1177/1071181320641145]
104. Karsh BT, Holden RJ, Or CK. Human factors and ergonomics of health information technology implementation. In: Carayon P, editor. Handbook of Human Factors and Ergonomics in Health Care and Patient Safety. 2nd ed. Boca Raton, FL, USA: CRC Press; 2011:249-264.
105. Or CK, Holden RJ, Valdez R. Human factors engineering and user-centered design for mobile health technology: enhancing effectiveness, efficiency, and satisfaction. Human-Automation Interaction: Mobile Computing 2022 (forthcoming).
106. Or CK, Liu K, So MK, Cheung B, Yam LY, Tiwari A, et al. Improving self-care in patients with coexisting type 2 diabetes and hypertension by technological surrogate nursing: randomized controlled trial. J Med Internet Res 2020 Mar 27;22(3):e16769 [FREE Full text] [doi: 10.2196/16769] [Medline: 32217498]
107. Or CK, Karsh BT. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009;16(4):550-560 [FREE Full text] [doi: 10.1197/jamia.M2888] [Medline: 19390112]
108. Hsieh YW, Wu CY, Lin KC, Chang YF, Chen CL, Liu JS. Responsiveness and validity of three outcome measures of motor function after stroke rehabilitation. Stroke 2009 Apr;40(4):1386-1391. [doi: 10.1161/STROKEAHA.108.530584] [Medline: 19228851]
109. Sears ED, Chung KC. Validity and responsiveness of the Jebsen-Taylor Hand Function Test. J Hand Surg Am 2010 Jan;35(1):30-37 [FREE Full text] [doi: 10.1016/j.jhsa.2009.09.008] [Medline: 19954898]
110. Schepers VP, Ketelaar M, Visser-Meily JM, Dekker J, Lindeman E. Responsiveness of functional health status measures frequently used in stroke research. Disabil Rehabil 2006 Sep 15;28(17):1035-1040. [doi: 10.1080/09638280500494694] [Medline: 16950733]
111. Santisteban L, Térémetz M, Bleton JP, Baron JC, Maier MA, Lindberg PG. Upper limb outcome measures used in stroke rehabilitation studies: a systematic literature review. PLoS One 2016 May 6;11(5):e0154792 [FREE Full text] [doi: 10.1371/journal.pone.0154792] [Medline: 27152853]

Abbreviations

ARAT: Action Research Arm Test
AS: Ashworth Scale

XSL•FO
RenderX

BBT: Box and Block Test
FIM: Functional Independence Measure
FMA-UE: Fugl-Meyer Assessment-Upper Extremity
JHFT: Jebsen Hand Function Test
mAS: Modified Ashworth Scale
MFT: Manual Function Test
MMT: Manual Muscle Testing
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
ROM: range of motion
SIS: Stroke Impact Scale
SMD: standardized mean difference
VR: virtual reality
WMFT: Wolf Motor Function Test

Edited by A Mavragani; submitted 07.09.20; peer-reviewed by M Demers, I Cikajlo, L Sheehy; comments to author 04.12.20; revised version received 19.10.21; accepted 18.04.22; published 20.06.22.

Please cite as:
Chen J, Or CK, Chen T
Effectiveness of Using Virtual Reality–Supported Exercise Therapy for Upper Extremity Motor Rehabilitation in Patients With Stroke: Systematic Review and Meta-analysis of Randomized Controlled Trials
J Med Internet Res 2022;24(6):e24111
URL: https://www.jmir.org/2022/6/e24111
doi:10.2196/24111
PMID:35723907

©Jiayin Chen, Calvin Kalun Or, Tianrong Chen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.06.2022. 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

Contexts and Outcomes of Proxy Online Health Information Seeking: Mixed Studies Review With Framework Synthesis

Reem El Sherif[1], MBChB, MSc; Pierre Pluye[1], MD, PhD; Fidelia Ibekwe[2], PhD

[1]Department of Family Medicine, McGill University, Montreal, QC, Canada

[2]School of Journalism & Communication, Aix-Marseille University, Aix-en-Provence, France

Corresponding Author:
Reem El Sherif, MBChB, MSc
Department of Family Medicine
McGill University
5858 Chemin de la Côte-des-Neiges
Montreal, QC, H3S 1Z1
Canada
Phone: 1 (514) 398 7375
Email: reem.elsherif@mail.mcgill.ca

Abstract

Background: High-quality online health information (OHI) can reduce unnecessary visits to health professionals and improve health. One of the ways that people use OHI is to support others with health conditions through proxy OHI seeking. Members of a person's social circle may help them overcome information-seeking barriers and illness challenges. There are several models on proxy information seeking. Yet, we know little about the use and outcomes of OHI on behalf of someone else.

Objective: The objectives of this paper are to explore and revise a framework on the context and outcomes of proxy OHI seeking

Methods: We conducted a mixed studies literature review integrating qualitative and quantitative evidence with thematic analysis of the findings of 28 studies, followed by framework synthesis incorporating the derived themes.

Results: We explored 4 main themes: (1) characteristics of proxy seekers, (2) context of proxy OHI seeking, (3) use of OHI to provide social support, and (4) outcomes of proxy OHI seeking. Our conceptual framework incorporates these themes and builds on previous work.

Conclusions: By better understanding how people use information together, information providers can adapt the information to meet all users' needs.

(J Med Internet Res 2022;24(6):e34345) doi:10.2196/34345

KEYWORDS

online health information; information seeking behavior; proxy information seeking; surrogate seekers; information outcomes; social support; health information; online information

Introduction

Background

Two-thirds (67%) of respondents to the American Health Information National Trends Survey between 2008 and 2017 reported turning to the internet first for health information [1]. Similarly, 69% of Canadians reported using the internet to search for health information in 2020 [2], and the proportion of adults seeking online health information (OHI) in other Organisation for Economic Cooperation and Development (OECD) countries more than doubled between 2008 and 2017 [3]. The use of OHI can improve quality of life and is generally associated with positive outcomes, such as increased empowerment of seekers and their families and improved health outcomes [4-7].

Based on the results of a recent systematic review on the outcomes of OHI seeking (hereafter, OHI outcomes), several contextual factors associated with these outcomes were identified, such as age, education, income, and eHealth literacy [8]. Another contextual factor is social support, defined broadly as "support accessible to an individual through social ties to other individuals, groups, and the larger community"[9]. Social support is an important factor because one of the ways people use OHI is to support family members or friends with health

XSL•FO
RenderX

conditions [10]. In fact, recent studies report that 61%-66% of OHI seekers are proxy seekers, meaning they seek OHI on behalf of someone else [11,12]. Moreover, findings from a study exploring internet use trends between 2008 and 2013 showed a significant increase in the use of family and friends to obtain health information [13].

However, while proxy information seeking has been explored in the literature, especially as it relates to health information, little is known about its relationship with the outcomes of OHI. This is a critical knowledge gap; previous research examining how to reduce negative outcomes of OHI suggests that OHI seekers may be able to overcome low eHealth literacy by discussing the information they find with others [14]. People are sometimes more likely to turn to their social circle to make sense of information they find rather than discuss it with a health professional [11,15]. Members of a person's social circle may help them overcome information-seeking barriers and illness challenges (eg, if they are too physically weak or mentally incapacitated to search themselves) [15].

By better understanding how people and their social circles use information together, information providers can better adapt the information to meet both their needs, and public health interventions can target patients' friends and family with information for dissemination and use [16]. Accordingly, the purpose of this paper is to contribute to our understanding of the role of social support in online health information outcomes by focusing on the outcomes of proxy OHI seeking.

This review will focus on the intersection of 3 main constructs: proxy information seeking, social support, and OHI outcomes.

Proxy Information Seeking

Information seeking encompasses "all the information that comes to a human being during a lifetime, not just in those moments when a person actively seeks information" [17]. In active information seeking mode, monitoring and directed searching are ways to answer known information needs (that are recognized and articulated). There are intervening variables that may be related to personal characteristics, social or interpersonal issues, or environmental considerations [18]. They can be defined as "those who seek information in a nonprofessional or informal capacity on behalf (or because) of others without necessarily being asked to do so" [15]. Proxy seekers may also be "experts," such as health librarians or health care professionals, with the specialized knowledge or skills to use the information with the person with whom they share a personal relationship [19].

The role of proxy information seeking has been explored in the literature and has also been referred to as surrogate seeking or lay information mediation [12,20]. In one of the earliest models on information seeking behavior, Wilson [21] used pathways to explain different patterns of information seeking. In his model, the user encounters "information systems" that can be technology (eg, the internet) and mediators, and these systems connect the user to "information resources" or actual information. Of 10 pathways proposed in this model, 2 indicate seeking that is "conducted by a mediator to fulfill an information request" [21]. This phenomenon is also described in McKenzie's [22] 2-dimensional model of information practices of women pregnant with twins. In her model, one of the modes of information practice is "by proxy," where the person interacts with information through another agent, including "intermediaries or gatekeepers" such as friends or family members.

Social Support

Social support is one of the positive products of "social relationships" that may have short- and long-term effects on health, for better and for worse, depending on their quality and quantity [23]. A 2004 model by Uchino [24] describes 2 broad dimension of support: structure and function. Structural aspects of support are the extent or composition of one's social network (size, contact, type, density, and strength) and the interconnections among them. Functions have 4 aspects that are highly related to each other: emotional, informational, tangible, and belonging. Most relevant to this review is informational support, which includes the provision of advice or guidance and may provide direction and carry an emotional message when received from a close source. Informational support could be construed as supportive, unsupportive, or mixed depending on the context [25-27].

Social support has consistently been linked to better health [24,28,29]. Several theories have been proposed to explain why this occurs; for example, social support can act as a mediator of stress that reduces its impact, thereby improving mental health [23]. Several studies have reported that those who perceive low social support experience increased stress and report a greater number of stressful events, while those who feel more satisfaction with their received social support report fewer emotional problems [30-33]. Another theory to explain the link between social support and better health is the provision of informational support, which encourages the receivers to manage their health. If we use pregnant women as an example, those who were more satisfied with perceived and received social support initiated prenatal care earlier than those who were less satisfied [34]. Pregnant women who received more informational support from people in their social network delivered babies with higher Apgar scores and higher birth weights [34,35].

Online Health Information Outcomes

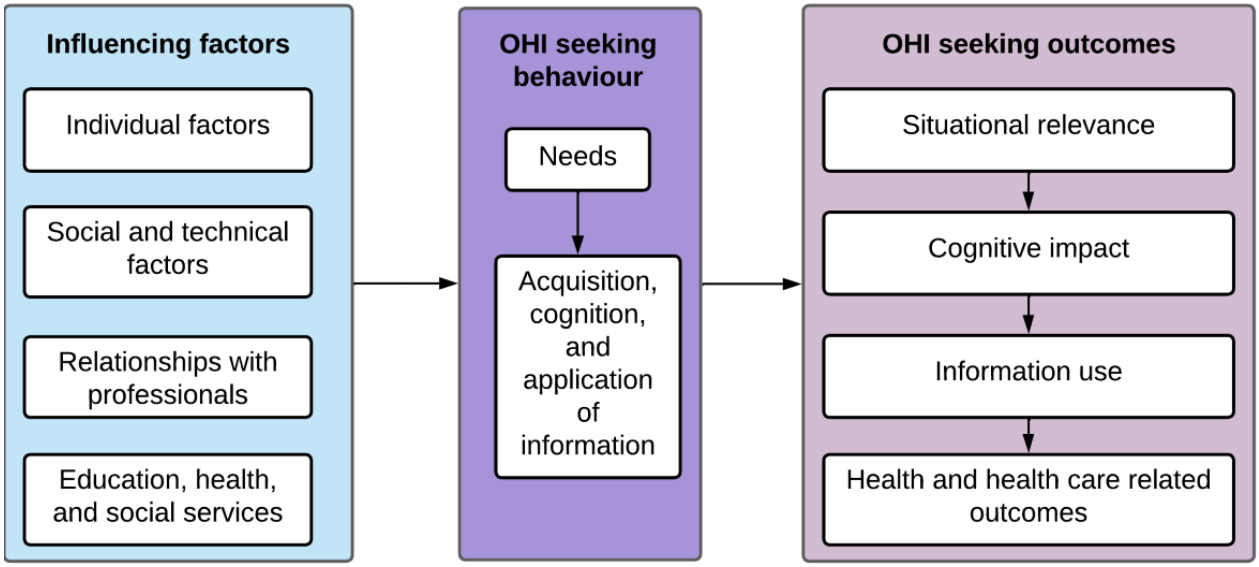
A theoretical framework on OHI outcomes and the factors associated with these outcomes was developed by Pluye and colleagues [8] based on a systematic review with a framework synthesis. This framework was derived from previous research by the authors and combines the information theory and psychosocial theory of behavior. It includes 4 types of contextual factors that influence OHI outcomes: (1) individual factors (eg, health literacy); (2) social and technical factors (eg, access to the internet); (3) relationships with professionals (eg, satisfaction with health care provider); and (4) education, health, and social services (eg, access to a family doctor). It also includes 4 levels of individual outcomes of OHI seeking: (1) situational relevance, (2) cognitive/affective impact (eg, being able to understand the information or not liking the information found), (3) use (eg, in discussions with a health care provider or to make a medical decision), and (4) subsequent health/well-being outcomes of use (eg, improved health or reduced worrying) of information.

XSL•FO
RenderX

These levels are presented in Figure 1. For each level, different types of outcomes were identified and validated using systematic mixed studies reviews and qualitative, quantitative, and mixed methods primary research studies [10,36,37].

However, this framework is focused exclusively on an individual perspective: it is the same person that starts the OHI seeking process and experiences the outcomes of this process. Studies that tested this framework therefore focused on people who used the OHI for their own health care and reported the health outcomes they themselves experienced. Little is known about what happens when the information need is to answer a question about someone else's health or what is involved when the information is used with someone else (for providing social support) [14]. Therefore, to adapt this framework to the context of proxy OHI seeking, we are interested in 4 sections of this framework: (1) influencing factors of OHI seeking, (2) OHI seeking behavior including information needs, (3) OHI use, and (4) outcomes of OHI use.

Figure 1. Online health information (OHI) outcomes conceptual framework.



Intersection of the 3 Concepts

There appears to be no comprehensive conceptual model on the outcomes of proxy OHI seekers using OHI to provide social support. Reifegerste et al [38] modified and extended the existing Comprehensive Model of Information Seeking (CMIS) with concepts of social network ties to predict proxy information seeking and the resulting social support intentions. They developed hypothetical scenarios (N=607) of people with varying severity in depression and with varying relationship closeness. Structural equation modeling was used to test the associations between the health-related factors (including demographics), proxy health information seeking intentions, and social support intentions. They hypothesized that support is the resulting action of proxy OHIS. This is an important study that modifies an existing information seeking model to proxy seeking; however, seeking and support were measured only as intentions. Moreover, the demographic characteristics were not found to be relevant, potentially due to the low variance of the study sample. Therefore, our review aims to build on this work by further exploring the context of proxy OHI seeking and the outcomes of using OHI to provide social support.

Methods

Design

A mixed studies review was conducted using a data-based convergent synthesis design in which qualitative and quantitative data were analyzed together using a qualitative thematic analysis [39,40]. A mixed studies review is ideal in this context because the evidence is from diverse fields of inquiry, and it uses diverse methods to provide a rich and highly practical understanding of complex health interventions [41,42]. Framework synthesis was then conducted to produce a revised conceptual framework.

Eligibility Criteria

Table 1 lists the inclusion and exclusion criteria that were deemed appropriate for identifying relevant studies.

XSL•FO
RenderX

Table 1. Inclusion and exclusion criteria.

	Inclusion criteria	Exclusion criteria
Research methods	Primary and secondary research (ie, qualitative, quantitative, and mixed methods empirical studies and literature reviews)	Not empirical research or a literature review (eg, commentary, editorials, reports)
OHI[a]	• Focus on online health information seeking • Online resource about health and medical topics	• No mention of OHI • Offline health information resources (eg, books or pamphlets) • Studies that tested specific online interventions (eg, testing the use of an e-kiosk or e-mental health services)
Proxy OHI seeking	Explore the phenomenon of proxy OHI seeking: • Characteristics of proxy seekers • Context of proxy OHI seeking • Use of OHI • Outcomes of OHI	• No mention of proxy OHI seeking • No mention of seekers that are physical members of the social circle that the person knows and is in contact with on a regular or semi-regular basis (eg, anonymous social media or online forum members) • Exclude parents of young children or surrogate decision-makers of incapacitated adults (eg, unconscious patients in an ICU[b])

[a]OHI: online health information.

[b]ICU: intensive care unit.

Sources and Search Strategy

Papers were searched in 5 databases (Medline, PsycInfo, CINAHL, LISA, and Scopus) from inception to May 25, 2021. A search strategy was compiled with the help of a health librarian and included 2 main concepts: OHI and proxy OHI seeking or social support. The term "surrogate seeking" was discovered after reviewing articles from the first 4 databases and was thus added to the Scopus search strategy. The sets were combined using Boolean operators depending on the database being searched, as presented in Table 2. The search was limited to English and French languages, with no limit on years. All the records were transferred to a reference manager software (EndNote x8) and duplicates were removed using the Bramer method [43]. After the selection stage, additional potentially relevant records were retrieved by tracking the citations (snowballing) of the selected documents.

Table 2. Search strategy.

Database	Date of latest search	Search terms	Records, n
Medline	May 20, 2021	*social support/ AND online.mp. AND "Health Information".af.	82
		"informational support".mp. AND online.mp. AND "Health Information".af.	14
CINAHL	May 20, 2021	"online health information" AND "social support"	16
		"online health information" AND "informational support"	5
PsycInfo	May 20, 2021	*social support/ AND online.mp. AND "Health Information".af.	141
		"informational support".mp. AND online.mp. AND "Health Information".af.	36
LISA	May 20, 2021	"proxy" AND "information seeking" AND "online health"	54
		"social support" AND "online health" AND Information	294
Scopus	May 20, 2021	"surrogate" or "proxy" AND "information seeking" AND "online health"	25
		mediator AND "online health information"	118

Selection of Relevant Studies

The 775 records were then imported into DistillerSR, a web-based application for conducting systematic reviews for selection [44]. For each record, eligibility codes were assigned according to the criteria described in Table 1. For every included record, the corresponding full-text publications were retrieved. Subsequently, full texts were imported into DistillerSR again and coded using the same eligibility criteria. Included studies were then exported into NVivo (Version 12).

Data Extraction and Synthesis of Included Studies

Characteristics of the included studies and results related to the role of social support in OHI seeking and outcomes were coded in NVivo. A deductive-inductive analytical approach was adopted for thematic analysis of the extracted evidence [45]. A coding manual was developed following the framework proposed by Pluye et al [8] that included (1) characteristics of proxy-OHI seekers, (2) context of proxy-OHI seeking, (3) use of OHI by proxy seekers, and (4) outcomes of OHI use for the seeker and recipient. The codes were then progressively clustered into major themes and subthemes.

XSL•FO
RenderX

Framework Synthesis

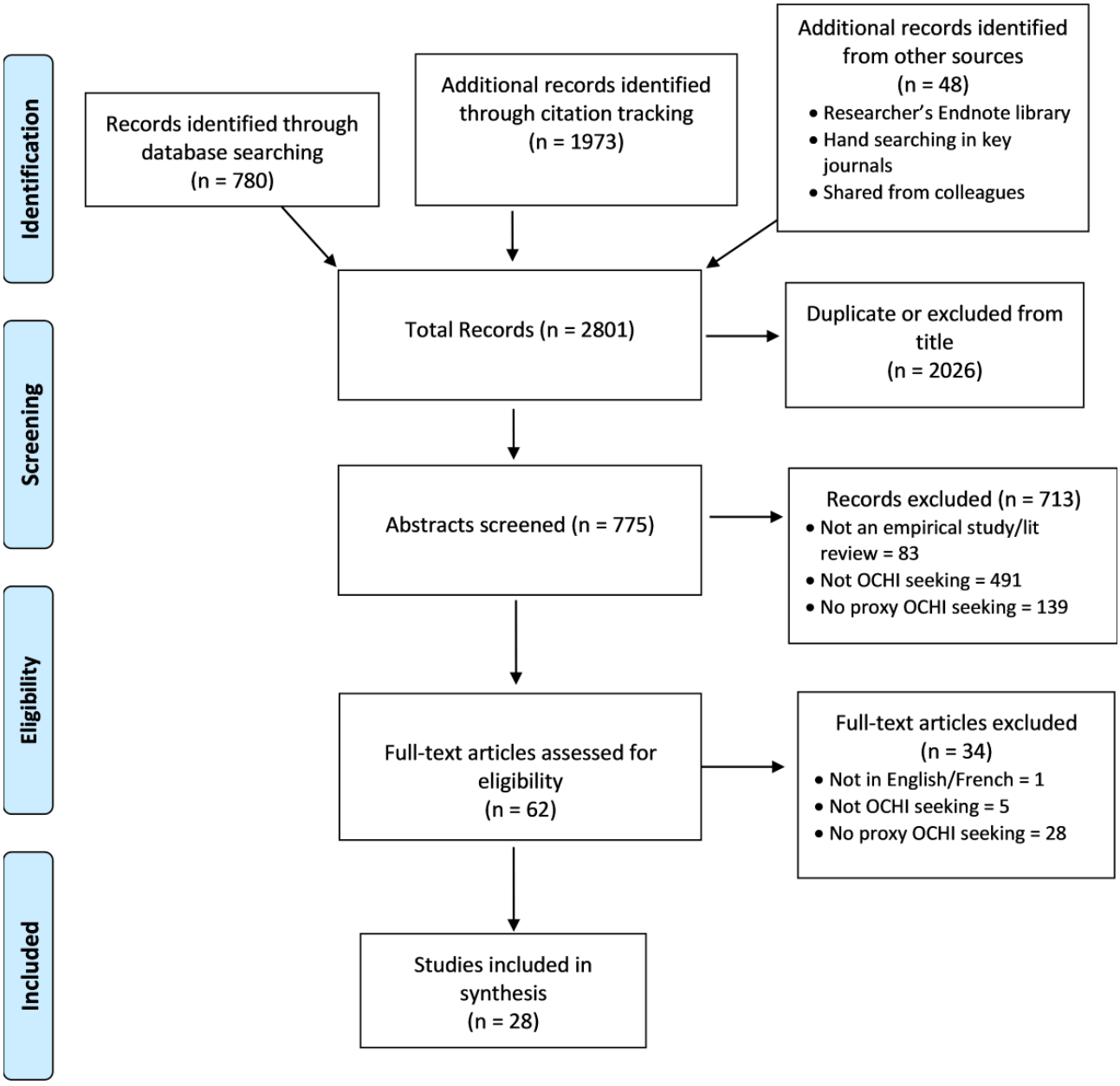
The initial framework in Figure 1 was revised following the qualitative synthesis stage. An iterative collaborative process was adopted over a series of meetings. All major themes were placed into textboxes and added to the figure representing the initial framework. Alternative figures were proposed until consensus was reached among the authors. The framework was then reviewed by 2 peer reviewers and presented at 2 research meetings (1 local and 1 international), and the feedback received was used to produce the final framework.

Results

Characteristics of Included Studies

Of 775 unique records identified in our search, 28 were deemed relevant and included in our review (Figure 2). Those referred to 15 (53.6%) quantitative studies (including 1 experimental study), 10 (35.7%) qualitative studies, 1 (3.6%) mixed methods study, and 2 (7.1%) systematic reviews. Over half (n=16, 57.1%) of the empirical studies were conducted in North America. The corresponding 28 full-text articles were divided into 3 groups depending on who the focus of the study was: OHI proxy seekers (n=9, 32.1%), OHI recipients (n=2, 7.1%), or both (n=17, 60.7%). Full details of the study characteristics are in presented in Multimedia Appendix 1.

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.



XSL•FO
RenderX

Characteristics of Proxy Seekers

The results of a telephone survey of 18,750 European citizens show that 61% of those seeking OHI searched on behalf of someone else, and of those, 26.6% exclusively searched on behalf of someone else. These surrogate OHI seekers were more likely to live with others and more likely to search on behalf of their partners, children, or other family members rather than for friends or colleagues [11]. This finding was echoed in several studies that reported that the proxy seeker was most often a member of the same household or with whom the person had close ties [12,15,46-51].

This was especially highlighted in relationships where the proxy seeker considered themselves responsible for someone else's health. We found 5 studies that focused on informal (unpaid) caregivers who reported higher and more constant proxy seeking behavior than noncaregivers [46,52,53]. A study exploring information seeking in families affected by multiple sclerosis describes the disease as a shared concern or responsibility that necessitates sharing information about it [54]. Dutta et al [55] described 3G households (3 generations of family members residing together) in Singapore, where the children and grandchildren play vital roles as sources of health information for grandparents.

Several other proxy seeker characteristics influenced OHI seeking behavior. One important factor is gender; 7 studies reported that most people who searched OHI on behalf of others were female [11,15,48,50,54,56,57]. Proxy seekers were generally younger and more educated [11,15,47,48,53,56,58] although 1 study reported that age, education, and income were not significant factors that influenced proxy OHI seeking behavior [59]. Another factor is related to the proxy seeker's experience with OHI: respondents in several of the included studies were reported as having higher health literacy [12,54] and engaging in frequent OHI seeking behavior [11].

Information Needs and Triggers of Proxy Seeking

OHI seeking was triggered by different reasons and at different times in the included studies (Table 3). The proxy seeker may be asked explicitly to search for OHI on behalf of someone who is unable to search for themselves, who has a complex health situation, or who needs to confirm information they had found online themselves [51,55,60,61]. On the other hand, more studies report that the proxy seeker initiates the search unsolicited out of interest [15,61], when they do not have enough information to support a person living with a health condition [47,54], immediately following a diagnosis [62-65], or following a visit with a health care provider [62,66]. Finally, the proxy seeker may also initiate the search themselves as a coping mechanism to help deal them with their emotions following the diagnosis of a loved one [53,61].

Table 3. Information needs and triggers of proxy seeking.

Code	Excerpt
Explicit request	"The carer may be asked to search for information on behalf of the person with cancer. This mostly occurs in situations where the patient does not have access to the internet or is not internet savvy or the person with cancer finds they are too ill to search." [61]
To make a decision	"Both patients and caregivers also mentioned that they surfed the internet again at specific moments later during the lung cancer treatment trajectory, such as during chemotherapy, at the appearance of new symptoms or disease progression, or when having to make a choice between 2 treatment options." [63]
To support someone with a health condition	"A high percentage of the 795 caregivers (87%) had used [the] internet to search for information about the disease of the patient they were taking care for in the last year prior to the survey." [47]
Out of interest or obligation	"For Gina, a 26-year-old Chinese participant, her role as a granddaughter constitutes her interpretation of HIS[a] as she mostly seeks out information for her grandparents. Jamila, a 37-year-old Malay woman, seeks out health information from the internet when one of her family members is not feeling well." [55]
Following a health care practitioner visit	"Patients and caregivers mentioned that their need to seek information often arose once they had time to rest and think about what they had been told, often at a time when their questions could not directly be answered by the treating specialist anymore: 'Once you have come home, you have forgotten half of what you have been told, which is exactly the moment you would want to ask something.'" [63]
Coping mechanism	"Carers also tended to act as 'gatekeepers' of information, and constantly sought new information as a means of coping." [53]

[a]HIS: health information seeker.

How Proxy Seekers Use OHI

Proxy seekers used OHI to better understand someone else's illness or to help themselves feel more empowered in their role as caregivers [49,64,65,67]. Several studies reported the sharing of information between caregiver and patient either directly by sending them a link or printout or indirectly by discussing the information found [49,50,57,60,64,68]. One study describes sharing and resharing the information among a social network so that it reaches a larger number of people [55] or so that a larger number of people are involved in making sense of the information [57].

One aspect of providing informational support involves acting as gatekeeper and controlling incoming information flow for the person [15]. An included literature review exploring the role of caregivers of cancer patients identified this role in 3 included studies, potentially as a way to manage the cancer experience of the patient [61]. Families developed strategies for controlling information sharing, either explicitly with the patient or

XSL•FO
RenderX

implicitly, especially if the information was potentially distressing or could lead to conflict [54,63].

Proxy seekers used the information in discussion with health care providers at a clinical visit [49,55,61,62,64]. This led to asking more questions and feeling more empowered during the visit, as well as involving the provider in the interpretation of the information [49,61,67]. In some cases, it led to requesting more testing or to trying a new treatment plan [62,69]. On the other hand, especially if the provider was not receptive to discussing the information, it also led to confronting or challenging the provider's decision [62].

Proxy seekers also used the information to provide emotional [51,62] and material support, especially as informal caregivers [46,61] to the person. They used the information to change that person's lifestyle; for example, mothers in 1 study cooked healthier food and encourage their families to walk together as a form of exercise [55]. In another study, the authors report that family members used the information to exert control on the patient, using techniques such as pushing or guilting [68].

Outcomes of OHI Use

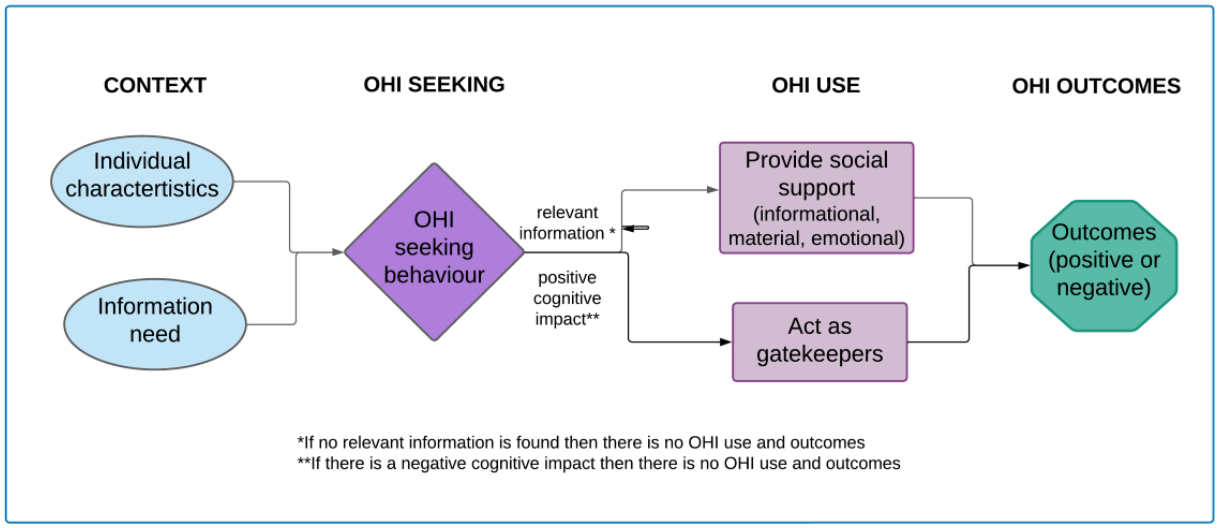
The outcomes reported by the included studies were overwhelmingly positive. Empowered by the informational they received, proxy seekers and effected individuals felt better informed and more confident, were able to discuss the information with their health care providers, and request different management options [61,62,69]. Information helped people make a health behavior change like quitting smoking [15]. It also helped lessen worries about their own health [15,66]. One study described a 87-year-old participant who reported she feels calmer when her grandchildren print out information and explain treatment options for her [55]. People described how having proxy seekers "care so much" about their health made them feel supported [51] and allowed them to have someone to talk to about their health [64].

Negative outcomes were rarely reported. A literature review found limited reports of patients' anxiety or decisions to refuse cancer treatment [61]. There were 2 studies that reported that the proxy seekers themselves experienced more anxiety, sometimes because of information overload [65,66]. The proxy seeker and the person did not always have the same approach to OHI: in situations where the person did not want to "know" or ignored the information, this led to tensions and conflict [54,68].

Revised Conceptual Framework

Figure 3 shows the revised conceptual framework after the review. The following paragraphs describe proxy seekers, their motivations, how they seek information, and their outcomes.

Figure 3. Outcomes of proxy online health information (OHI) seeking framework.



Who Proxy Seekers Are

Proxy seekers are more likely to be female and are also more likely to share health information with others, as they are considered the "central nodes" of health information within a community [70,71]. Moreover, they are more likely to be more educated, with higher eHealth literacy, and frequent internet users in general. Proxy seekers are likely to be in frequent contact with the people for whom they are seeking OHI and to report strong social ties with these people (eg, family members of the same household).

Why And When Does Proxy Seeking Occurs

The OHI seeking process is triggered by an explicit or implicit information need. Explicit information needs may be communicated to the proxy seeker with or without a request for informational support. Proxy seekers who are also informal caregivers may initiate OHI seeking as part of their caregiving responsibilities. The proxy seeker may also initiate the search themselves out of curiosity, for reassurance, or as a coping mechanism to help deal with their emotions following a diagnosis of their loved ones.

XSL•FO
RenderX

How Proxy Seekers Use Information

When proxy seekers find a situationally relevant information object that they understand or agree with (examples of positive cognitive impacts on the seeker), they can use it to provide social support for someone else. This support is most commonly informational: either by sharing the OHI found directly or discussing it with the person to help them make sense of it. Support may also be emotional or material, such as offering to cook meals. The proxy seeker also acts as an information gatekeeper by filtering the information for the person to reduce stress due to information overload.

Outcomes of OHI Use by Proxy Seekers

Using the information will lead to separate outcomes experienced by the person and the proxy seeker, which are generally positive; for example, feeling more confident discussing the information at a clinical visit. In situations where the information is conflicting or unsolicited, it may lead to negative outcomes such as increased worrying or worsening of an interpersonal relationship.

Discussion

Principal Results

To our knowledge, this is the first review to explore the outcomes of proxy OHI seeking and use of OHI to provide social support to others. We adapted a framework on individual OHI outcomes to proxy seekers and described and explained the context, use, and outcomes. Although there are 2 included reviews that reported interesting results, they did not fully address our question: the first explored the role of the internet in supporting and informing caregivers of people with cancer [61], and the second explored how informal caregivers of children with health care needs used internet-based health care services and resources [72]. Another relevant review that explored the proxy OHI seeking behavior of parents for their children and describing a conceptual model was not included in our review because parents are also proxy decision-makers for their children [73]. Another recent study adapted the existing Comprehensive Model of Information Seeking to surrogate health information seeking but did not explore the outcomes of social support [38].

Comparison to Existing Models on OHI Seeking Outcomes

In his revised 1996 model, Wilson [74] added "information processing and use." Our conceptual framework goes further and, in addition to describing the context of information seeking behavior by the proxy seeker, also explores OHI use and outcomes. Similar to the OHI outcomes framework by Pluye [8], our framework includes factors that influence information seeking behavior and leads to 4 levels of outcomes. The use of OHI in our framework revolves around types of social support, and the health and health care–related outcomes are reported by both the proxy seeker and the affected person. Moreover, we identified 2 additional consequences of informational support: sharing misleading information and acting as a gatekeeper to the information.

Our findings echo those of other studies exploring offline proxy health information seeking. In situations where the information need is explicit and the proxy seeker has high health literacy, informational support is associated with positive emotional support, and other outcomes are generally positive. First, people who can discuss the information they found with others are more likely to better understand the information, use that information to make decisions about their health care, and experience better health outcomes such as reduced worries [75-78]. Other potential outcomes include improvement in the receiver's health, buffering of potential negative outcomes, and increase in perceived social support [9,32,79]. This is especially true if the provider has higher health literacy than the receiver, in that they are better able to explain, contextualize, or validate the information [80,81]. Some people may prefer information avoidance, defined as "any behavior designed to prevent or delay the acquisition of available but potentially unwanted information" [82], which may lead to tensions between the proxy seeker and the affected person.

Second, for the seekers themselves, these outcomes include a change in their relationship with the person (improved or worsened) and feeling more involved in the health care of others [83]. Moreover, social support providers who reported feeling more satisfied with their interaction with the person and who felt better about themselves after providing informational support were more likely to continue doing so and more likely to seek information from other sources [83]. Negative outcomes for the seekers reported include increased anxiety due to information overload. This is defined as "when the information processing demands on time…exceed the supply or capacity of time available for such processing" [84].

In situations where the informational support is unsolicited and the person does not feel that the information is relevant to their situation, interpersonal tensions may develop [14]. This may also occur in relation to sharing sensitive or intimate information with family members; for example, a study examining the effects of discussing information on sexuality and contraception on mother-daughter relationships reported that a strain in the relationship may develop [85]. In addition, sharing misleading health information from unreliable sources may also lead to negative health outcomes, as described in 2 recent systematic reviews [86,87]. More specifically, in this context, the seekers do not intend to cause harm and are in fact spreading misinformation that may lead to delayed care, decreased quality of life, and increased risk of mortality.

Limitations

There are some limitations to our review. Unlike in a systematic review, only 1 reviewer carried out the selection phase, so some relevant studies may have been missed. However, our goal was to revise a framework and not necessarily to be exhaustive (in contrast to the needs of a systematic that aims to measure effectiveness of an intervention). Similar to other reviews, there may have been underreporting of negative outcomes due to publication bias. Finally, systematically reviewing all the models on information seeking behavior was beyond the scope of this review, but we reviewed and discussed the most common models with a specialized expert librarian.

XSL•FO
RenderX

Directions For Future Research

Most studies on information seeking behavior do not explore how the information is used by proxy seekers, and what happens next [88]. While this review explores the outcomes of OHI proxy seeking, few studies report outcomes for the seekers themselves. As such, future empirical studies can focus on these outcomes from the seekers' perspectives. Furthermore, little is known about which contextual factors or seeker characteristics are associated with positive and negative OHI outcomes. Future studies can test our framework in different contexts, revise it, and propose research-based solutions to help the proxy seekers use OHI with others.

Conclusion

The outcomes of proxy OHI seeking constitute an important topic for both information specialists and health care practitioners. Members of a person's social circle may help them overcome information-seeking barriers and illness challenges (eg, when they are too physically weak or mentally incapacitated to search themselves) [15]. People are sometimes more likely to turn to members of their social circle to make sense of OHI they find rather than discuss it with a health care professional [11]. By better understanding how affected people and their social circle use OHI together, OHI providers can better adapt their platforms and information to meet both their needs, and health care practitioners can target patients' social circles with information for dissemination and use [16]. Potential public health intervention strategies can focus improving proxy OHI seeking and OHI use to promote positive outcomes for proxy seekers and the people they seek for through strategies that help proxy OHI seekers find relevant OHI, evaluate it, and use it appropriately. Strategies can also include extending social support networks for people without an effective social circle by identifying social support interventions from previous work that may be applicable in the context of proxy OHI seeking.

Acknowledgments

The authors gratefully acknowledge Vera Granikov and Quan Nha Hong for their help in reviewing this manuscript and providing constructive feedback. RES was granted a doctoral award from the Canadian Institute of Health Research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of Included Studies.
[DOCX File , 42 KB - jmir_v24i6e34345_app1.docx]

References

1. Finney Rutten LJ, Blake KD, Greenberg-Worisek AJ, Allen SV, Moser RP, Hesse BW. Online Health Information Seeking Among US Adults: Measuring Progress Toward a Healthy People 2020 Objective. Public Health Rep 2019;134(6):617-625 [FREE Full text] [doi: 10.1177/0033354919874074] [Medline: 31513756]
2. Table 22-10-0137-01 Selected online activities by gender, age group and highest certificate, diploma or degree completed. Statistics Canada. 2020. URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=2210013701 [accessed 2021-05-31]
3. ICT Access and Usage by Households and Individuals database on the Internet. OECD Telecommunications and Internet Statistics. 2017. URL: https://www.oecd-ilibrary.org/content/data/b9823565-en [accessed 2021-05-31]
4. Amante DJ, Hogan TP, Pagoto SL, English TM, Lapane KL. Access to care and use of the Internet to search for health information: results from the US National Health Interview Survey. J Med Internet Res 2015 Apr 29;17(4):e106 [FREE Full text] [doi: 10.2196/jmir.4126] [Medline: 25925943]
5. Prescott J, Mackie L. "You sort of go down a rabbit hole...you're just going to keep on searching": a qualitative study of searching online for pregnancy-related information during pregnancy. J Med Internet Res 2017 Jun 05;19(6):e194 [FREE Full text] [doi: 10.2196/jmir.6302]
6. Case D, Given L. In: Mai JE, editor. Looking For Information: A Survey of Research on Information Seeking, Needs, and Behavior 4th ed. London, UK: Emerald Group Publishing; Dec 21, 2016:2284-2286.
7. Pluye P, El Sherif R, Gonzalez-Reyes A, Turcotte E, Schuster T, Bartlett G, et al. Outcomes of equity-oriented, web-based parenting information in mothers of low socioeconomic status compared to other mothers: participatory mixed methods study. J Med Internet Res 2020 Nov 10;22(11):e22440 [FREE Full text] [doi: 10.2196/22440] [Medline: 33170125]
8. Pluye P, El Sherif R, Granikov V, Hong QN, Vedel I, Galvao MCB, et al. Health outcomes of online consumer health information: A systematic mixed studies review with framework synthesis. J Assoc Inf Sci Technol 2019 Jul 30;70(7):643-659. [doi: 10.1002/asi.24178] [Medline: 31423458]
9. Lin N, Simeone RS, Ensel WM, Kuo W. Social support, stressful life events, and illness: a model and an empirical test. J Health Soc Behav 1979 Jun;20(2):108-119. [Medline: 479524]

XSL•FO
RenderX

10. Pluye P, Grad R, Repchinsky C, Jovaisas B, Johnson-Lafleur J, Carrier M, et al. Four levels of outcomes of information-seeking: A mixed methods study in primary health care. J Am Soc Inf Sci Tec 2012 Dec 10;64(1):108-125 [FREE Full text] [doi: 10.1002/asi.22793]
11. Reifegerste D, Bachl M, Baumann E. Surrogate health information seeking in Europe: influence of source type and social network variables. Int J Med Inform 2017 Jul 01:7-14 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.04.006]
12. Cutrona SL, Mazor KM, Vieux SN, Luger TM, Volkman JE, Finney Rutten LJ. Health information-seeking on behalf of others: characteristics of "surrogate seekers". J Cancer Educ 2015 Mar;30(1):12-19 [FREE Full text] [doi: 10.1007/s13187-014-0701-3] [Medline: 24989816]
13. Massey PM. Where do U.S. adults who do not use the internet get health information? examining digital health information disparities from 2008 to 2013. J Health Commun 2016 Nov 23;21(1):118-124. [doi: 10.1080/10810730.2015.1058444] [Medline: 26166484]
14. El Sherif R, Pluye P, Thoër C, Rodriguez C. Reducing negative outcomes of online consumer health information: qualitative interpretive study with clinicians, librarians, and consumers. J Med Internet Res 2018 May 04;20(5):e169 [FREE Full text] [doi: 10.2196/jmir.9326] [Medline: 29728350]
15. Abrahamson JA, Fisher KE, Turner AG, Durrance JC, Turner TC. Lay information mediary behavior uncovered: exploring how nonprofessionals seek health information for themselves and others online. J Med Libr Assoc 2008 Oct;96(4):310-323 [FREE Full text] [doi: 10.3163/1536-5050.96.4.006] [Medline: 18974809]
16. Kim W, Kreps GL, Shin C. The role of social support and social networks in health information-seeking behavior among Korean Americans: a qualitative study. Int J Equity Health 2015 Apr 28;14(1):40 [FREE Full text] [doi: 10.1186/s12939-015-0169-8] [Medline: 25927546]
17. Bates M. Toward an integrated model of information seeking and searching. In: New Review of Information Behaviour Research. 2002 Presented at: Fourth international Conference on Information Needs, Seeking and Use in Different Contexts; September 11, 2002; Lisbon, Portugal p. 1-15.
18. Wilson T. Models in information behaviour research. J Doc 1999 Aug;55(3):249-270. [doi: 10.1108/EUM0000000007145]
19. Wyatt S, Henwood F, Hart A, Smith J. The digital divide, health information and everyday life. New Media Soc 2016 Jun 30;7(2):199-218. [doi: 10.1177/1461444805050747]
20. Wathen N, Wyatt S, Harris R. Mediating Health Information: The Go-Betweens in a Changing Socio-Technical Landscape. London, UK: Palgrave Macmillan; 2008.
21. Wilson TD. On user studies and information needs. J Doc 1981 Jan;37(1):3-15. [doi: 10.1108/eb026702]
22. McKenzie P. A model of information practices in accounts of everyday life information seeking. J Doc 2003;59:19-40. [doi: 10.1108/00220410310457993]
23. Umberson D, Montez JK. Social relationships and health: a flashpoint for health policy. J Health Soc Behav 2010 Oct 08;51 Suppl(1_suppl):S54-S66. [doi: 10.1177/0022146510383501] [Medline: 20943583]
24. Uchino B. Social Support and Physical Health: Understanding the Health Consequences of Relationships. London, UK: Yale University Press; 2004:0300127987.
25. Dubois S, Loiselle C. Cancer informational support and health care service use among individuals newly diagnosed: a mixed methods approach. J Eval Clin Pract 2009 Apr;15(2):346-359. [doi: 10.1111/j.1365-2753.2008.01013.x] [Medline: 19335496]
26. Loiselle CG, Lambert SD, Dubois S. Beyond the mere dichotomy of active search versus avoidance of information about the self. J Med Libr Assoc 2006 Oct;94(4):375. [Medline: 17082826]
27. McKinley C, Wright P. Informational social support and online health information seeking: Examining the association between factors contributing to healthy eating behavior. Comput Hum Behav 2014 Aug;37:107-116 [FREE Full text] [doi: 10.1016/j.chb.2014.04.023]
28. Drentea P, Moren-Cross JL. Social capital and social support on the web: the case of an internet mother site. Sociol Health Illn 2005 Nov;27(7):920-943 [FREE Full text] [doi: 10.1111/j.1467-9566.2005.00464.x] [Medline: 16313523]
29. House JS. Social isolation kills, but how and why? Psychosom Med 2001;63(2):273-274. [doi: 10.1097/00006842-200103000-00011] [Medline: 11292275]
30. Oprescu F, Campo S, Lowe J, Andsager J, Morcuende JA. Managing uncertainty in the context of clubfoot care: exploring the value of uncertainty management theory and the sense of virtual community. Iowa Orthop J 2013;33:142-148. [Medline: 24027474]
31. Hamlett KW, Pellegrini DS, Katz KS. Childhood chronic illness as a family stressor. J Pediatr Psychol 1992 Feb;17(1):33-47. [doi: 10.1093/jpepsy/17.1.33] [Medline: 1545320]
32. Dunst C, Trivette C, Cross A. Mediating influences of social support: personal, family, and child outcomes. Am J Ment Defic 1986 Jan;90(4):403-417. [Medline: 2418680]
33. Kiecolt-Glaser JK, Fisher LD, Ogrocki P, Stout JC, Speicher CE, Glaser R. Marital quality, marital disruption, and immune function. Psychosom Med 1987;49(1):13-34. [doi: 10.1097/00006842-198701000-00002] [Medline: 3029796]
34. Cutrona CE, Suhr JA. Controllability of stressful events and satisfaction with spouse support behaviors. Commun Res 2016 Jun 30;19(2):154-174. [doi: 10.1177/009365092019002002]

XSL•FO
RenderX

35. Guillory J, Niederdeppe J, Kim H, Pollak JP, Graham M, Olson C, et al. Does social support predict pregnant mothers' information seeking behaviors on an educational website? Matern Child Health J 2014 Nov 27;18(9):2218-2225. [doi: 10.1007/s10995-014-1471-6] [Medline: 24671467]
36. Pluye P, Granikov V, Bartlett G, Grad RM, Tang DL, Johnson-Lafleur J, et al. Development and content validation of the information assessment method for patients and consumers. JMIR Res Protoc 2014 Feb 18;3(1):e7 [FREE Full text] [doi: 10.2196/resprot.2908] [Medline: 24550180]
37. Bujold M, El Sherif R, Bush PL, Johnson-Lafleur J, Doray G, Pluye P. Ecological content validation of the Information Assessment Method for parents (IAM-parent): A mixed methods study. Eval Program Plann 2018 Feb;66:79-88 [FREE Full text] [doi: 10.1016/j.evalprogplan.2017.09.011] [Medline: 29053984]
38. Reifegerste D, Blech S, Dechant P. Understanding information seeking about the health of others: applying the comprehensive model of information seeking to proxy online health information seeking. J Health Commun 2020 Feb 01;25(2):126-135. [doi: 10.1080/10810730.2020.1716280] [Medline: 32009552]
39. Pluye P, Hong QN, Bush PL, Vedel I. Opening-up the definition of systematic literature review: the plurality of worldviews, methodologies and methods for reviews and syntheses. J Clin Epidemiol 2016 May;73:2-5. [doi: 10.1016/j.jclinepi.2015.08.033] [Medline: 26898706]
40. Hong QN, Pluye P, Bujold M, Wassef M. Convergent and sequential synthesis designs: implications for conducting and reporting systematic reviews of qualitative and quantitative evidence. Syst Rev 2017 Mar 23;6(1):61 [FREE Full text] [doi: 10.1186/s13643-017-0454-2] [Medline: 28335799]
41. Pluye P, Hong QN. Combining the power of stories and the power of numbers: mixed methods research and mixed studies reviews. Annu Rev Public Health 2014;35:29-45. [doi: 10.1146/annurev-publhealth-032013-182440] [Medline: 24188053]
42. Grant M, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Info Libr J 2009 Jun;26(2):91-108 [FREE Full text] [doi: 10.1111/j.1471-1842.2009.00848.x] [Medline: 19490148]
43. Bramer WM, Giustini D, de Jonge GB, Holland L, Bekhuis T. De-duplication of database search results for systematic reviews in EndNote. J Med Libr Assoc 2016 Jul;104(3):240-243 [FREE Full text] [doi: 10.3163/1536-5050.104.3.014] [Medline: 27366130]
44. DistillerSR.Version 2.35.: Evidence Partners; 2021. URL: https://www.evidencepartners.com/ [accessed 2021-02-28]
45. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. Int J Qual Methods 2016 Nov 29;5(1):80-92. [doi: 10.1177/160940690600500107]
46. Sadasivam RS, Kinney RL, Lemon SC, Shimada SL, Allison JJ, Houston TK. Internet health information seeking is a team sport: analysis of the Pew Internet Survey. Int J Med Inform 2013 Mar;82(3):193-200. [doi: 10.1016/j.ijmedinf.2012.09.008] [Medline: 23149121]
47. Chua GP, Ng QS, Tan HK, Ong WS. Caregivers of cancer patients: what are their information-seeking behaviours and resource preferences? Ecancermedicalscience 2020 Jul 10;14:1068. [doi: 10.3332/ecancer.2020.1068] [Medline: 32728384]
48. Cutrona SL, Mazor KM, Agunwamba AA, Valluri S, Wilson PM, Sadasivam RS, et al. Health information brokers in the general population: an analysis of the Health Information National Trends Survey 2013-2014. J Med Internet Res 2016 Jun 03;18(6):e123 [FREE Full text] [doi: 10.2196/jmir.5447] [Medline: 27260952]
49. Kirschning S, von Kardorff E, Merai K. Internet use by the families of cancer patients—help for disease management? J Public Health 2006 Oct 3;15(1):23-28. [doi: 10.1007/s10389-006-0070-4]
50. Nicholas D, Huntington P, Gunter B, Russell C, Withey R. The British and their use of the web for health information and advice: a survey. Aslib Proceedings: New Information Perspectives 2003;55(5/6):261-276.
51. Zhao YC, Zhao M, Song S. Online health information seeking behaviors among older adults: systematic scoping review. J Med Internet Res 2022 Feb 16;24(2):e34790 [FREE Full text] [doi: 10.2196/34790] [Medline: 35171099]
52. Bangerter LR, Griffin J, Harden K, Rutten LJ. Health information-seeking behaviors of family caregivers: analysis of the Health Information National Trends Survey. JMIR Aging 2019 Jan 14;2(1):e11237 [FREE Full text] [doi: 10.2196/11237] [Medline: 31518309]
53. James N, Daniels H, Rahman R, McConkey C, Derry J, Young A. A study of information seeking by cancer patients and their carers. Clin Oncol (R Coll Radiol) 2007 Jun;19(5):356-362. [doi: 10.1016/j.clon.2007.02.005] [Medline: 17399963]
54. Mazanderani F, Hughes N, Hardy C, Sillence E, Powell J. Health information work and the enactment of care in couples and families affected by multiple sclerosis. Sociol Health Illn 2019 Feb 24;41(2):395-410. [doi: 10.1111/1467-9566.12842] [Medline: 30677163]
55. Dutta MJ, Kaur S, Luk P, Lin J, Lee ST. Health information seeking among Singaporeans: roles and collective contexts. Health Commun 2018 Apr 02;33(4):433-442. [doi: 10.1080/10410236.2016.1278493] [Medline: 28151015]
56. Oh YS. Predictors of self and surrogate online health information seeking in family caregivers to cancer survivors. Soc Work Health Care 2015 Dec;54(10):939-953. [doi: 10.1080/00981389.2015.1070780] [Medline: 26671245]
57. Turner A, Osterhage K, Taylor J, Hartzler A, Demiris G. A closer look at health information seeking by older adults and involved family and friends: design considerations for health information technologies. AMIA Annu Symp Proc 2018;2018:1036-1045 [FREE Full text] [Medline: 30815147]
58. Li H. Informal caregivers' use of the internet for caregiving information. Soc Work Health Care 2015 Jul 17;54(6):532-546. [doi: 10.1080/00981389.2015.1045577] [Medline: 26186424]

XSL•FO
RenderX

59. Cutrona SL, Mazor KM, Vieux SN, Luger TM, Volkman JE, Finney RLJ. Health information-seeking on behalf of others: characteristics of "surrogate seekers". J Cancer Educ 2015 Mar;30(1):12-19 [FREE Full text] [doi: 10.1007/s13187-014-0701-3] [Medline: 24989816]
60. Carpenter DM, Elstad EA, Sage AJ, Geryk LL, DeVellis RF, Blalock SJ. The relationship between partner information-seeking, information-sharing, and patient medication adherence. Patient Educ Couns 2015 Jan;98(1):120-124. [doi: 10.1016/j.pec.2014.10.001] [Medline: 25455797]
61. Kinnane NA, Milne DJ. The role of the Internet in supporting and informing carers of people with cancer: a literature review. Support Care Cancer 2010 Sep;18(9):1123-1136. [doi: 10.1007/s00520-010-0863-4] [Medline: 20336326]
62. Dolce MC. The internet as a source of health information: experiences of cancer survivors and caregivers with healthcare providers. Oncol Nurs Forum 2011 May;38(3):353-359. [doi: 10.1188/11.ONF.353-359] [Medline: 21531685]
63. Schook RM, Linssen C, Schramel FM, Festen J, Lammers E, Smit EF, et al. Why do patients and caregivers seek answers from the internet and online lung specialists? A qualitative study. J Med Internet Res 2014 Feb;16(2):e37 [FREE Full text] [doi: 10.2196/jmir.2842] [Medline: 24496139]
64. Simon C, Schramm S. Cancer and the computerized family: towards a clinical ethics of "indirect" internet use. Med Health Care Philos 2008 Sep 19;11(3):337-341. [doi: 10.1007/s11019-008-9127-1] [Medline: 18283559]
65. Coder M. Information needs and information seeking by family members and friends of terminal cancer patients: an exploratory study. J Hosp Librariansh 2020 Feb 04;20(1):1-26. [doi: 10.1080/15323269.2020.1702838]
66. Bouju P, Tadié JM, Uhel F, Letheulle J, Fillatre P, Lavoué S, et al. Internet use by family members of intensive care unit patients: a pilot study. Intensive Care Med 2014 Aug 28;40(8):1175-1176. [doi: 10.1007/s00134-014-3371-z] [Medline: 24972885]
67. Tonsaker T, Law S, Ormel I, Nease C, Bartlett G. Engaging caregivers: exploring perspectives on web-based health information. Fam Pract 2017 Aug 01;34(4):479-484. [doi: 10.1093/fampra/cmw084] [Medline: 27543794]
68. Brown LK, Veinot TC. Information behavior and social control: toward an understanding of conflictual information behavior in families managing chronic illness. J Assoc Inf Sci Technol 2020 Apr 28;72(1):66-82. [doi: 10.1002/asi.24362]
69. Coffey NT, Cassese J, Cai X, Garfinkel S, Patel D, Jones R, et al. Identifying and understanding the health information experiences and preferences of caregivers of individuals with either traumatic brain injury, spinal cord injury, or burn injury: a qualitative investigation. J Med Internet Res 2017 May 10;19(5):e159 [FREE Full text] [doi: 10.2196/jmir.7027] [Medline: 28490418]
70. Altizer KP, Grzywacz JG, Quandt SA, Bell R, Arcury TA. A qualitative analysis of how elders seek and disseminate health information. Gerontol Geriatr Educ 2014 Nov 22;35(4):337-353. [doi: 10.1080/02701960.2013.844693] [Medline: 24188253]
71. Colon-Ramos U, Atienza AA, Weber D, Taylor M, Uy C, Yaroch A. Practicing what they preach: health behaviors of those who provide health advice to extensive social networks. J Health Commun 2009 Mar 16;14(2):119-130. [doi: 10.1080/10810730802659111] [Medline: 19283537]
72. Park E, Kim H, Steinhoff A. Health-related internet use by informal caregivers of children and adolescents: an integrative literature review. J Med Internet Res 2016 Mar 03;18(3):e57 [FREE Full text] [doi: 10.2196/jmir.4124] [Medline: 26940750]
73. Kubb C, Foran HM. Online health information seeking by parents for their children: systematic review and agenda for further research. J Med Internet Res 2020 Aug 25;22(8):e19985 [FREE Full text] [doi: 10.2196/19985] [Medline: 32840484]
74. Wilson T. Information behaviour: An interdisciplinary perspective. Inf Process Manag 1997 Jul;33(4):551-572. [doi: 10.1016/s0306-4573(97)00028-9]
75. Tanis M, Hartmann T, Te Poel F. Online health anxiety and consultation satisfaction: A quantitative exploratory study on their relations. Patient Educ Couns 2016 Jul;99(7):1227-1232. [doi: 10.1016/j.pec.2016.01.021] [Medline: 26873545]
76. 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]
77. Iverson S, Howard K, Penney B. Impact of internet use on health-related behaviors and the patient-physician relationship: a survey-based study and review. J Am Osteopath Assoc 2008;108(12):699-711.
78. Thapa DK, Visentin DC, Kornhaber R, West S, Cleary M. The influence of online health information on health decisions: A systematic review. Patient Educ Couns 2021 Apr;104(4):770-784. [doi: 10.1016/j.pec.2020.11.016] [Medline: 33358253]
79. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. Psychol Bull 1985 Sep;98(2):310-357. [Medline: 3901065]
80. Gerber BS, Eiser AR. The patient physician relationship in the Internet age: future prospects and the research agenda. J Med Internet Res 2001;3(2):E15 [FREE Full text] [doi: 10.2196/jmir.3.2.e15] [Medline: 11720957]
81. Fox S, Duggan M. Health online 2013. Pew Research Center. 2013. URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2021-05-31]
82. Sweeny K, Melnyk D, Miller W, Shepperd JA. Information avoidance: who, what, when, and why. Rev Gen Psychol 2010 Dec 01;14(4):340-353. [doi: 10.1037/a0021288]
83. Hether HJ, Murphy ST, Valente TW. It's better to give than to receive: the role of social support, trust, and participation on health-related social networking sites. J Health Commun 2014 Dec 25;19(12):1424-1439. [doi: 10.1080/10810730.2014.894596] [Medline: 24766297]

XSL•FO
RenderX

84. Schick A, Gordon L, Haka S. Information overload: A temporal approach. Accounting, Organizations and Society 1990 Jan;15(3):199-220. [doi: 10.1016/0361-3682(90)90005-F]
85. Amsellem-Mainguy Y. Prescrire et proscrire des conduites, véhiculer des normes: les mères comme actrices privilégiées de prévention en matière de sexualité et de contraception. Recherches familiales 2006;3(1):49-59 [FREE Full text] [doi: 10.3917/rf.003.0049]
86. 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]
87. Swire-Thompson B, Lazer D. Public health and online misinformation: challenges and recommendations. Annu Rev Public Health 2020 Apr 02;41(1):433-451. [doi: 10.1146/annurev-publhealth-040119-094127] [Medline: 31874069]
88. Case DO, O'Connor LG. What's the use? Measuring the frequency of studies of information outcomes. J Assn Inf Sci Tec 2015 Jan 30;67(3):649-661. [doi: 10.1002/asi.23411]

Abbreviations

CMIS: Comprehensive Model of Information Seeking
OECD: Organisation for Economic Cooperation and Development
OHI: online health information

Edited by R Kukafka; submitted 18.10.21; peer-reviewed by D Reifegerste, H Kim; comments to author 30.11.21; revised version received 18.02.22; accepted 21.03.22; published 24.06.22.

Please cite as:
El Sherif R, Pluye P, Ibekwe F
Contexts and Outcomes of Proxy Online Health Information Seeking: Mixed Studies Review With Framework Synthesis
J Med Internet Res 2022;24(6):e34345
URL: https://www.jmir.org/2022/6/e34345
doi:10.2196/34345
PMID:35749210

©Reem El Sherif, Pierre Pluye, Fidelia Ibekwe. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.06.2022. 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

The Syndemic of Inequity and COVID-19 in Virtual Care

Esha Ray Chaudhuri[1,2], MA, MEd, PhD

[1]Calgary, AB, Canada
[2]See Acknowledgments

Corresponding Author:
Esha Ray Chaudhuri, MA, MEd, PhD
Calgary, AB
Canada
Phone: 1 5878903913
Email: ER2Consult@gmail.com

Related Article:

Comment on: http://jmir.org/2022/1/e29876/

Abstract

The critical intersections of structural inequities and vulnerabilities of marginalized populations, particularly those engaging the social gradient of minority ethnic communities, are revealed in the syndemic approach to COVID-19. Although proposals for cultural interventions to improve virtual care provide relevant measures, they may not address the root cause of the disparate impacts of a pandemic on population subgroups. The common misperception of equality as synonymous with equity further impedes the efficacy of digital health in quality-of-care initiatives, as it systemically fails to acknowledge the disparate realities of marginalized populations, while intending to benefit all. This commentary suggests that an alignment of the health care system with Canada's pluralist principles would support a paradigm shift in transforming virtual care into an equitable standard as envisioned by Pham and colleagues in their paper, "The Future of Virtual Care for Older Ethnic Adults Beyond the COVID-19 Pandemic."

(J Med Internet Res 2022;24(6):e37717) doi:10.2196/37717

KEYWORDS

COVID-19; virtual care; health equity; syndemic; aging; ethnicity; older adults; equity; digital health; diversity; ethnic minority groups; minority groups; older ethnic adults

Introduction

In a fundamental sense, the vision for transforming virtual care from that of an exclusive service that benefits only a few to that of a standard for providing equitable care for all [1] echoes the age-old debate between policy variations on the *zip code* and the *genetic code*. This commentary aims to further develop the key theme discussed by Pham et al [1]—engaging the "reimagining" of virtual care for older ethnic adults—by considering the syndemic nature of COVID-19 and the intersection of cultural interventions in care and equity in virtual care.

The Context of the COVID-19 Syndemic

Well before the onset of COVID-19 and other major pandemics in the past 2 decades, studies on the design and evaluation of eHealth interventions recognized the challenges and implications of the interdisciplinary nature of the field [2].

The emergency management of health care in the pandemic era inadvertently proved the critical role of the social determinants of health through data about the rapid viral spread in largely marginalized, resource-challenged communities [3]. Based on the presence of similar contexts like the digital divide along pathways of structural inequities, COVID-19 has been characterized as a syndemic as opposed to a pandemic [4], emphasizing the intersections of the contributing demographic, social, economic, and environmental factors of the pandemic.

A summary review of the literature suggests that despite a significant number of health care and digital health reform initiatives, which address the disparities experienced by marginalized ethnic communities [5], few have addressed the need for a systemic transformation based on equity. Even the unique perspectives of reorienting the health care culture to its original *benevolent* foundation appear to sustain in principle the context of ethnic minority populations' vulnerabilities [6].

XSL•FO
RenderX

Syndemic theory advances the examination of health and health care disparities while emphasizing the contexts of social and economic systems in these processes. As such, the theory provides a critical alternative to conventional systemic reform culture. It recognizes how disparities in social realities are accountable for not only *shaping* the marginal experience but also for its *distribution* across subgroups of populations.

Aging, Ethnicity, and the Equity Paradox

In the current design, implementation, and evaluation of virtual care, the ethnicity context represents one of the several dimensions of equity, such as aging, gender, etc. To evaluate the impact of equity in digital health, it is essential that the determinants be addressed within the synergistic lens of intersectionality; the interface of the factors with structural inequities of exclusive policies presents an additional dimension.

With the increasing diversity of the population, the need to recognize equity becomes imperative. The perception of ethnic diversity as a "strength" and "asset" as found in different sectors of social planning, such as business, industry, or labor, provides an important contrast to the typical safety-net approaches to vulnerable populations of ethnic minorities in health equity and digital health studies [7].

Within studies about the intersections of ethnicity, aging, and equity [5], inclusion of ethnic older adults enhances the generic minority data measured by metrics that assume homogeneity of vulnerabilities. The ethnically nuanced care expectations of older adults and, more importantly, the cultural values that frame the expectations, are seldom contextually related to any specific equity dimension.

Pham et al [1] offer a good example of leveraging the strength of ethnic diversity to enhance quality of care. The paper presents important insights about unique cultural elements of filial piety and kinship values prevalent in Asian families. The distinct cultural norms explain the common practice adopted by family members, including adult children, who often volunteer to care for their aged parents even at the cost of sacrificing their professional careers. The study proposes a formal care partner role for family members to help older adults navigate the digital health system. The observation not only advances the potential of catalyzing diversity as a "strength" for quality care but also identifies normative variations, which are seldom acknowledged in people-centered care policies. Within the intersectionalities of ethnicity, aging, and quality care principles, the intervention model provides a compelling argument for the segregation of data for "older ethnic adults"; it further reflects the need for digital determinants to distinguish between assumptions of typical "safety-net" traits of (minority) ethnic adults and their actual role as unique partners in strengthening the scope, scale, and equity of the health system and digital health.

Despite a noticeable increase in the acknowledgment of ethnically diverse data for the effective diagnosis of disease profiles and trajectories, ethnic patients remain the "subject" of studies rather than their architects. In the general discourse of health disciplines, data on ethnic minority groups are routinely aggregated with *vulnerabilities* and *marginalization concerns.* Yet, research about the impact of equity in the transformation of health systems, and more critically understanding the role of intersector approaches to challenges and opportunities of equity, would allow digital health to become more inclusive and sustainable.

Quoting science philosopher Thomas Kuhn, Meskó and colleagues [8] describe inequities in health as *anomalies* within the traditional paradigm of health that cannot be addressed without a shift in the structure of the system. This misalignment of the framework and the vision is described as a *paradox.* The conventional insular norms of health care systems act as barriers to equity, disregarding pluralism as deviation. The design of virtual care is susceptible to translating systemic inequities that may be embedded in existing models of health care. Crucial reform initiatives of cultural interventions [1] or designing methodical frameworks for equity analysis in digital health [9] are promising approaches for improving the equality aspects of health care, even if within the traditional systemic norms. To the extent the initiatives align with the culture-specific norms of the conventional system, targeted reform initiatives present good alternatives for improving the efficacy of care. Yet, in the absence of intentional transformative approaches, cultural variations in normative principles of health equity and digital health would continue to be interpreted as part of generic data, *measured by metrics of seemingly homogeneous vulnerabilities.* The unmet care needs and, more importantly, the cultural values that frame these needs—often acknowledged as proverbial cracks and gaps in quality of care—remain unrecognized as upstream factors and are seldom identified as a rationale for transformation.

Equity Issues in Virtual Care

The concept of digital health equity is complex and multidimensional. It integrates a comprehensive consideration of individual contexts, the social determinants of health, and the enabling environment [10,11].

In the case of virtual care, the *complexity* increases with the introduction of digital determinants into the equation. The synoptic review of various models for health equity by Shaw et al [5] provides a glimpse of this complexity in their discussion of the levels of the digital divide, where individuals in the final level, who have access to technology and possess digital literacy, in addition to having competencies in digital navigation, are still not always able to achieve quality outcomes.

As the syndemic approach illustrates, epidemiologic assumptions about health equity generally address clinical-level efficiencies in the care quality of vulnerable ethnic minority groups, and seldom introduce the social lens into the equation. Inequities of digital health originate when the conventional approach gets instinctively coded into algorithms of digital technology, despite its innovative performance in various fields of medicine [11].

As long as the foundation of virtual care—the traditional health care system—remains unaware of its systemic cultural bias, innovative digital technologies and tools will mirror inequities. Virtual care represents a unique medium of care service delivery and, as such, it can effectively design technical solutions for

XSL•FO
RenderX

access issues for all Canadians, including ethnic older adults and minority language–speaking patients, through the creation of appropriate user-friendly platforms for overcoming barriers to participation in digital health. Yet, substantive accommodation of ethnicity in contemporary discourse on cultural equity in health and digital health designs requires a shift in paradigm in the content of health care policy principles and its strategic priorities and action imperatives, which should resonate with the values of *all* patients and consumers *as aligned with the principles of pluralism*. Nothing less will help to achieve the purpose described by Pham et al [1] of catalyzing the transition from an exclusive service to an equitable standard.

Acknowledgments

ERC is is an Independent Equity Analyst in Calgary, Alberta, Canada.

Conflicts of Interest

None declared.

References

1. Pham Q, El-Dassouki N, Lohani R, Jebanesan A, Young K. The future of virtual care for older ethnic adults beyond the COVID-19 pandemic. J Med Internet Res 2022 Jan 07;24(1):e29876 [FREE Full text] [doi: 10.2196/29876] [Medline: 34994707]
2. Pagliari C. Design and evaluation in eHealth: challenges and implications for an interdisciplinary field. J Med Internet Res 2007 May 27;9(2):e15 [FREE Full text] [doi: 10.2196/jmir.9.2.e15] [Medline: 17537718]
3. Verma A, Towfighi A, Brown A, Abhat A, Casillas A. Moving towards equity with digital health innovations for stroke care. Stroke 2022 Mar;53(3):689-697. [doi: 10.1161/STROKEAHA.121.035307] [Medline: 35124973]
4. Horton R. Offline: COVID-19 is not a pandemic. The Lancet 2020 Sep 26;396(10255):874 [FREE Full text] [doi: 10.1016/S0140-6736(20)32000-6] [Medline: 32979964]
5. Shaw J, Brewer LC, Veinot T. Recommendations for health equity and virtual care arising from the COVID-19 pandemic: Narrative review. JMIR Form Res 2021 Apr 05;5(4):e23233 [FREE Full text] [doi: 10.2196/23233] [Medline: 33739931]
6. Brewer LC, Fortuna KL, Jones C, Walker R, Hayes SN, Patten CA, et al. Back to the future: Achieving health equity through health informatics and digital health. JMIR Mhealth Uhealth 2020 Jan 14;8(1):e14512 [FREE Full text] [doi: 10.2196/14512] [Medline: 31934874]
7. Strength through diversity - 7th policy forum, 2 March 2020. Organization for Economic Cooperation and Development. 2020 Mar 02. URL: https://www.oecd.org/education/strength-through-diversity/strength-through-diversity-7th-policy-forum-march-2020.htm [accessed 2022-02-17]
8. Meskó B, Drobni Z, Bényei É, Gergely B, Győrffy Z. Digital health is a cultural transformation of traditional healthcare. Mhealth 2017;3:38 [FREE Full text] [doi: 10.21037/mhealth.2017.08.07] [Medline: 29184890]
9. Crawford A, Serhal E. Digital health equity and COVID-19: The innovation curve cannot reinforce the social gradient of health. J Med Internet Res 2020 Jun 02;22(6):e19361 [FREE Full text] [doi: 10.2196/19361] [Medline: 32452816]
10. Global strategy on digital health 2020-2025. World Health Organization. 2021. URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf [accessed 2022-02-15]
11. Rodriguez JA, Shachar C, Bates DW. Digital inclusion as health care - supporting health care equity with digital-infrastructure initiatives. N Engl J Med 2022 Mar 24;386(12):1101-1103. [doi: 10.1056/NEJMp2115646] [Medline: 35302722]

Edited by T Leung; submitted 03.03.22; this is a non–peer-reviewed article;accepted 28.04.22; published 01.06.22.

Please cite as:
Chaudhuri ER
The Syndemic of Inequity and COVID-19 in Virtual Care
J Med Internet Res 2022;24(6):e37717
URL: https://www.jmir.org/2022/6/e37717
doi:10.2196/37717
PMID:35512269

©Esha Ray Chaudhuri. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 01.06.2022. 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

XSL•FO
RenderX

information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Short Paper

Outpatient Care Among Users and Nonusers of Direct-to-Patient Telehealth: Observational Study

Alison Cuellar[1,2], BA, PhD; J Mary Louise Pomeroy[1], MPH; Sriteja Burla[3], MA; Anupam B Jena[2,4,5], MD, PhD

[1]Department of Health Administration and Policy, George Mason University, Fairfax, VA, United States

[2]National Bureau of Economic Research, Cambridge, MA, United States

[3]Department of Economics, George Mason University, Fairfax, VA, United States

[4]Department of Health Care Policy, Harvard Medical School, Boston, MA, United States

[5]Department of Medicine, Massachusetts General Hospital, Boston, MA, United States

Corresponding Author:
Alison Cuellar, BA, PhD
Department of Health Administration and Policy
George Mason University
MS 1J3
4400 University Dr
Fairfax, VA, 22030
United States
Phone: 1 703 993 5048
Email: aevanscu@gmu.edu

Abstract

Background: Expansion of telehealth insurance coverage is hampered by concerns that such coverage may encourage excessive use and spending.

Objective: The aim of this paper is to examine whether users of telehealth services rely more on other forms of outpatient care than nonusers, and to estimate the differences in payment rates.

Methods: We examined claims data from a large national insurer in 2017. We limited our analysis to patients with visits for 3 common diagnoses (N=660,546). We calculated the total number of visits per patient, overall, and by setting, and adjusted for patient- and county-level factors.

Results: After multivariable adjustment, telehealth-visit users, compared to nonusers, had 0.44 fewer visits to primary care, 0.11 fewer visits to emergency departments, and 0.17 fewer visits to retail and urgent care. All estimates are statistically significant at $P<.001$. Average payment rates for telehealth visits were lower than all other settings.

Conclusions: These findings suggest that telehealth visits may substitute rather than add to in-person care for some types of care. Our study suggests that telehealth visits may offer an efficient and less costly alternative.

(J Med Internet Res 2022;24(6):e37574) doi:10.2196/37574

KEYWORDS

telemedicine; insurance; policy; telehealth; user; primary care; outpatient; claims; in-person; virtual; insurer; coverage

Introduction

Telemedicine has dramatically changed health care delivery since COVID-19, offering safety and convenience [1]. To promote greater access to care, federal and state policy makers temporarily removed several telehealth barriers, including those related to insurance coverage and reimbursement [2,3]. However, concerns exist about making these changes permanent. On the one hand, telehealth has the potential to improve access to care (eg, by helping patients to overcome barriers such as transportation and childcare). On the other hand, expansion of telehealth insurance coverage is hampered by concerns that such coverage may encourage excess use and spending. Such concerns are heightened when payment is on a per-visit, fee-for-service basis [4].

Most direct-to-patient (DTP) telehealth visits have been found to occur outside of regular business hours, suggesting that convenience and accommodation may be key considerations [5]. In a survey of adults conducted in 2019 before the COVID-19 pandemic, 49% of adults reported being willing or

XSL•FO
RenderX

very willing to use video visits [6]. Previous studies have found that DTP telehealth users are more likely to live in urban areas, be younger, and less likely to have comorbid conditions than the general population, indicating that access and affordability may not be the key drivers of telehealth visit use [5]. Evidence is limited with respect to the potential for telehealth visits to serve as a substitute for in-person care rather than as a complement to it. If telehealth visits substitute for in-person care, this could mean health care savings; however, if telehealth visits complement and add to in-person visits, then this would increase health care expenditures. Early evidence for telehealth acute respiratory illnesses found that telehealth visits represented additional use rather than replacing visits to other providers [7]. A more recent study of telehealth found no differences in total outpatient visits after hospital discharge, early in the pandemic [8]. In a survey of people experiencing homelessness, 29.1% self-reported they would have sought care in an emergency department (ED) if they had not had access to telehealth [9]. Additional evidence on whether telehealth complements or substitutes for in-person care is lacking.

We examined claims data from a large national insurer that offered DTP telehealth visits as a covered benefit. We assessed whether telehealth visits were associated with differences in the use of office-based primary care, retail and urgent care clinics, and EDs.

Methods

Data and Measures

Our cross-sectional study used 2017 private insurance claims data for continuously enrolled members ages 18-64 years who were offered telehealth services through a third-party DTP vendor. We limited our analysis to the top 3 most common claims diagnoses for telehealth visits in order to increase comparability across patients and sites. The most common diagnoses were respiratory infections, diseases of the urinary system, and eye disorders. We further limited our analysis to telehealth visits, office-based primary care, urgent or retail clinics, and EDs based on claim codes. We calculated the total number of visits per member who experienced at least one visit with the target diagnoses. Total visits were calculated overall and by setting. Separately, we calculated mean insurer paid amounts for evaluation and management visits for each setting.

Analysis

We estimated patient-level multivariable, negative binomial regression models in which the dependent variable was the total number of visits in a given care setting (separate regressions were run for primary care, urgent or retail clinics, and ED visits). Independent variables included an indicator for having had a telehealth visit that year, an indicator for having high-deductible health insurance coverage, and an indicator for having had a primary care office visit in the prior year. Additional control variables included age, sex, a continuous measure of illness burden using claims-based Episode Treatment Groups, a rural county indicator, county-level measures of total primary care physicians and EDs obtained from claims, and county-level counts of retail clinics and urgent care centers. We also adjusted for county-level demographics and commute times from the American Community Survey. For each visit type, we calculated regression-adjusted total visits by whether the member had a telehealth visit. Analysis was conducted using Stata (Version 15, StataCorp).

Ethics Approval

The National Bureau of Economic Research institutional review board determined this study to fall under Exemption #4 as detailed at 45 CFR Part 46 Subpart A Section 46.101 [10]. As such, it has been exempted from review.

Results

Overall, 660,546 members with the selected diagnoses had, on average, 0.56 visits for primary care, 0.60 for retail or urgent care centers, 0.13 for EDs, and 0.04 for telehealth care in 2017 (Table 1). The median insured paid amounts were US $40 for telehealth visits, US $87 for primary care, US $113 for urgent and retail clinics, and US $812 for ED visits.

After multivariable adjustment (Figure 1), telehealth-visit users had fewer visits to primary care compared with nonusers (0.13 vs 0.57, adjusted difference 0.44; 95% CI 0.44 to 0.45), EDs (0.03 vs 0.14, adjusted difference 0.11; 95% CI 0.11 to 0.11), and retail and urgent care (0.17 vs 0.62, adjusted difference 0.45; 95% CI 0.45 to 0.46). These estimates of number of visits control for the individual having a primary care visit in the prior year, age, gender, illness burden measured from claims-based Symmetry groupings, whether their insurance plan had a high deductible, rural status, and other county characteristics of the patient. The sample includes patients with respiratory infections, diseases of the urinary system, and eye disorders. All estimates are statistically significant at $P<.001$.

XSL•FO
RenderX

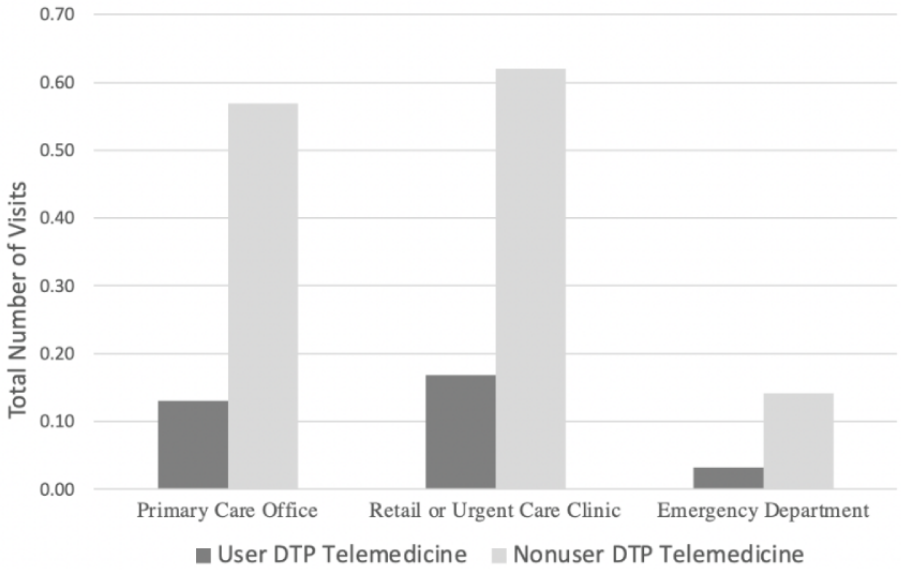
Table 1. Characteristics of the study population.

Annual visits and population characteristics	All (N=660,546)	Telehealth visit nonuser (N=639,975)	Telehealth visit user (N=20,571)	Difference (95% CI)
Annual visits				
Primary care visits	0.56	0.57	0.11	0.46 (0.45, 0.47)
Retail and urgent visits	0.60	0.62	0.18	0.44 (0.43, 0.45)
ED[a] visits	0.13	0.13	0.03	0.10 (0.10, 0.11)
Individual-level characteristics				
Age (years)	41.5	41.6	39.3	2.3 (2.2, 2.5)
Female, n (%)	410,199 (62.1)	396,785 (62.0)	13,515 (65.7)	–3.7 (–4.3, –3.0)
Primary care office visit, prior year, n (%)	280,732 (42.5)	274,549 (42.9)	6151 (29.9)	13.1 (12.5, 13.7)
High deductible health plan, n (%)	292,622 (44.3)	280,949 (43.9)	11,911 (57.9)	–14.0 (–14.7, –13.3)
Illness burden score	1.76	1.77	1.51	0.25 (0.23, 0.28)
County-level characteristics				
White, n (%)	480,877 (72.8)	465,902 (72.8)	15,140 (73.6)	–0.8 (–1.0, –0.6)
<18 years of age, n (%)	153,907 (23.3)	149,114 (23.3)	4875 (23.7)	–0.4 (–0.4, –0.3)
>65 years of age, n (%)	93,798 (14.2)	90,876 (14.2)	2859 (13.9)	0.3 (0.3, 0.4)
Bachelor's degree or higher, n (%)	227,228 (34.4)	220,791 (34.5)	6974 (33.9)	0.5 (0.4, 0.7)
Median household income (in thousands of US $)	65.2	65.3	64.1	1.2 (0.9, 1.4)
Private insurance, n (%)	457,098 (69.2)	442,863 (69.2)	14,132 (68.7)	0.5 (0.4, 0.7)
Rural county, n (%)	7927 (1.2)	7680 (1.2)	226 (1.1)	0.2 (0.0, 0.3)
Mean travel time to work (min)	27.3	27.3	26.8	0.45 (0.39, 0.52)
Primary care providers per 1000 members	8.63	8.66	7.79	0.86 (0.1, 0.1)
Retail and urgent care per 100,000 population	4.42	4.42	4.41	0.03 (–0.03, 0.05)[b]
EDs per 100,000 population	0.776	0.775	0.785	0.010 (–0.021, 0.001)[b]

[a]ED: emergency department.

[b]All differences had *P* values <.001 except these measures where *P*<.05.

Figure 1. Outpatient use for direct-to-patient (DTP) telemedicine users and nonusers.



XSL•FO
RenderX

Discussion

Among the beneficiaries of a large insurer treated for 1 of 3 common acute outpatient conditions, telehealth visit users had lower use of primary care, retail clinic and urgent care, and ED visits for those conditions. These findings suggest that telehealth visits may substitute rather than add to in-person care in some settings, although the extent of substitution or addition is unknown due to the possibility of unmeasured confounders.

In this study, insurer payment rates for third-party, DTP telehealth visits were lower than payment rates for visits in other settings, although the scope of care was narrow. Prior to the pandemic, 6 states had telemedicine parity laws, which mandated that private insurers reimburse telehealth visits on par with in-person visits [11]. By fall 2021, a total of 21 states had reimbursement parity laws for commercial insurance [12]. As these parity laws have become more common, they reduce the payment differences between in-person care and care delivered by DTP telemedicine networks, lowering potential savings to insurers.

Use of all forms of telemedicine is increasingly rapidly, offering the potential for significant improvements in health care access [13], particularly as greater investments are made to expand broadband access for rural areas and low-income families through recently passed federal legislation. Our study suggests that for some conditions, telehealth visits may be an efficient and less costly alternative to care in other settings.

Study limitations include analysis of a large, national insurer in a single year, a specific telehealth service, and select conditions, which may not generalize or apply to more specialized forms of telemedicine, including recent telehealth care that has been audio only. Commercial claims data also do not contain racial and ethnic information. Further, our study was observational and measured associations between telehealth visits and other forms of visit. Future work should examine other forms of televisit and additional patient populations.

Acknowledgments

Support was provided by the Agency for Healthcare Research and Quality (R01HS025163, AC).

Conflicts of Interest

ABJ reports receiving (in the last 36 months) consulting fees unrelated to this work from Bioverativ, Merck/Sharp/Dohme, Janssen, Edwards Life Sciences, Novartis, Amgen, Eisai, Otsuka Pharmaceuticals, Vertex Pharmaceuticals, Celgene, Sanofi Aventis, Precision Health Economics, and Analysis Group. ABJ also reports receiving (in the last 36 months) income unrelated to this work from hosting the podcast Freakonomics, MD, and from book rights to Doubleday Books. The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript. AC, JMLP, and SB declare no conflicts of interest.

References

1. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
2. Telehealth: Delivering Care Safely During COVID-19. US Department of Health and Human Services. 2020. URL: https://www.hhs.gov/coronavirus/telehealth/index.html [accessed 2022-01-01]
3. Shachar C, Engel J, Elwyn G. Implications for Telehealth in a Postpandemic Future: Regulatory and Privacy Issues. JAMA 2020 Jun 16;323(23):2375-2376. [doi: 10.1001/jama.2020.7943] [Medline: 32421170]
4. Expansion of telehealth in Medicare public meeting. MedPAC. URL: http://www.medpac.gov/-public-meetings-/meeting-details/september-2020-public-meeting [accessed 2020-12-03]
5. Fischer SH, Ray KN, Mehrotra A, Bloom EL, Uscher-Pines L. Prevalence and Characteristics of Telehealth Utilization in the United States. JAMA Netw Open 2020 Oct 01;3(10):e2022302 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.22302] [Medline: 33104208]
6. Jain T, Mehrotra A. Comparison of Direct-to-Consumer Telemedicine Visits With Primary Care Visits. JAMA Netw Open 2020 Dec 01;3(12):e2028392 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.28392] [Medline: 33289842]
7. Ashwood JS, Mehrotra A, Cowling D, Uscher-Pines L. Direct-To-Consumer Telehealth May Increase Access To Care But Does Not Decrease Spending. Health Aff (Millwood) 2017 Mar 01;36(3):485-491. [doi: 10.1377/hlthaff.2016.1130] [Medline: 28264950]
8. Bressman E, Russo A, Werner RM. Trends in Outpatient Care and Use of Telemedicine After Hospital Discharge in a Large Commercially Insured Population. JAMA Health Forum 2021 Nov 12;2(11):e213685-e213391 [FREE Full text] [doi: 10.1001/jamahealthforum.2021.3685] [Medline: 33196765]
9. Adams CS, Player MS, Berini CR, Perkins S, Fay J, Walker L, et al. A Telehealth Initiative to Overcome Health Care Barriers for People Experiencing Homelessness. Telemed J E Health 2021 Aug 01;27(8):851-858. [doi: 10.1089/tmj.2021.0127] [Medline: 34297907]
10. HHS.gov. Office for Human Research Protections (OHRP). 45 CFR 46. US Department of Health and Human Services. URL: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/index.html [accessed 2022-05-30]

XSL•FO
RenderX

11. State Telehealth Laws and Reimbursement Policies: A Comprehensive Scan of the 50 States and the District of Columbia. Center for Connected Health Policy. 2019. URL: https://www.cchpca.org/sites/default/files/2019-10/50%20State%20Telehealth%20Laws%20and%20Reibmursement%20Policies%20Report%20Fall%202019%20FINAL.pdf [accessed 2020-05-26]
12. State Telehealth Laws and Reimbursement Policies: Summary Chart, Fall 2021. Center for Connected Health Policy. 2021. URL: https://www.cchpca.org/2021/10/Fall2021_StateSummaryChart_FINAL.pdf [accessed 2022-01-10]
13. Benda NC, Veinot TC, Sieck CJ, Ancker JS. Broadband Internet Access Is a Social Determinant of Health!. Am J Public Health 2020 Aug;110(8):1123-1125. [doi: 10.2105/ajph.2020.305784]

Abbreviations

DTP: direct-to-patient
ED: emergency department

Edited by G Eysenbach; submitted 25.02.22; peer-reviewed by D Madhusudhan; comments to author 06.05.22; revised version received 18.05.22; accepted 20.05.22; published 06.06.22.

Please cite as:
Cuellar A, Pomeroy JML, Burla S, Jena AB
Outpatient Care Among Users and Nonusers of Direct-to-Patient Telehealth: Observational Study
J Med Internet Res 2022;24(6):e37574
URL: https://www.jmir.org/2022/6/e37574
doi:10.2196/37574
PMID:35666556

©Alison Cuellar, J Mary Louise Pomeroy, Sriteja Burla, Anupam B Jena. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.06.2022. 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

Using Social Media for Clinical Research: Recommendations and Examples From the Brown-Lifespan Center for Digital Health

Elizabeth M Goldberg[1,2,3], ScM, MD; Rochelle K Rosen[3,4,5], PhD; Don S Dizon[3,6,7], MD; Kirsten J Langdon[3,5,8], PhD; Natalie M Davoodi[9], MPH; Tyler B Wray[3,10], PhD; Nicole R Nugent[3,5], PhD; Shira I Dunsiger[3,4], PhD; Megan L Ranney[1,3,4], MPH, MD

[1]Department of Emergency Medicine, Warren Alpert Medical School of Brown University, Providence, RI, United States

[2]Department of Health Services, Policy & Practice, Brown University School of Public Health, Providence, RI, United States

[3]Brown-Lifespan Center for Digital Health, Providence, RI, United States

[4]Department of Behavioral and Social Sciences, School of Public Health, Brown University, Providence, RI, United States

[5]Department of Psychiatry and Human Behavior, Warren Alpert Medical School of Brown University, Providence, RI, United States

[6]Department of Medical Oncology, Warren Alpert Medical School of Brown University, Providence, RI, United States

[7]Lifespan Cancer Institute, Rhode Island Hospital, Providence, RI, United States

[8]Department of Psychiatry, Rhode Island Hospital, Providence, RI, United States

[9]Department of Emergency Medicine, Providence, RI, United States

[10]Center for Alcohol and Addiction Studies, Department of Behavioral and Social Sciences, School of Public Health, Brown University, Providence, RI, United States

Corresponding Author:
Megan L Ranney, MPH, MD
Department of Behavioral and Social Sciences
School of Public Health
Brown University
121 S Main Street
Providence, RI, 02903
United States
Phone: 1 (401) 863 3375
Email: megan_ranney@brown.edu

Abstract

Social media integration into research has increased, and 92% of American social media participants state they would share their data with researchers. Yet, the potential of these data to transform health outcomes has not been fully realized, and the way clinical research is performed has been held back. The use of these technologies in research is dependent on the investigators' awareness of their potential and their ability to innovate within regulatory and institutional guidelines. The Brown-Lifespan Center for Digital Health has launched an initiative to address these challenges and provide a helpful framework to expand social media use in clinical research.

(J Med Internet Res 2022;24(6):e35804) doi:10.2196/35804

KEYWORDS

social media; Twitter; Facebook; clinical research; privacy; institutional review board; regulations; regulation; guideline; big data

Introduction

Social media includes technologies that allow multidirectional communication via web-based networks (Facebook), microblogs (Twitter), video sharing sites (YouTube), blogs, and other forums [1]. A 2021 Pew Research Center survey found that 72% of American adults use some form of social media, with that figure surpassing 80% among those under 50 years of age [2]. As social media use increases, its integration into and relevance for clinical research has also increased. These web-based channels offer a low- or no-cost venue for recruitment [3-6], more ready venues for volunteer engagement [7], and greater generalizability owing to the potential of web-based tools to access diverse or marginalized communities [6,8].

XSL•FO
RenderX

Beyond these aspects, social media also offers tremendous opportunities for clinical researchers. First, it provides the opportunity to increase knowledge about clinical research in a way that encourages the public to learn and discuss issues. For instance, during the COVID-19 pandemic, the National Science Foundation funded COVID Info Commons [9], a "convergence accelerator" that promotes federally funded research on COVID-19 on its Twitter account, provides the public with a search engine to find National Science Foundation–funded COVID research, holds monthly seminars over Zoom, accessible to the public, on research in progress, and posts recorded seminars to YouTube with Spanish and American Sign Language interpretation. Second, social media also allows for the delivery of interventions in an innovative way [8] and is a potential source of real-world evidence that can be accessed to generate new hypotheses or identify unmet needs in various clinical communities [10,11].

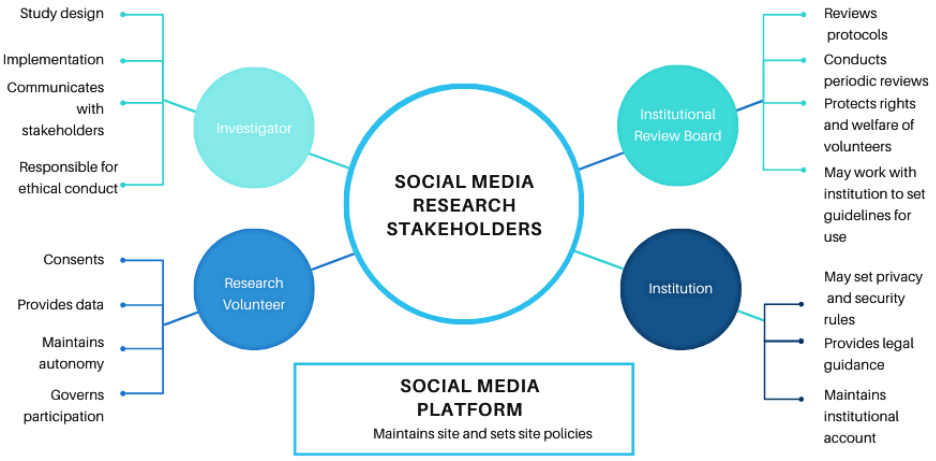
Despite the potential applications that can be used, social media research still faces significant barriers to its effective use. Principal among them is the lack of uniformity in how research proposals are reviewed at a local and national level and the lack of guidance available to researchers seeking to explore social media; this in turn may result in the unintended consequence of discouraging new and established researchers from incorporating social media into their own work.

The Brown-Lifespan Center for Digital Health (CDH) is a hub where researchers, clinicians, administrators, entrepreneurs, and business representatives from Brown University and its affiliated hospital partners collaboratively design, test, and deploy digital solutions to the society's most pressing health challenges. In this paper, we review the issues facing investigators and institutions related to research using social media technologies. We propose a roadmap for researchers, agencies, and institutions to integrate social media as a tool for completing clinical research studies. We focus on elements of social media use investigators should be cognizant of, issues institutional review boards (IRBs) should address, and suggest institutional procedures to facilitate safe and responsible social media use for clinical research.

Social Media and the Clinical Researcher: Concerns and Considerations

While the potential role of social media in research has been established, issues have been raised by investigators, including ethics, privacy, consent, and confidentiality concerns for participants [12]. Additionally, whether and how communities on social media represent real-world or offline communities is a concern [13], especially as older and underresourced individuals may lack access to broadband or familiarity with social media channels; this inequality is often termed the "digital divide" [14]. Moreover, new social media platforms are constantly emerging, introducing dynamically changing impacts to participants and researchers, while also reshaping use patterns of more established tools. The stakeholders in social media research and their key roles are illustrated in Figure 1.

Figure 1. Social media research stakeholders.



Institutional Perspectives on Social Media Research: Concerns and Considerations

The use of social media for the purposes of human subjects research remains an area of concern for many institutions. The major reason this remains the case is that social media is not designed for ethical, human subject–approved research. Rather, it is designed, by intent, for public use. Further, social media platforms often include data agreements between the platform's creators and all users that allow third parties to data mine in order to influence people using these platforms (eg, by targeted advertising or streamlined content). As a result, the intention for social media platforms is diametrically opposed to the most basic tenets of clinical research, including but not limited to the importance of deidentification and human subject protection. These concerns were succinctly brought forth in a recent paper by Vallury et al [15] in detailing their experience assessing public attitudes related to abortion in Australia. The authors describe how the lead researcher of a study on abortion stigma experienced "a barrage of harassment on and beyond social media" when her web-based research went "viral." The lessons learned include the need for a supportive and coordinated institutional response to plan for and manage web-based and offline mental and physical health and safety risks. They recommend the development of training, guidelines, and policies

XSL•FO
RenderX

to address the practical and ethical aspects of using social media for research.

Social media research requires an understanding of the following: the ethics of using web-based data as research data; the responsibilities of the researcher to participants both during and following the study; ensuring diversity and equity in who can access the study; and the risks and consequences to the researcher and the institution (particularly if the subject matter reflects politically or socially controversial topics). Ultimately, while the approach to social media research must be based on traditional understandings of good clinical practice and protections (for participants and for researchers), social media reflects an ever-changing environment that institutions must be prepared to recognize and effectively respond to.

The CDH Proposal on Social Media Research Applications

As social media research proliferates, research practices in digital health and social media must similarly be regularly reviewed and updated so that they evolve as well. In short, best practices and local research guidelines need to be established and regularly updated to facilitate the protection of research study volunteers and the investigators involved.

In Table 1, we listed several critical questions that researchers should address during the *design* of a study that uses social media and provided suggestions as to how each can be approached. At the earliest stages, it is incumbent on research teams to establish norms for their social media research, including how to safeguard identifiable information, ensure privacy of both the study team and research participants, and maintain confidentiality of research documents. These plans should be provided as written documentation to the local IRB. Table 2 includes examples of how CDH-affiliated faculty used social media for research, including references.

XSL•FO
RenderX

Table 1. Questions for investigators to address during study planning.

Category	Issue	Critical questions	Suggested approaches
Approach	Recruitment	• Will participants be recruited via traditional means (in research facilities, over the phone, or by flyers), by social media, or both? Will other strategies, such as crowdsourced or gig economy social media recruitment be used? • What social media platform (eg, LinkedIn, Twitter, Instagram, Facebook, or Discord) will be used? Rationale? • How will the investigator approach sampling? • What social media account (eg, related to a research lab, an institution, or an investigator) will the research team use? Provide rationale. • Will the participants be compensated?	• Provide data on the demographics of the participants as these may vary depending on the social media network employed. • Share how the participants will be routed from social media sites to Health Insurance Portability and Accountability Act–compliant data collection software sites to obtain further information. • Obtain letters of agreement from social media account participants to collaborate (eg, institutional or influencer). • Include social media community members in research design and implementation whenever feasible and appropriate.
Research team	Expertise	• Who on your team has expertise in social media use?	• Team members should have experience in social media research. If not available, ensure collaborators are involved who do.
Research plan	Dissemination	• Will data sets collected over social media be shared? With whom? How? • How will participants be informed of study progress and results?	• Unless specifically approved otherwise, only share deidentified data. • Informing participants of the study results is the responsibility of the research team.
Human subjects protection	Privacy and confidentiality	• How will personal identifiers including social media account names be protected by the research team? • What data will be obtained from social media? • Will account analytics, such as on Twitter or videos, be used? • What consent process will be used prior to data acquisition? • Will teams verify the identities of social media participants? How? • How will teams deidentify the accounts? • Will the research team engage with participants via social media?	• Be aware of the platform's privacy and confidentiality policy [12]. • Clarify what data are available publicly versus what data are available only with consent. • If electronic consent will be used, describe the consent process and how participant comprehension is verified, and provide strategies to verify that the participant meets the eligibility criteria of the study. • Clarify and assess understanding of protocols for social media posting (eg, participant-posted photos and video will be part of the research record), including who can create or add content and who will be able to see or use it. • Define provisions to reduce risk of breach of confidentiality.
Human subjects protection	Security	• How will information be collected and stored? • How will the team ensure that third parties will not have access to information about the participant's interests or affiliations?	• Describe the process for the collection of public versus private data, and whether third-party services will be used to facilitate data collection. • Specify that third parties will not have access to answers to investigator-posted surveys or screening instruments [4]. • Be aware and describe relevant institutional policies on social media use.
Human subjects protection	Risks	• How will disclosures of self-harm, trolling or other harmful comments, and other human subject concerns be monitored and identified? What is the crisis mitigation plan if disclosures are identified?	• Describe strategies for mitigating and addressing risks to participants (eg, as described by Nicholas et al [10] in "risk detection"), including frequency of monitoring, anonymity of subjects, and crisis mitigation plans. • Disclose to participants that you will not be monitoring their responses in real time, and provide them with a document or create a blanket post that lists resources for immediate help.
Human subjects protection	Recruitment	• If material will be posted on social media for purposes of recruitment: • Where will the ads be posted? • Will the ads be targeted to certain demographics? How? • How will ambient privacy be maintained?	• Provide examples of the kinds of ads or communication that may be used in the study. In order to harness the social media networks' full potential to build community, investigators may need to be agile, and it is not feasible to submit verbatim advertisements and communication to the institutional review board.

XSL•FO
RenderX

Category	Issue	Critical questions	Suggested approaches
Human subjects protection	Equity and diversity	• What strategies will be used to ensure recruitment includes women, minorities, and other underrepresented communities?	• Describe plans to ensure equitable access to recruitment and estimate likelihood of recruitment of demographic subgroups. • Consider the fact that recruitment techniques that enroll web-based participants looking for paid work (eg, through MTurk or Craigslist) may result in more demographically diverse participants than those that use a recruiting ad (eg, Facebook) [16].
Protection of the study team	Risks	• What are the foreseeable risks to the study team in the conduct of this research? • What is the plan to mitigate these risks?	• Restrict social media communication to handles specific to the study, not to any one individual on the research team. • Avoid using personal social media handles to communicate research-related activities.

Table 2. Examples from the Center for Digital Health faculty illustrating how social media can be used for clinical research.

Use category	Study topic and authors	Notes
Recruitment	Telehealth in Older Adults, Goldberg et al [17]	Physicians were recruited into qualitative interviews through advertisements posted on Twitter, Facebook physician groups, and specialty society and physician listservs
Identified individuals for an intervention	A Cyberbullying Media-Based Prevention Intervention for Adolescents on Instagram: Pilot Randomized Controlled Trial, Kutok et al [18]	Recruited a national sample of adolescents with a history of past-year cybervictimization through Instagram for a randomized control trial delivered via an app-based program.
Idea generation, iterative improvement of app based on participant feedback, and dissemination	MyCovidRisk—a free app to help individuals assess their risk of being infected with COVID-19, Goldberg et al [19]	A collaboration was formed between 2 investigators after a Twitter conversation about the need of an app that assists the public with assessing COVID-19 risk. Then, the investigators crowdsourced opinions on risk categories and what was considered an "acceptable" risk by the public on Twitter. The investigators shared a beta version of the app on Twitter and modified the design and content based on public feedback. Finally, information about how to access the app was advertised on Twitter and other social media channels.
Performed a needs assessment	The Needs of Women Treated for Ovarian Cancer: Results From a #gyncsm Twitter Chat, Thomas et al [20]	Investigators obtained IRB[a] approval to conduct a tweet chat asking women about survivorship from ovarian cancer. Questions were asked surrounding needs after cancer treatment, and the responses were analyzed quantitatively and qualitatively.
Used Twitter to obtain data from users in a specific location and analyzed the results qualitatively	#PuertoRicoSeLevanta: A Closer Look at the Language Used on the First-Year Anniversary of Hurricane Maria, Rodríguez-Guzmán et al [21]	In order to examine psychological processes 1 year after Hurricane Maria and understand the differences in reactions depending on location, the research team collected tweets using hashtags associated with Hurricane Maria and geomapping. They used Linguistic Inquiry and Word Count software (LIWC2015, Pennebaker) to conduct a quantitative linguistic analysis of the sample of tweets.
Created a novel data set using crowdsourcing	Crowdsourcing from Scratch: A Pragmatic Experiment in Data Collection by Novice Requesters, Papoutsaki et al [22]	Used crowdsourcing techniques and Amazon Mechanical Turk to create a data set of all Computer Science faculty in the 50 top Computer Science graduate programs. This project yielded guidelines that novice requesters can use who are new to using crowdsourcing for data collection and extraction from the web.
Obtained insights on human affect	Sochiatrist: Signals of Affect in Messaging Data, Massachi et al [23]	Extracted social media data and deidentified them to understand how messages can serve as a proxy for changes in a person's affect.

[a]IRB: institutional review board.

Practical Guidance for Institutional Review Boards

We recommend that IRBs develop policies surrounding the appropriate and safe use of social media in clinical research. Sharing these guidelines with researchers who plan to use social media in their studies will help ensure consistency and can be useful for investigators and IRBs alike to improve efficiency and reduce the need for revisions. Gelinas et al [24] created an IRB checklist for evaluating social media recruitment proposals that can be a valuable resource for this purpose. Below, we summarize major considerations related to recruitment, benefits, risks, and informed consent.

Recruitment

The IRB application should specify which social media sites will be used and why, whether advertisements of the study will be used, and how targeted recruitment will be conducted, if

XSL•FO
RenderX

applicable. For instance, recruitment advertisements posted on social media may draw global participation. Because of this, investigators should specify how they will ensure only eligible participants in the preferred geographic region will be recruited. Typically, this can be addressed by the inclusion of appropriate eligibility criteria as a part of a screening survey. The steps in the IRB process pertaining to social media–based recruitment methods are illustrated in Figure 2.

When using social media to recruit, researchers should put safeguards in place. Investigators should be aware that they may receive survey replies from fictitious accounts or be the target of harassment or other trolling behaviors seeking to discredit the study. Several methods exist to address these concerns [15], including the following: (1) offering compensation for users to verify that they are indeed who they are and delaying payment until completed; (2) regular and routine monitoring of advertisements and posts related to the study; (3) understanding the policies governing privacy, harassment, and reporting on the channels being used; (4) adopting mechanisms to moderate posting on public forums related to the research; (5) if surveys are used, users should take advantage of security measures to prevent fraud and mitigate malfeasance (eg, they can use Completely Automated Public Turing Test question types to prevent bots from submitting survey responses); in addition, proactively monitoring times to completion can be useful, as in our experience, completion rates are typically very fast for fictitious accounts; and (6) it is important to monitor the referring link to determine if links are being reshared for fraudulent purposes. Several surveys also allow investigators to prevent multiple submissions from one device by using cookies.

Figure 2. Steps in the institutional review board (IRB) process pertaining to social media–based recruitment methods.

Researchers should acknowledge and disclose terms of sites or apps before advertising recruitment on them to avoid third-party data sharing. For example, research participants who engage with study advertisements could—depending on the type of advertisement and social media site—allow third-party sites to collect information about their interests or affiliations unknowingly [25]. Participants are often unaware of the terms of the apps and sites they use regularly. Investigators should be knowledgeable of these terms before they advertise their studies on these sites. In one study of men who have sex with men [25], participants had few concerns about data being shared anonymously with researchers but expressed more concerns with data being sold to third-party partners. However, research participants evidenced substantial variability in privacy concerns

XSL•FO
RenderX

and comfort with sharing different types of data, suggesting a need to gain consent for data sharing for specific types of data.

The investigator should also specify how potential participants' privacy will be protected during the advertising process. For example, a study recruiting people with a history of substance use disorder should be careful not to inadvertently violate participants' privacy by advertising imagery or language that labels a potential participant as a person with an addiction. Research participants may not be forthcoming or truthful with their answers if they are particularly concerned about evading law enforcement [26], which may raise data quality concerns that should be addressed by the investigator. Investigators should educate their study participants about the limits of their confidentiality as needed.

Benefits and Risks of Social Media Use for Research

The benefits of social media use for research include (1) ease of recruitment; (2) increased engagement by social media participants; (3) rapid sharing of information in a way that is intuitive for participant; (4) and building of web-based communities; moreover (5) several studies have shown that participants feel web-based participation in research is more private than in-person participation.

The risks of using social media for research include the following: (1) third-party use of data such as tracking participants' clicks on advertisements; (2) breach of confidentiality through intentional or unintentional sharing of data by participants or the study team; and (3) exposure to malicious content; however, if the investigator is only using social media to recruit, there is no additional exposure to malicious content outside of what is seen from scrolling through your feed.

Informed Consent

Researchers using social media for their studies may choose to offer research volunteer electronic informed consent (e-consent) if there is no waiver of consent. e-Consent refers to the use of electronic systems and processes to inform research participants of information related to the study and obtain and document their consent. In guidance prepared jointly by the Department of Health and Human Services and the Food and Drug Administration [27] for investigators, sponsors, and IRBs, the following recommendations were made for e-consent: (1) e-consent should be designed to convey information about the study to the research volunteer or their legally authorized representative in language that is understandable; (2) e-consent should allow navigation forward or backward so that participants can review information, and hyperlinks can be used to view further detail; (3) participants should have the option to use paper-based consent or be assisted by study personnel if they cannot use the e-consent technology; (4) study personnel should verify identity through a state-issued identification, the use of personal questions, biometric methods, or visual methods. Verification using these techniques may not be necessary in social behavioral minimal risk research studies; (5) opportunities to ask questions and consider participation are necessary; questions can be answered in person, over the phone, or by video conferencing, but should be answered prior to consent; (6) investigators should assess understanding of the study (eg, by including questions that test understanding or through other methods to gauge individuals' comprehension of all elements of the consent; (7) participants should obtain a copy of the informed consent; and (8) IRBs should review the usability of the e-consent material to ensure they are easy to navigate and should review any optional questions or other methods used to gauge comprehension of key study elements.

An important aspect of consent relates to vulnerable populations, including but not limited to children and prisoners. For these participants, it will be important to request that researchers provide information to prevent coercion and a means to affirm consent, respectively. Finally, provisions for re-consent are necessary if the child comes of age during the study or if cognition improves or worsens during a longitudinal study in older adults [28].

Confidentiality, Security, and Privacy

Investigators should describe how privacy protections are put in place for participants. For instance, are apps "sandboxed" so that apps on the same device cannot obtain data that the participant enters into the research app? If data are being collected, where will they be stored and who maintains access? For instance, most volunteers understand that if they post on social media sites publicly, their information will be discoverable by any user of the social media site. However, volunteers may not know that if they use more private ways to communicate with the research team, their information can still be retained by the platform. Twitter, for instance, allows participants to use "Direct Messages" to have nonpublic conversations on the platform. While these direct messages are not public per se, Twitter still stores and processes the communication and information shared in these messages [29]. For instance, links shared in direct messages are scanned for malicious content. Further, Twitter will not use the content of your message; however, information about whom you communicated with and when will be examined to better understand platform usage in an effort to generate more relevant content. Volunteers should also be aware that even if they delete their copy of the direct message, recipients (in this case, the research team) will retain their own copy, which they can duplicate, store, or reshare.

Other relevant questions include the following: Will participants have a right to view or edit their data? Moreover, how do you protect the privacy of parties who have not consented? For instance, for studies on Facebook, if you are an investigator and you are observing a research volunteer's feed, you may see comments on the feed by their unconsented friends. It is important that researchers have a plan to include or exclude data from people not consented. These details require careful thought and consideration prior to initiating recruitment via social media platforms to ensure the protection of human subjects.

XSL•FO
RenderX

Investigators should also consider that third parties may develop novel ways to broach the security of platforms and exploit the identifiers of account holders. For instance, the administrator of a Facebook group, consisting of individuals who tested positive for breast-cancer mutations, discovered a Chrome extension that allowed marketers to scrape the membership lists of closed Facebook groups [30]. Facebook had previously added tools to make the membership lists of closed groups private, and they were unaware of this Chrome extension until the group administrator worked with a security researcher to submit the information to Facebook. Facebook then sent a cease-and-desist letter to the Chrome extension.

Statistical Analysis

IRBs should be aware of unique uses and analytic techniques for social network analysis. Social network analysis often involves large samples and can have substantial computational requirements. For instance, in a study aiming to discover emergent web-based communities of cannabis participants for public health surveillance [8], investigators performed social network analysis by first finding the actors of interest, 6 cannabis dispensaries in Oakland, and then discovering accounts that follow these 6 accounts and their followers. Then, participant information was collected from these accounts such as friend counts, follower counts, and account creation date. The total number of accounts collected by these means included 2.2 million participants. Then, researchers used stochastic block modeling to infer network structure with the purpose of uncovering hidden populations of cannabis consumers. After manual coding, communities of illicit, recreational, and medical cannabis participants were identified. This analysis helped researchers examine a research question and illicit use patterns that would be challenging and costly to discover without social media analyses. However, these methods are computationally complex and require expertise in big data (analysis and data management) beyond what many investigators may need for traditional clinical research studies. Investigators need to be skilled in these advanced statistical techniques, such as stochastic block modeling and high-dimensional multilevel models, as well as qualitative content analysis, in order to identify spam and fraudulent accounts and ensure the validity of their findings. It is important to note that this level of work often requires significant server space and power to run the analyses; this availability may vary depending on institutional resources.

Institutional Procedures to Facilitate Safe and Effective Social Media Use

Institutions may opt to publish social media guides when used for research to help investigators follow institutional privacy and security recommendations and to help them follow best practices in social media use. For instance, the Harvard Clinical and Translational Science Center publishes a guide, "The Use of Social Media in Recruitment to Research: A Guide for Investigators and IRBs," that summarizes their laws and regulations, including Health Insurance Portability and Accountability Act, advises on recruitment techniques that follow their social media guidelines, and assists investigators in designing procedures that respect ethical norms [31]. The University of South Florida provides specific parameters for their faculty and staff to guide the development of a social media presence [32]. A mixed methods study including interviews with 5 institutional offices and 15 subject-matter experts at the University of Florida suggests that a centrally managed social media account for communicating with participants and initiating advertising campaigns could be successful to facilitate participant enrollment in health and clinical research studies [33]. Some institutions list social media accounts that have pre-existing approval for research usage [34,35]. However, if an institutional account is not already approved, it is recommended that the social media or public relations team from the institution work with the IRB and Human Subjects Protective Program to agree on guidelines for social media use in clinical research. Given that terms of agreement often include legal jargon, which may be confusing for investigators and research volunteers, it can be helpful to involve the institution's legal team to help with interpreting terms of the chosen social media platforms.

Conclusion

The right to privacy and data security is a fundamental aspect of clinical research that must be considered in the social media space. Researchers are expected to uphold the principles of trust and respect by approaching the aims and details of the study with transparency and refraining from collecting data about potential participants in ways unknown to the social media participant. Communication between the research team and research participants should be carried out in such a way to avoid breaches of confidentiality or exposing personal information in the public domain. Because communication may be frequent and cannot always be completely scripted on social media sites, it is beneficial for IRBs and institutions to agree to norms that allow the investigator to have flexibility to communicate with research participants in a manner that is consistent with the study aims and the IRB protocol. Finally, given the ever-changing terms of use and privacy policies on social media sites, it is critical for study teams to maintain awareness of such policies and develop plans to ensure ongoing compliance.

Further work is needed to (1) identify what unique safeguards may be necessary for individuals with special situations that make them more vulnerable to exploitation (eg, undocumented individuals, minors, and sex workers), (2) develop recruitment techniques and interventions tailored to special populations who are traditionally disadvantaged by the digital divide (eg, older individuals and rural persons), (3) suggest ways researchers can best recruit volunteers and access data from social media sites while being sensitive to the diverse privacy needs of volunteers (eg, different comfort levels with disclosure), (4) ensure all stakeholders understand the limitations of different platforms' privacy policies, and (5) develop best techniques to disclose and increase the comprehension of yet unidentified vulnerabilities in platforms that can be exploited by third parties.

XSL•FO
RenderX

Social media can be a valuable tool for clinical research recruitment, retention, data collection, and dissemination. However, as an open and shareable entity, there is a possible dissonance between traditional research ethics and the public use of social media sites. Social media research stakeholders should be aware that our understanding of the ideal privacy policies and other safeguards for volunteers are still evolving and will likely never be static. Regulatory agencies, such as IRBs, and funding agencies should share clear guidelines for social media use in research to enhance innovation and ensure privacy and efficiency.

Acknowledgments

EMG, RKR, DSD, KJL, NMD, and MLR contributed to writing the manuscript, editing, and redrafting. EMG, RKR, and DSD devised, wrote, and organized the contributions of this manuscript. TBW, NRN, and SID provided critical edits and revisions in writing the manuscript. All authors have read, edited, and approved the current version of the manuscript. We would like to acknowledge all the Brown-Lifespan Center for Digital Health faculty and staff leadership, including Michael Armey, Don Dizon, Shira Dunsiger, Zahrie Ernst, Mustafa Mazlumoglu, Ross Hilliard, Jeff Huang, Rebecca Laferriere, John Pateña, Rochelle Rosen, Allison Seeley, Anthony Spirito, Kimani Toussaint Jr., and Tyler Wray. EMG received funding for this work from the National Institute on Aging (K76 AG-059983). KJL received funding from the National Institute on Drug Abuse (K23DA046482), the Centers of Biomedical Research Excellence (P20GM125507), and the Center for AIDS Research (P30AI042853).

Conflicts of Interest

DSD is a consultant for Pfizer and KJL is a consultant for Noom, Inc. The other authors have no conflicts to declare.

References

1. NIH Office of Intramural Research. Guidance Regarding Social Media Tools. National Institutes of Health. 2016 Feb 16. URL: https://www.nih.gov/health-information/nih-clinical-research-trials-you/guidance-regarding-social-media-tools [accessed 2021-10-22]
2. Auxier B, Anderson M. Social Media Use in 2021. Pew Research Center. 2021 Apr 07. URL: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ [accessed 2021-10-26]
3. Kutok ER, Doria N, Dunsiger S, Patena JV, Nugent NR, Riese A, et al. Feasibility and Cost of Using Instagram to Recruit Adolescents to a Remote Intervention. J Adolesc Health 2021 Nov;69(5):838-846. [doi: 10.1016/j.jadohealth.2021.04.021] [Medline: 34059428]
4. Fisher CB, Bragard E, Bloom R. Ethical Considerations in HIV eHealth Intervention Research: Implications for Informational Risk in Recruitment, Data Maintenance, and Consent Procedures. Curr HIV/AIDS Rep 2020 Jun;17(3):180-189 [FREE Full text] [doi: 10.1007/s11904-020-00489-z] [Medline: 32358768]
5. Khatri C, Chapman SJ, Glasbey J, Kelly M, Nepogodiev D, Bhangu A, STARSurg Committee. Social media and internet driven study recruitment: evaluating a new model for promoting collaborator engagement and participation. PLoS One 2015 Mar 16;10(3):e0118899 [FREE Full text] [doi: 10.1371/journal.pone.0118899] [Medline: 25775005]
6. Batterham PJ. Recruitment of mental health survey participants using Internet advertising: content, characteristics and cost effectiveness. Int J Methods Psychiatr Res 2014 Jun;23(2):184-191 [FREE Full text] [doi: 10.1002/mpr.1421] [Medline: 24615785]
7. Grajales F, Clifford D, Loupos P, Okun S, Quattrone S, Simon M, et al. Social Networking Sites and the Continuously Learning Health System: A Survey. National Academy of Medicine. 2021. URL: https://nam.edu/perspectives-2014-social-networking-sites-and-the-continuously-learning-health-system-a-survey/ [accessed 2021-11-08]
8. Baumgartner P, Peiper N. Utilizing Big Data and Twitter to Discover Emergent Online Communities of Cannabis Users. Subst Abuse 2017;11:1178221817711425 [FREE Full text] [doi: 10.1177/1178221817711425] [Medline: 28615950]
9. COVID Information Commons. Columbia University in the City of New York. 2021. URL: https://covidinfocommons.datascience.columbia.edu/ [accessed 2021-10-26]
10. Nicholas J, Onie S, Larsen ME. Ethics and Privacy in Social Media Research for Mental Health. Curr Psychiatry Rep 2020 Nov 23;22(12):84. [doi: 10.1007/s11920-020-01205-9] [Medline: 33225404]
11. Mittelstadt BD, Floridi L. The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. Sci Eng Ethics 2016 May;22(2):303-341. [doi: 10.1007/s11948-015-9652-2] [Medline: 26002496]
12. Moreno MA, Goniu N, Moreno PS, Diekema D. Ethics of social media research: common concerns and practical considerations. Cyberpsychol Behav Soc Netw 2013 Oct;16(9):708-713 [FREE Full text] [doi: 10.1089/cyber.2012.0334] [Medline: 23679571]
13. 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]

XSL•FO
RenderX

14. Dizon DS, Sedrak MS, Lewis MA, Cook E, Fisch MJ, Klemp JR, SWOG Digital Engagement Committee. Incorporating Digital Tools to Improve Clinical Trial Infrastructure: A White Paper From the Digital Engagement Committee of SWOG. JCO Clin Cancer Inform 2018 Dec;2:1-8 [FREE Full text] [doi: 10.1200/CCI.17.00122] [Medline: 30652537]
15. Vallury KD, Baird B, Miller E, Ward P. Going Viral: Researching Safely on Social Media. J Med Internet Res 2021 Dec 13;23(12):e29737 [FREE Full text] [doi: 10.2196/29737] [Medline: 34898450]
16. Antoun C, Zhang C, Conrad FG, Schober MF. Comparisons of Online Recruitment Strategies for Convenience Samples. Field Methods 2015 Sep 16;28(3):231-246. [doi: 10.1177/1525822X15603149]
17. Goldberg EM, Jiménez FN, Chen K, Davoodi NM, Li M, Strauss DH, et al. Telehealth was beneficial during COVID-19 for older Americans: A qualitative study with physicians. J Am Geriatr Soc 2021 Nov;69(11):3034-3043 [FREE Full text] [doi: 10.1111/jgs.17370] [Medline: 34245165]
18. Kutok ER, Dunsiger S, Patena JV, Nugent NR, Riese A, Rosen RK, et al. A Cyberbullying Media-Based Prevention Intervention for Adolescents on Instagram: Pilot Randomized Controlled Trial. JMIR Ment Health 2021 Oct 15;8(9):e26029 [FREE Full text] [doi: 10.2196/26029] [Medline: 34524103]
19. Goldberg EM, Bingaman CS, Perera S, Ranney ML. MyCOVIDRisk app: development and utilisation of a COVID-19 risk assessment and mitigation application. BMJ Innov 2021 Mar 30;7(2):363-367. [doi: 10.1136/bmjinnov-2021-000672]
20. Thomas TH, Nauth-Shelley K, Thompson MA, Attai DJ, Katz MS, Graham D, et al. The Needs of Women Treated for Ovarian Cancer: Results From a #gyncsm Twitter Chat. J Patient Cent Res Rev 2018;5(2):149-157 [FREE Full text] [doi: 10.17294/2330-0698.1592] [Medline: 31413999]
21. Rodríguez-Guzmán VM, García-Ramírez GM, Bogen KW, Orchowski LM, Nugent N. #PuertoRicoSeLevanta: A Closer Look at the Language Used on the First-Year Anniversary of Hurricane Maria. J Technol Behav Sci 2021 Jun;6(2):358-364. [doi: 10.1007/s41347-020-00167-2] [Medline: 34337146]
22. Papoutsaki A, Guo H, Metaxa-Kakavouli D, Gramazio C, Rasley J, Xie W, et al. Crowdsourcing from Scratch: A Pragmatic Experiment in Data Collection by Novice Requesters. In: Proceedings of the AAAI Conference on Human Computation and Crowdsourcing.: AAAI Press; 2015 Sep 23 Presented at: Third AAAI Conference on Human Computation and Crowdsourcing; November 8-11, 2015; San Diego, California, USA p. 140-149 URL: https://www.aaai.org/ocs/index.php/HCOMP/HCOMP15/paper/viewFile/11582/11436
23. Massachi T, Fong G, Mathur V, Pendse SR, Hoefer G, Fu JJ, et al. Sochiatrist. Proc. ACM Hum.-Comput. Interact 2020 Oct 14;4(CSCW2):1-25. [doi: 10.1145/3415182]
24. 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]
25. Rendina HJ, Mustanski B. Privacy, Trust, and Data Sharing in Web-Based and Mobile Research: Participant Perspectives in a Large Nationwide Sample of Men Who Have Sex With Men in the United States. J Med Internet Res 2018 Jul 04;20(7):e233 [FREE Full text] [doi: 10.2196/jmir.9019] [Medline: 29973332]
26. Barratt MJ, Potter GR, Wouters M, Wilkins C, Werse B, Perälä J, et al. Lessons from conducting trans-national Internet-mediated participatory research with hidden populations of cannabis cultivators. Int J Drug Policy 2015 Mar;26(3):238-249. [doi: 10.1016/j.drugpo.2014.12.004] [Medline: 25576247]
27. Office for Civil Rights (OCR). Notification of Enforcement Discretion for Telehealth Remote Communications During the COVID-19 Nationwide Public Health Emergency. HHS.gov. 2021 Jan 20. URL: https://www.hhs.gov/hipaa/for-professionals/special-topics/emergency-preparedness/notification-enforcement-discretion-telehealth/index.html# [accessed 2021-03-21]
28. Hunold KM, Goldberg EM, Caterino JM, Hwang U, Platts-Mills TF, Shah MN, Academy for Geriatric Emergency Medicine. Inclusion of older adults in emergency department clinical research: Strategies to achieve a critical goal. Acad Emerg Med 2022 Mar;29(3):376-383. [doi: 10.1111/acem.14386] [Medline: 34582613]
29. Twitter. Twitter Privacy Policy. Twitter. 2022 Jun 10. URL: https://twitter.com/en/privacy#chapter1.4 [accessed 2022-03-24]
30. Zhang S. Facebook Groups as Therapy. The Atlantic. 2018 Oct 26. URL: https://www.theatlantic.com/technology/archive/2018/10/facebook-emotional-support-groups/572941/ [accessed 2022-03-24]
31. Regulatory Foundations, Ethics, and Law Program. The Use of Social Media in Recruitment to Research: A Guide for Investigators and IRBs. Harvard Clinical and Translational Science Center. Boston, Massachussets: The Harvard Clinical and Translational Science Center; 2017. URL: https://catalyst.harvard.edu/wp-content/uploads/regulatory/Social_Media_Guidance.pdf [accessed 2021-11-08]
32. Introduction to Social Media. University of South Florida. 2022. URL: https://www.usf.edu/ucm/marketing/intro-social-media.aspx [accessed 2021-11-08]
33. Flood-Grady E, Solberg LB, Baralt C, Meyer M, Stevens J, Krieger JL. Engaging Institutional Stakeholders to Develop and Implement Guidelines for Recruiting Participants in Research Studies Using Social Media: Mixed Methods, Multi-Phase Process. J Med Internet Res 2021 Oct 08;23(10):e23312 [FREE Full text] [doi: 10.2196/23312] [Medline: 34623319]
34. Social Media Policy. Iona College. 2021. URL: https://www.iona.edu/offices/information-technology/information-and-policies/social-media-policy [accessed 2021-11-08]

XSL•FO
RenderX

35. Toronto Metropolitan University Research Ethics Board. Guidelines for Recruitment of Research Participants. Toronto Metropolitan University. Toronto, ON: Toronto Metropolitan University; 2017 Nov. URL: https://www.torontomu.ca/content/dam/research/documents/ethics/guidelines-for-recruitment-of-research-participants.pdf [accessed 2022-05-31]

Abbreviations

CDH: Center for Digital Health
e-consent: electronic informed consent
IRB: institutional review board

Edited by T Leung; submitted 17.12.21; peer-reviewed by R Davis-Martin, M Thompson, M Katz; comments to author 31.01.22; revised version received 11.04.22; accepted 02.05.22; published 13.06.22.

Please cite as:
Goldberg EM, Rosen RK, Dizon DS, Langdon KJ, Davoodi NM, Wray TB, Nugent NR, Dunsiger SI, Ranney ML
Using Social Media for Clinical Research: Recommendations and Examples From the Brown-Lifespan Center for Digital Health
J Med Internet Res 2022;24(6):e35804
URL: https://www.jmir.org/2022/6/e35804
doi:10.2196/35804
PMID:35700012

©Elizabeth M Goldberg, Rochelle K Rosen, Don S Dizon, Kirsten J Langdon, Natalie M Davoodi, Tyler B Wray, Nicole R Nugent, Shira I Dunsiger, Megan L Ranney. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.06.2022. 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

Triage Errors in Primary and Pre–Primary Care

Hai Nguyen[1,2], PhD; Andras Meczner[1], MD; Krista Burslam-Dawe[3], DipHE Nursing; Benedict Hayhoe[3,4], MD, MRCGP

[1]Your.MD Ltd, London, United Kingdom

[2]Health Services and Population Research, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[3]eConsult Ltd, London, United Kingdom

[4]Department of Primary Care, School of Public Health, Imperial College London, London, United Kingdom

Corresponding Author:
Hai Nguyen, PhD
Your.MD Ltd
5th Floor Lincoln House
296-302 High Holborn
London, WC1V 7JH
United Kingdom
Phone: 44 7847349099
Email: hai@livehealthily.com

Abstract

Triage errors are a major concern in health care due to resulting harmful delays in treatments or inappropriate allocation of resources. With the increasing popularity of digital symptom checkers in pre–primary care settings, and amid claims that artificial intelligence outperforms doctors, the accuracy of triage by digital symptom checkers is ever more scrutinized. This paper examines the context and challenges of triage in primary care, pre–primary care, and emergency care, as well as reviews existing evidence on the prevalence of triage errors in all three settings. Implications for development, research, and practice are highlighted, and recommendations are made on how digital symptom checkers should be best positioned.

(J Med Internet Res 2022;24(6):e37209) doi:10.2196/37209

KEYWORDS

triage errors; pre-primary care; digital symptom checker; primary care; viewpoint; triage; symptom checker; emergency care

Introduction

Across health care settings globally, the inability of supply (health care resources) to meet demand (the need of individuals for health care advice) means significant limitations exist on access to medical assessments and treatments. Safe, effective, and fair distribution of health care resources therefore requires some form of filtering and direction, or triage, of individuals within health care services based on type or severity of symptoms and/or initial likely diagnoses.

Emerging health technologies have the potential to provide answers to this problem, in supporting the initial assessment of individuals presenting with symptoms to ensure that they access the right area of the health system with the appropriate degree of urgency. Digital symptom checkers represent one approach, providing users with triage recommendations based on their presenting symptoms and responses to screening questions. However, the extent to which digital symptom checkers can safely be used alongside or in place of existing forms of initial medical assessment is currently unclear, with the potential significance of error in triage recommendation being substantial.

In this article, we discuss existing evidence on triage errors in pre–primary care (using digital symptom checkers), in comparison with primary care and emergency care, and provide recommendations on how digital symptom checkers might be best positioned to support users and existing health systems.

What Are Triage Errors?

The Oxford English Dictionary defines triage as "the assignment of degrees of urgency of need in order to decide the order of treatment of a large number of injured or ill patients." The sorting of patients into emergency, urgent, nonurgent, and self-care categories becomes essential in all health care settings where there is a need to manage allocation of limited health care resources [1].

Triage errors can be described as either undertriage or overtriage. Undertriage occurs when the level of urgency of an individual's

XSL•FO
RenderX

condition is underestimated [2] and they are allocated to less urgent health services or treatments than they need, potentially resulting in worsening of their condition. Overtriage refers to inappropriate allocation of health care resources to individuals whose health care needs are less significant [2]. This may lead to unnecessary use of scarce health resources and may also have a direct detrimental impact on affected individuals through unneeded (and potentially harmful) investigations or treatments [3,4].

Triage in Pre–Primary Care: Context and Challenges

Triage is likely to take place at many stages of a patient's symptomatic and diagnostic journey, from initial awareness of symptoms through to final established diagnosis and definitive management or resolution of symptoms. Experiencing symptoms is common and frequently does not require medical assessment or treatment [5]. Most individuals will filter and prioritize symptoms that they experience based on factors including personal health beliefs, previous experiences, and informal sources of health information, and seek health care based on the perceived severity of their symptoms/condition, as well as local health system rules, access, and availability.

It has been suggested that the "pre–primary care" health sector, where individuals have reached the stage of considering seeking formal advice on their symptoms but have not yet seen a physician, should be the target of new technological approaches to triage [6]. Building on contemporary interest in self-care, the use of digital technologies to provide detailed and accurate advice and triage provision to support individuals in "self-triage" could enable them to manage their medical problems themselves where possible, or direct them to services of a type and urgency appropriate to their symptoms or condition [7,8].

Digital forms of consultation and triage lack any opportunity for physical examination or for other human interaction, where subtle cues may be picked up. Fully digital consultation systems often lack access to users' medical histories and are entirely dependent on the data entered by users at the time of consultation. These limitations mean that errors are inevitable. Although face-to-face consultation is often viewed as the gold standard of primary care, it is not free from limitations. These might arise from biases and cultural differences between the clinicians and the patients (for instance, some patients may be reluctant to have blood drawn due to their religious beliefs) [9,10]. To consider the acceptability or otherwise of such errors, it is necessary to understand the extent of error in existing health care triage, both through face-to-face and telephone consultations.

Existing Evidence on Triage Errors

Triage Errors in Primary Care

Studies that investigate triage errors in primary care are scarce. A systematic review assessing the safety of telephone triage in out-of-hours care compared with standard face-to-face doctor assessment suggested that triage was safe in 97% (95% CI 96.5%-97.4%) of all patients contacting out-of-hours care and in 89% (95% CI 86.7%-90.2%) of patients with high urgency [11]. This reduced to 46% (95% CI 42.7%-49.8%) when high-risk groups were examined [11]. A triage system in Belgium reported a comparable level of accuracy (98%) when a new French-language algorithm was used [12]. This seems to be consistent with reported rates of triage errors since the 1970s [13]. However, a more recent study in Belgium that compared the triage decisions made by telephone operators and those made by physicians showed a lower level of accuracy [14]. The correctness of the advice given by the operator according to the physicians was 71%, with 12% underestimation of urgency and 17% overestimation [14].

Although some primary care telephone triage is done by doctors, much is done by nurses, sometimes using computer-based clinical decision support systems [15]. A study assessing the safety of telephone triage in general practitioner cooperatives found that triage nurses estimated the level of urgency correctly in 69% of total patients and underestimated the level of urgency in 19% of them [16]. A similar study in the Netherlands reported a comparable rate of triage errors (ie, the level of care was underestimated in 17% of the patients and overestimated in 19%). In Belgium, both the undertriage and overtriage rates were slightly lower, at 10% and 13% of all patients who contacted the out-of-hour telephone service, respectively [17]. In the same study, general practitioners and nurses were found to agree on the level of urgency in 77% of all contacts [17].

Triage Errors in Emergency Care

In emergency department settings, triage error rates appear to be markedly higher. Tam et al [18] found that triage accuracy in a number of multicentered and single-centered studies was only around 60%, with about 23% of cases undertriaged. A similar rate of triage errors was indicated in a US study, where emergency nurse triage accuracy was recorded for 54% of patients with acute myocardial infarction [19]. Better triage accuracy was recorded in a study in South Korea, where retrospective comparison of records of patients admitted to two emergency departments with a gold standard method (based on a 5-level triage scale reviewed by medical experts) [20] found disagreement in 14.7% of the cases (10% overtriage and 5% undertriage). A comparable 17% triage error rate was reported in a study in Brazil using similar methods [21]. Although triage accuracy varied across studies and there is no standardized acceptable triage rate for all patients, the American College of Surgeons has suggested an acceptable rate of undertriage for trauma patients of 5% and 25%-35% for overtriage [18,22]. It is worth noting that relatively high overtriage rates may be seen in emergency care settings where access to rapid imaging or other investigations allows for subsequent "downgrading" of triage.

Triage Errors in Digital Symptom Checkers

The accuracy of digital symptom checkers in providing triage has been met with skepticism. There is limited evidence in this area, but vignette studies have suggested that triage error rates have been shown to be high for digital symptom checkers [23,24]. One study compared 12 publicly available symptom checkers and reported that only 51% of triage decisions for the top 5 diagnoses were correct [23]. However, this is the mean

XSL•FO
RenderX

rate of errors, which may be skewed by a wide range of triage accuracy between the least and most accurate symptom checkers (22%-72%) [23]. The rates of triage errors increase with condition urgency [23,25]. The level of urgency was found to be appropriately assessed in a small proportion of emergency cases with ophthalmic diagnoses (39%, 95% CI 14%-64%) [26]. When applied in emergency department settings, symptom checkers were reported to be inadequately sensitive to emergency cases, with triage accuracy between 45%-75% of total patients [27-30]. However, in a recent study using digital patient self-triage in a hospital emergency department, a digital tool showed higher sensitivity to high-acuity conditions and similar specificity for low-acuity conditions when compared with standard nurse triage using the Manchester Triage System; it also tended to result in overtriage of patients when compared with standard nurse triage [31].

Triage advice provided by symptom checkers is found to be more risk averse than that provided by health care professionals [30,32], with 85% of the users advised to see their doctor in one study [33]. However, in a 5-year follow-up evaluation study, it was observed that symptom checkers in 2020 are less risk averse (odds of 1.11:1, overtriage errors to undertriage errors) than in 2015 (odds of 2.82:1) [24]. Triage errors in emergencies, nevertheless, are still high, with 40% of emergency cases being missed by symptom checkers [24].

Although most studies regarding the accuracy of symptom checkers were carried out through clinical vignettes [23,24], some clinical trials have been conducted to compare the rates of triage error of face-to-face consultation with a physician and digital symptom assessment technologies [34-37]. Results from these clinical trials show that while symptom checkers did not perform as well as face-to-face consultation, correct triage for certain health conditions was still achieved in a higher proportion of patients than expected [37]. Some symptom checkers were reported to attain a sensitivity level of over 50% [36], consistent with previous findings [23].

Evidence of triage error rates in primary care, in emergency care, and by symptom checkers is summarized in Table 1.

Table 1. Triage errors in primary care, in emergency care, and by digital symptom checkers.

	Overtriage	Undertriage
Primary care	13%-19%	10%-19%
Emergency care	10%-35%	5%-23%
Symptom checker	No specific rate of overtriage reported. Mean rate of triage accuracy reported to be around 50%, with a range of 22%-72%	No specific rate of undertriage reported. Mean rate of triage accuracy reported to be around 50%, with a range of 22%-72%

Discussion

Summary

Triage error rates in primary and emergency care vary widely across the literature [38], and differing settings and definitions of triage across settings make comparison difficult. The overall level of accuracy of out-of-hour telephone triage was between 69% and 98%. Undertriage rates ranged from 10%-19% in primary care setting and 5%-23% in emergency setting. Overtriage rates ranged from 13%-19% in primary care setting and 10%-35% in emergency setting. Based on limited evidence, digital symptom checkers have relatively low triage accuracy, with a mean error rate of around 50% [25,30]. However, this is likely skewed by outliers caused by the most and least accurate tools, ranging from 22%-72% [23]. Although the errors tend to be over- rather than undertriage, with users advised to visit a doctor in 85% of cases in one study even when symptoms were appropriate for self-care [33], symptom checkers are increasingly less risk averse [24].

Limitations

It is worth noting that this article is not a formal systematic review, thus no specific strategies or selection criteria were applied to our literature search. This might result in potentially relevant studies being missed, despite our best effort to ensure appropriate studies regarding triage accuracy were included. However, from our consideration of the literature, we observed a high level of heterogeneity among the rates of triage errors across studies. The heterogeneity of triage error rates in primary care, pre–primary care, and emergency care is attributable to a number of factors. Most importantly, case mix and approach to/purpose of triage differ substantially across these settings. The number and type of conditions considered in each study also differed. Although the majority of studies included a mix of acute and chronic conditions, some only considered one type of disease (eg, chronic mental health disorders). Studies that assessed triage accuracy in more conditions were more likely to report higher error rates. In addition, the methods used to identify triage errors were heterogeneous. Eight methods were commonly employed in assessing triage accuracy, namely autopsies, patient and provider surveys, standardized patients, second reviews, diagnostic testing audit, malpractice claims, case reviews, and voluntary reports [3]. Studies that used different methods were found to report significantly different rates of errors [3]. Finally, there appeared to be a lack of clarity in the definition and comparison of triage errors. Some studies did not specify whether the triage errors were overtriage or undertriage. This lack of clarity and consistency means it is not possible to draw conclusions or make clear recommendations as to an acceptable error rate for symptom checkers.

Implications for Development and Practice

Consideration of triage error in primary care is particularly timely in the current unprecedented public health context. The recent COVID-19 pandemic has challenged the ability of health systems worldwide to meet demand, with services in some countries completely overwhelmed. A pressing need to avoid all but the most urgent and essential health service use, and to limit face-to-face interaction between health care professionals and members of the public to an absolute minimum, has led to

XSL•FO
RenderX

the adoption of a "remote total triage" system in primary care using telephone and online consulting in many countries [39].

Whether the digital symptom checkers' level of performance for triage is acceptable depends on the purposes for which they are used [25]. If symptom checkers are seen as a replacement for seeing physicians, they would currently be an inferior alternative [25]. However, if used by individuals to gather quick and accessible information about particular conditions, they are likely to be superior to self-directed internet searches using online search engines [25]. This is especially appropriate when only the best-performing symptom checkers with low triage error rates are used. It is also worth noting that artificial intelligence technology is constantly improving, potentially making it possible for triage made by digital symptom checkers to become more accurate and thus become a safe and useful addition to traditional face-to-face consultations.

Although seeking to avoid unnecessary burden on health services, the lack of available background information and inability to include information from physical examinations or nonverbal cues means that any remote assessment system will likely need to take a risk averse approach to triage. Thus, it is arguably appropriate that digital triage tools adopt this approach.

Implications for Research and Development

Most studies assessing the triage error rates among symptom checkers are conducted through clinical vignettes. The preparation and evaluation of vignettes need to be standardized to allow for external validity and comparability. Furthermore, clinical trials where symptom checkers' rates of triage error are compared with those of face-to-face consultation should be encouraged. This method not only enables the assessment of triage accuracy but also allows the examination of users' compliance with triage advice and possible benefits for the health care system.

There is little evidence on users' compliance with triage advice, in either traditional forms of triage or that given by digital symptom checkers, nor is there data on consequences of symptom checker errors. Additionally, little is currently known about patient expectations and health beliefs in relation to digital diagnostic and triage tools. It seems likely that most individuals would place lower weighting on the advice of a symptom checker than a human clinician, and use the information provided by these tools as part of their decision-making process. However, in times of increasing reliance on digital technology, it is possible that some individuals may have greater trust in these tools than might be expected. Research is clearly needed to clarify these questions, but developers should assume a relatively high degree of reliance of users on the recommendations that symptom checkers provide.

In addition, well-conducted research to understand the clinical effectiveness of digital symptom checkers in effectively triaging individuals (ie, offering appropriate self-care advice or assigning to appropriate services) is urgently needed. To inform decisions of users and policy makers adequately, this must incorporate formal comparison with existing provision, with a focus on primary care telephone and online triage.

Recommendations

1. Digital symptom checkers should largely position themselves in the pre–primary care triage/self-care area—evidence does not currently support the ability of artificial intelligence to provide effective consultations at the level of those that would normally take place in traditional face-to-face primary care or emergency department setting.
2. Digital symptom checkers can be appropriately promoted as a safer alternative/effective addition to existing sources of information (such as self-directed online searches) for individuals prior to seeking formal health advice.
3. Providers of digital symptom checkers should seek to ensure that triage error rates fall within the lower thresholds demonstrated in existing evidence of those in primary care telephone triage (acknowledging the substantial limitations of this literature).
4. Developers of symptom checkers should make efforts to expand the evidence base in this area, including establishing systems to gain user feedback on triage accuracy/appropriateness, as well as engaging with academic partners to carry out formal research. Findings in terms of limitations and error rates should be clearly publicized and highlighted to users.
5. Current methods employed to study symptom checkers' triage accuracy such as case vignette studies should be standardized to allow for external validity and comparability. Clinical trials where key outcome measures include the accuracy of both outcome conditions and triage should also be conducted. A clear distinction between over- and undertriage should be made to provide data for safety monitoring and economic evaluation.

Conclusion

There is very limited evidence and no clear gold standard comparison for triage errors in digital symptom checkers, meaning that it is not possible to make recommendations on an acceptable error rate. Positioning symptom checkers in the self-care/pre–primary care triage setting therefore seems to be most appropriate and where they can likely add value for individuals experiencing symptoms. Industry and academics should work together to develop the necessary evidence, and efforts should be made to collect user feedback and outcomes data. Until clearer comparisons with existing care are available, digital symptom checkers and triage tools should appropriately continue to take a risk averse approach in the recommendations they give to users.

XSL•FO
RenderX

Acknowledgments

BH is grateful for the support of the National Institute for Health and Care Research (NIHR) under the Applied Research Collaborations programme for North West London. The views expressed in this publication are those of the authors and not necessarily those of the National Health Service (NHS), the NIHR, or the Department of Health.

Authors' Contributions

HN and BH conceptualized and wrote the paper. AM provided insights into the accuracy of digital symptom checkers and contributed to writing the paper. KBD provided further insights in relation to emergency department triage. All authors read and approved the final manuscript.

Conflicts of Interest

HN works as an epidemiologist at Your.MD Ltd. AM works as VP Clinical Operations at Your.MD Ltd. BH is a general practitioner working in the National Health Service (NHS), and works for eConsult Health Ltd, a provider of asynchronous consultations for NHS primary, secondary, and urgent/emergency care. KBD is a registered general nurse working in the NHS and also works for eConsult Health Ltd as a Clinical Director for Urgent and Emergency Care.

References

1. Iserson KV, Moskop JC. Triage in medicine, part I: Concept, history, and types. Ann Emerg Med 2007 Mar;49(3):275-281. [doi: 10.1016/j.annemergmed.2006.05.019] [Medline: 17141139]
2. Cook CH, Muscarella P, Praba AC, Melvin WS, Martin LC. Reducing overtriage without compromising outcomes in trauma patients. Arch Surg 2001 Jul 01;136(7):752-756. [doi: 10.1001/archsurg.136.7.752] [Medline: 11448384]
3. Graber ML. The incidence of diagnostic error in medicine. BMJ Qual Saf 2013 Oct 15;22 Suppl 2(Suppl 2):ii21-ii27 [FREE Full text] [doi: 10.1136/bmjqs-2012-001615] [Medline: 23771902]
4. Hildebrandt DE, Westfall JM, Fernald DH, Pace WD. Harm resulting from inappropriate telephone triage in primary care. J Am Board Fam Med 2006 Sep 01;19(5):437-442 [FREE Full text] [doi: 10.3122/jabfm.19.5.437] [Medline: 16951292]
5. Elnegaard S, Andersen RS, Pedersen AF, Larsen PV, Søndergaard J, Rasmussen S, et al. Self-reported symptoms and healthcare seeking in the general population--exploring "The Symptom Iceberg". BMC Public Health 2015 Jul 21;15(1):685 [FREE Full text] [doi: 10.1186/s12889-015-2034-5] [Medline: 26195232]
6. Carr-Brown J, Berlucchi M. Pre-Primary Care: An Untapped Global Health Opportunity. Your.MD. 2016. URL: https://assets.ctfassets.net/iqo3fk8od6t9/3NZzZ8F7tUi9QPCljAd1YH/43f0e8e262a35f7715ec4b96d1257312/report.pdf [accessed 2022-05-27]
7. Oliver D. David Oliver: Why force GP streaming on NHS emergency departments? BMJ 2020 Mar 18;368:m992. [doi: 10.1136/bmj.m992] [Medline: 32188596]
8. El-Osta A, Webber D, Gnani S, Banarsee R, Mummery D, Majeed A, et al. The Self-Care Matrix: A unifying framework for self-care. Self Care 2019;10(3):38-56 [FREE Full text]
9. FitzGerald C, Hurst S. Implicit bias in healthcare professionals: a systematic review. BMC Med Ethics 2017 Mar 01;18(1):19 [FREE Full text] [doi: 10.1186/s12910-017-0179-8] [Medline: 28249596]
10. Johnson RL, Saha S, Arbelaez JJ, Beach MC, Cooper LA. Racial and ethnic differences in patient perceptions of bias and cultural competence in health care. J Gen Intern Med 2004 Feb;19(2):101-110 [FREE Full text] [doi: 10.1111/j.1525-1497.2004.30262.x] [Medline: 15009789]
11. Huibers L, Smits M, Renaud V, Giesen P, Wensing M. Safety of telephone triage in out-of-hours care: a systematic review. Scand J Prim Health Care 2011 Dec 29;29(4):198-209 [FREE Full text] [doi: 10.3109/02813432.2011.629150] [Medline: 22126218]
12. Brasseur E, Servotte J, Donneau A, Stipulante S, d'Orio V, Ghuysen A. Triage for out-of-hours primary care calls: a reliability study of a new French-language algorithm, the SALOMON rule. Scand J Prim Health Care 2019 Jun 29;37(2):227-232 [FREE Full text] [doi: 10.1080/02813432.2019.1608057] [Medline: 31033368]
13. Albin SL, Wassertheil-Smoller S, Jacobson S, Bell B. Evaluation of emergency room triage performed by nurses. Am J Public Health 1975 Oct;65(10):1063-1068. [doi: 10.2105/ajph.65.10.1063] [Medline: 1163704]
14. Morreel S, Colliers A, Remmen R, Verhoeven V, Philips H. How accurate is telephone triage in out-of-hours care? An observational trial in real patients. Acta Clin Belg 2022 Apr 30;77(2):301-306. [doi: 10.1080/17843286.2020.1839719] [Medline: 33124524]
15. Bunn F, Byrne G, Kendall S. The effects of telephone consultation and triage on healthcare use and patient satisfaction: a systematic review. Br J Gen Pract 2005 Dec;55(521):956-961 [FREE Full text] [Medline: 16378566]
16. Graversen DS, Christensen MB, Pedersen AF, Carlsen AH, Bro F, Christensen HC, et al. Safety, efficiency and health-related quality of telephone triage conducted by general practitioners, nurses, or physicians in out-of-hours primary care: a quasi-experimental study using the Assessment of Quality in Telephone Triage (AQTT) to assess audio-recorded telephone calls. BMC Fam Pract 2020 May 09;21(1):84 [FREE Full text] [doi: 10.1186/s12875-020-01122-z] [Medline: 32386511]

XSL•FO
RenderX

17. Philips H, Van Bergen J, Huibers L, Colliers A, Bartholomeeusen S, Coenen S, et al. Agreement on urgency assessment between secretaries and general practitioners: an observational study in out-of-hours general practice service in Belgium. Acta Clin Belg 2015 Oct;70(5):309-314. [doi: 10.1179/2295333715Y.0000000017] [Medline: 25819448]
18. Tam HL, Chung SF, Lou CK. A review of triage accuracy and future direction. BMC Emerg Med 2018 Dec 20;18(1):58 [FREE Full text] [doi: 10.1186/s12873-018-0215-0] [Medline: 30572841]
19. Sanders SF, DeVon HA. Accuracy in ED triage for symptoms of acute myocardial infarction. J Emerg Nurs 2016 Jul;42(4):331-337. [doi: 10.1016/j.jen.2015.12.011] [Medline: 26953510]
20. Moon S, Shim JL, Park K, Park C. Triage accuracy and causes of mistriage using the Korean Triage and Acuity Scale. PLoS One 2019 Sep 6;14(9):e0216972 [FREE Full text] [doi: 10.1371/journal.pone.0216972] [Medline: 31490937]
21. Hinson JS, Martinez DA, Schmitz PSK, Toerper M, Radu D, Scheulen J, et al. Accuracy of emergency department triage using the Emergency Severity Index and independent predictors of under-triage and over-triage in Brazil: a retrospective cohort analysis. Int J Emerg Med 2018 Jan 15;11(1):3 [FREE Full text] [doi: 10.1186/s12245-017-0161-8] [Medline: 29335793]
22. Zachariasse JM, van der Hagen V, Seiger N, Mackway-Jones K, van Veen M, Moll HA. Performance of triage systems in emergency care: a systematic review and meta-analysis. BMJ Open 2019 May 28;9(5):e026471 [FREE Full text] [doi: 10.1136/bmjopen-2018-026471] [Medline: 31142524]
23. 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]
24. 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 May 10;24(5):e31810 [FREE Full text] [doi: 10.2196/31810] [Medline: 35536633]
25. 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]
26. Shen C, Nguyen M, Gregor A, Isaza G, Beattie A. Accuracy of a popular online symptom checker for ophthalmic diagnoses. JAMA Ophthalmol 2019 Jun 01;137(6):690-692 [FREE Full text] [doi: 10.1001/jamaophthalmol.2019.0571] [Medline: 30973602]
27. Yu SWY, Ma A, Tsang VHM, Chung LSW, Leung S, Leung L. Triage accuracy of online symptom checkers for Accident and Emergency Department patients. Hong Kong Journal of Emergency Medicine 2019 Apr 16;27(4):217-222. [doi: 10.1177/1024907919842486]
28. Luger TM, Houston TK, Suls J. Older adult experience of online diagnosis: results from a scenario-based think-aloud protocol. J Med Internet Res 2014 Jan 16;16(1):e16 [FREE Full text] [doi: 10.2196/jmir.2924] [Medline: 24434479]
29. Bisson LJ, Komm JT, Bernas GA, Fineberg MS, Marzo JM, Rauh MA, et al. Accuracy of a computer-based diagnostic program for ambulatory patients with knee pain. Am J Sports Med 2014 Oct 29;42(10):2371-2376. [doi: 10.1177/0363546514541654] [Medline: 25073597]
30. 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]
31. Dickson SJ, Dewar C, Richardson A, Hunter A, Searle S, Hodgson LE. Agreement and validity of electronic patient self-triage (eTriage) with nurse triage in two UK emergency departments: a retrospective study. Eur J Emerg Med 2022 Feb 01;29(1):49-55. [doi: 10.1097/MEJ.0000000000000863] [Medline: 34545027]
32. 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]
33. Nijland N, Cranen K, Boer H, van Gemert-Pijnen JEWC, Seydel ER. Patient use and compliance with medical advice delivered by a web-based triage system in primary care. J Telemed Telecare 2010 Jan 19;16(1):8-11. [doi: 10.1258/jtt.2009.001004] [Medline: 20086260]
34. Martin SS, Quaye E, Schultz S, Fashanu OE, Wang J, Saheed MO, Prem Ramaswami, et al. A randomized controlled trial of online symptom searching to inform patient generated differential diagnoses. NPJ Digit Med 2019 Nov 11;2(1):110 [FREE Full text] [doi: 10.1038/s41746-019-0183-0] [Medline: 31728417]
35. Knitza J, Muehlensiepen F, Ignatyev Y, Fuchs F, Mohn J, Simon D, et al. Patient's perception of digital symptom assessment technologies in rheumatology: results from a multicentre study. Front Public Health 2022 Feb 22;10:844669 [FREE Full text] [doi: 10.3389/fpubh.2022.844669] [Medline: 35273944]
36. 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]
37. Proft F, Spiller L, Redeker I, Protopopov M, Rodriguez VR, Muche B, et al. Comparison of an online self-referral tool with a physician-based referral strategy for early recognition of patients with a high probability of axial spa. Semin Arthritis Rheum 2020 Oct;50(5):1015-1021. [doi: 10.1016/j.semarthrit.2020.07.018] [Medline: 32911279]

XSL•FO
RenderX

38. Schiff GD, Kim S, Abrams R, Cosby K, Lambert B, Elstein AS, et al. Diagnosing diagnosis errors: lessons from a multi-institutional collaborative project. In: Henriksen K, Battles JB, Marks ES, Lewin DI, editors. Advances in Patient Safety: From Research to Implementation (Volume 2: Concepts and Methodology). Rockville, MD: Agency for Healthcare Research and Quality (US); 2005.
39. Advice on how to establish a remote 'total triage' model in general practice using online consultations. National Health Service. 2020. URL: https://www.england.nhs.uk/coronavirus/documents/advice-on-how-to-establish-a-remote-total-triage-model-in-general-practice-using-online-consultations/ [accessed 2022-05-27]

Abbreviations

NHS: National Health Service
NIHR: National Institute for Health and Care Research

Edited by R Kukafka; submitted 10.02.22; peer-reviewed by J Knitza, M Kopka; comments to author 08.03.22; revised version received 21.03.22; accepted 04.04.22; published 24.06.22.

Please cite as:
Nguyen H, Meczner A, Burslam-Dawe K, Hayhoe B
Triage Errors in Primary and Pre–Primary Care
J Med Internet Res 2022;24(6):e37209
URL: https://www.jmir.org/2022/6/e37209
doi:10.2196/37209
PMID:35749166

©Hai Nguyen, Andras Meczner, Krista Burslam-Dawe, Benedict Hayhoe. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.06.2022. 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

What are Digital Public Health Interventions? First Steps Toward a Definition and an Intervention Classification Framework

Julian Wienert[1,2,3*], BA, MA, PhD; Tina Jahnel[1,4], BA, MA, PhD; Laura Maaß[1,5*], BA, MA

[1]Research Cluster Evaluation, Leibniz ScienceCampus Bremen Digital Public Health, Bremen, Germany

[2]Social Epidemiology, Prevention and Evaluation, Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany

[3]Health Psychology, Social Sciences, IU International University for Applied Sciences, Bad Reichenhall, Germany

[4]Department of Health Services Research, Institute for Public Health and Nursing, University of Bremen, Bremen, Germany

[5]Health, Long-Term Care and Pensions, Research Center on Inequality and Social Policy (SOCIUM), Bremen, Germany

[*]these authors contributed equally

Corresponding Author:
Julian Wienert, BA, MA, PhD
Health Psychology
Social Sciences
IU International University for Applied Sciences
Kaiserplatz 1
Bad Reichenhall, 83435
Germany
Phone: 49 1725858645
Email: julian.wienert@iu.org

Abstract

Digital public health is an emerging field in population-based research and practice. The fast development of digital technologies provides a fundamentally new understanding of improving public health by using digitalization, especially in prevention and health promotion. The first step toward a better understanding of digital public health is to conceptualize the subject of the assessment by defining what digital public health interventions are. This is important, as one cannot evaluate tools if one does not know what precisely an intervention in this field can be. Therefore, this study aims to provide the first definition of digital public health interventions. We will merge leading models for public health functions by the World Health Organization, a framework for digital health technologies by the National Institute for Health and Care Excellence, and a user-centered approach to intervention development. Together, they provide an overview of the functions and areas of use for digital public health interventions. Nevertheless, one must keep in mind that public health functions can differ among different health care systems, limiting our new framework's universal validity. We conclude that a digital public health intervention should address essential public health functions through digital means. Furthermore, it should include members of the target group in the development process to improve social acceptance and achieve a population health impact.

(J Med Internet Res 2022;24(6):e31921) doi:10.2196/31921

KEYWORDS

digital health; digital Public Health; digital public health interventions; digital health technologies; mHealth; eHealth; participatory approach; framework; mobile phone

Introduction

Background

As digitization plays a large role in an increasing number of health systems, digital public health is an emerging field for population-based research and practice. The fast development of both hardware- and software-based digital technologies provides a fundamentally new understanding of improving public health, which can be achieved through digitalization, especially in prevention and health promotion. For example, digital technologies may improve physical activity levels, dietary intake, posture, and mental well-being via sensors and apps [1]. Technological innovations in apps for tracking health-related behavior, monitoring potential health risks, and communication and interaction have rapidly changed many aspects of public health [2]. However, not all of these interventions might achieve a health impact at a population level by displaying effectiveness

XSL•FO
RenderX

in randomized clinical trials and efficacy under quasi-experimental real-world circumstances.

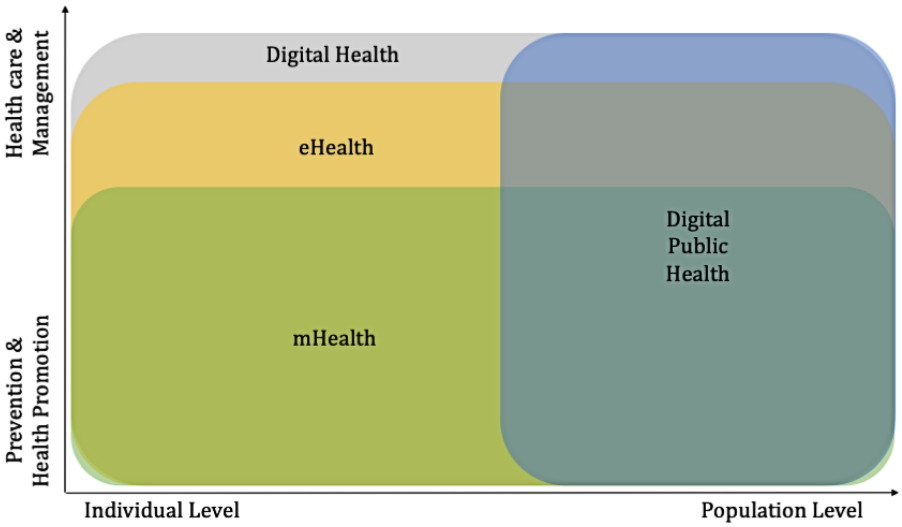
Although there is a need for evaluation methods that address the many challenges that arise with digitization (eg, fast-paced development), it is challenging to assess digital public health interventions as these may span from population health surveillance to the prevention of specific diseases, and they develop faster than analog interventions [3]. Moreover, companies and institutions often develop digital tools based on market evaluations, expected profits, and technological possibilities but not based on the public's needs and preferences. To improve digital public health interventions' effectiveness and efficacy, we first need to understand what they entail and how they are defined. However, to our knowledge, no definition for digital public health interventions exists to date. Only by doing that will we gather meaningful, valid, and reliable results on their effectiveness and efficiency. Thus, the aim of this viewpoint is to offer a definition for digital public health interventions.

Before defining digital public health interventions, we need to explain the differences between eHealth, mobile health (mHealth), digital health, and digital public health. This is necessary to highlight the differences between digital health interventions (DHIs) and digital public health interventions. Following this, we will build a framework that might help to identify, structure, and classify digital public health interventions. Our definition and framework will rely on existing approaches from public health [4], digital health [5], and user-centered design [6]. The first section of building the framework will explain why we chose the selected models; we decided to use the Essential Public Health Functions (EPHFs) by the World Health Organization (WHO) for the public health level [4], the updated version of the *Evidence Standards Framework for Digital Health Technologies* by the National Institute for Health and Care Excellence (NICE) [5], and the *Participatory Health Research* approach by Wright [6]. After clarifying the reasons for choosing named approaches, we will explain each model and how they are related to digital public health and may be used for digital public health purposes. After setting the theoretical background for our definition, we will illustrate our findings using the German Corona App as an example to validate our digital public health intervention criteria. We will conclude with a definition for digital public health interventions and use this to propose a digital public health intervention classification framework (Multimedia Appendix 1).

Differences Among eHealth, mHealth, Digital Health, and Digital Public Health

Terms such as eHealth, mHealth, or digital health are used in the context of the digitization of public health. Since 2019, few papers have also referred to the term digital public health. Given the multitude of terms and definitions in digital health, it is essential to understand the considerable heterogeneity of how such terms interrelate with each other and where digital public health might find its place in the terminological canon of digital health. Therefore, the following section will define the named terms and summarize their core fields of action and the target group's level, as seen in Figure 1.

Figure 1. Core field of action and target group level of mHealth, eHealth, digital health, and digital public health [7-15]. mHealth: mobile health.



An article on *eHealth* concepts based on an extensive literature search [7] confirms the lack of consensus on the meaning of *eHealth* as possibly the first word in this field. A 2005 study found 51 different definitions of *eHealth* [8]. This lack of consensus highlights the importance of a shared understanding of terms. More recent studies emphasize that because of its immense dynamics, the field of *eHealth* is challenging to define [9]. Most definitions share the use of information and communication technologies (eg, the internet) for health topics. Their focus mostly lies on delivering health services rather than health promotion and disease prevention [10-12]. Some definitions also highlight the importance of user-centered approaches for facilitating health services in *eHealth* [10,12]. The word *mHealth* aims more directly at a particular technology, namely smartphones and mobile sensors, in their health significance. Thus, *mHealth* is defined more precisely overall, although different technologies are used here [13]. As a part of *eHealth*, the focus of *mHealth* lies on wireless and mobile technologies and their use in enhancing health-related science, treatment, and ultimately health status [9].

XSL•FO
RenderX

Fatehi et al [14] stated that in 2020, there were >90 different definitions for *digital health*. They concluded that *digital health* includes eHealth, mHealth, self-tracking, wearable devices, artificial intelligence, and information systems in health care, focusing on health and not technology. *Digital health* focuses on the health of individuals (eg, patients) to improve health care with technology [14]. This is where *digital health* and digital public health differ, as digital public health aims to improve health and well-being at the population level. Nevertheless, digital public health uses the same technologies that are also used to improve individual health care; however, its purposes change. A recent publication by Zeeb et al [15] provides the first overview of what digital public health might be. It serves as a starting point for developing a better understanding of digital public health interventions by providing a short introduction to central terms. Thus, following the article by Zeeb et al [15], the authors propose the following definition for digital public health in distinction to the other abovementioned fields and outline for which fields they see it relevant (own translation from German):

> *DiPH [...] focuses on the development, application, and knowledge interest on Public Health and thus on prevention, health promotion, and the related basic sciences such as epidemiology. Primary clinical and individual patient-related aspects are not in the foreground, unlike, for example, telemedicine with its concrete application in an individual treatment and care context. However, it should be noted that the term DiPH has not yet prevailed over others such as eHealth and mHealth. Also, this is hard to expect given the diversity and dynamics of the terms used to date. Where, however, the focus of digitization and health is on population, prevention, and health promotion, including a conscious analysis of health inequalities, DiPH can offer a clearer classification than some other terms in this field.*

Although the definition by Zeeb et al [15] serves as a good starting point for the discussion, it mainly focuses on the primary level of prevention (ie, preventing a disease or injury before it occurs). Although public health, in its essence, comprises 3 levels of prevention, according to this definition, secondary (eg, reducing the impact of a disease or injury after it occurred) and tertiary preventions (eg, rehabilitation) are not explicitly mentioned by Zeeb et al [15] as part of digital public health. The central challenge of defining digital public health is the integration of digital development and technologies into public health concepts and use them to achieve public health goals rather than redefining and reconceptualizing public health in the face of technological advancements [5,15].

To develop a first working definition and classification framework for digital public health interventions, established models for both digital health [5] and public health [4] were assessed to combine aspects of these models to develop a more holistic definition of digital public health interventions. Here we suggest that digital public health, as a complex intervention, should be viewed from different perspectives. As such, our proposed classification is a combination of the elements of already existing models. Specifically, the EPHF by the WHO provided us with an overview of the necessary core functions of public health, which might also be addressed by digital means. Our definition will explicitly include the area of health promotion (ie, focusing on health resources) and all three levels of prevention (ie, primary prevention to reduce the risk of disease development, secondary prevention as screening and early diagnosis, and tertiary prevention for rehabilitation), as all these levels are included in the EPHF by the WHO [4]. We will then link our definition to the participation approach: a user-centered model for the development of digital interventions to increase the acceptance of digital public health interventions among target groups [6]. This will create a framework that follows the concepts of public health (goals).

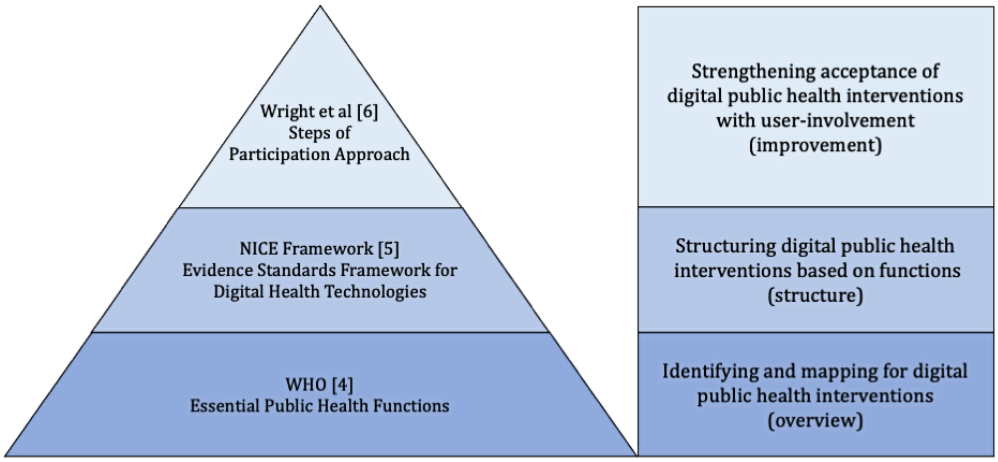
Choice of Included Models

Following a narrative approach and based on the authors' expertise and experience in the field of Public Health, three layers were identified: (1) overview, which is a larger operational layer where the central functions of digital public health are mapped; (2) structure, which is a layer that focuses on structuring digital public health activities (eg, by functions); and (3) improvement, which is a layer that specifically includes the individual perspective in the development and use of digital public health interventions. For each layer, a framework was identified based on the author's expertise and previous experience with the frameworks and a nonsystematic literature search for alternative frameworks. The frameworks included are as follows: the EPHF by the WHO, which offers a macro view of public health topics [4]; the *Evidence Standards Framework for Digital Health Technologies* by the NICE, which categorizes DHIs for the UK setting [5]; and the *Steps of Participation Approach*, which was suggested by Wright [6].

Together, the EPHF and the NICE framework will build the base of mapping and structuring digital public health interventions. We then use the *Steps of Participation Approach* suggested by Wright [6]. This is a user-centered approach for intervention development that aims to increase acceptance within the target group. The approach provides well-described and clear-cut categories for target group involvement—participation and nonparticipation alike. Together, with all 3 models aligned, a conceptual pyramid for digital public health intervention classification is formed (Figure 2).

XSL•FO
RenderX

Figure 2. Conceptual pyramid for a framework of digital public health interventions [4-6]. NICE: National Institute for Health and Care Excellence; WHO: World Health Organization.



EPHFs by the WHO

A way of addressing public health goals to affect population health is using *the EPHF* [4]. Following the WHO report *EPHFs, health systems and health security: developing conceptual clarity and a WHO road map for action,* these functions can be separated into cross-cutting, horizontal functions, roughly based on the building blocks approach to health systems, and service-based, vertical functions comprising the traditional public health services provided by modern health systems [4]. Although there is no precise definition for each part, as they depend on each health care system or region, the WHO report identified some significant categories that most EPHF share (Textbox 1).

Textbox 1. Essential public health functions according to the World Health Organization [4].

Essential public health functions

Horizontal functions

- Governance (eg, public health management, policy, and planning or quality assurance in health services)
- Financing: establishment of sustainable organizational structures, institutional capacity, and policy making
- Human resources: development and management of human resources
- Health information systems: population health surveillance and monitoring
- Research: development of a national public health research agenda, allocation of resources for research, integration of research activities into public health, capacity building for innovation, and dissemination to translate research findings into policy and practice
- Social participation and health communication: social participation, community partnership, community engagement and/or (digital) public health communication, and design public health services around people's needs

Vertical functions

- Health protection: regulations and legal protections (for workers, patients, consumers, and the environment)
- Health promotion: community and social participation, intersectoral collaboration, measures to address behavioral risk factors (tobacco, alcohol, diet, and physical activity), and the social determinants of health and health education
- Disease prevention: services provided within the health care system and targeting communicable diseases
- Health care: With specific functions for quality assurance and access, universal health coverage is a defining feature depending on the World Health Organization region or country (ie, the European region emphatically excludes most health care services from the public health remit because of strong roots in the principles of universal access as opposed to the United States or Western Pacific regions).
- Preparedness for public health emergencies: encompass any sudden, large-scale, negative impact on public health arising from outbreaks, natural disasters, severe weather events, migratory flows, accidents, terrorism, or other environmental or human causes
- Other vertical functions: A wide variety of specific vertical functions are given importance in different countries. In part, this reflects an overall approach to developing frameworks that list essential services rather than broader functions per se. At the same time, the vertical positions were chosen to reflect national priorities.

These are just a few examples of fields for action in both (analog) public health interventions and digital public health interventions. However, we could not apply all EPHF to every setting. They depend strongly on specific health care systems, which differ among countries. In general, the WHO regards public health interventions primarily as an effort or policy that attempts to improve mental, social, and physical health at the population level by including and addressing EPHF [4].

XSL•FO
RenderX

Analogous to public health interventions, digital public health interventions have the potential to include and address horizontal as well as vertical functions. The governmental regulation of mHealth apps as medical devices with the possibility of reimbursement, as started in Germany in December 2020 [16], may be one way of applying horizontal EPHF of governance to digital public health. Various countries have also developed proximity-tracing apps as tools for population health surveillance and monitoring during the SARS-CoV-2 pandemic [17,18]. The last example of applying the horizontal EPHF to digital public health is the digitization of health care systems in total. This leads to a redesign around people's needs and expectations in, for instance, web-based consultation services or telemedicine for people in rural areas who do not have access to health care professionals [19]. As for vertical public health functions, apps and wearables for self-monitoring, step counting, and fitness tracking can serve as examples of vertical digital public health functions. Their goal is to promote health and a healthy lifestyle [20,21].

Level of Interaction: NICE Framework

Applying EPHF as a cornerstone for the identification and mapping of digital public health interventions provides an initial overview of the field of digital public health. The next step is to further structure such interventions based on their functional classification proposed by NICE's *Evidence Standards Framework for Digital Health Technologies* [5]. This applied framework describes the types and levels of evidence needed to show the effectiveness and expected economic impact of a DHI. Various publications have used this framework for their digital health technology assessment, which confirmed our resolution that this framework is not only well-known but also well-used in the scientific field of digital health [22-24]. The NICE framework aims to establish standardized criteria that can assess DHIs by providing a functional classification and stratification into evidence tiers. This separation illustrates the main functions of the types of interventions that we expect to be the most widely developed (Textbox 2).

Textbox 2. Functional classification and stratification into evidence tiers [5].

Stratification into evidence tiers and functional classification

Evidence tier C: interventions

- Preventive behavior change: address public health issues (eg, smoking, eating, alcohol, sexual health, sleeping, and exercise)
- Self-manage: allows people to self-manage a specific condition; may include behavior change techniques
- Treat: provides treatment and guides treatment
- Active monitoring: tracking patient location, using wearables to measure, record, and/or transmit data about a specific condition
- Calculate: a calculator that affects treatment, diagnosis, or care
- Diagnose: diagnose a specific condition; guides diagnosis

Evidence tier B: understanding and communicating

- Inform: provides information (about a condition or general health and lifestyle), resources, or activities to the public, patients, or clinicians
- Simple monitoring: includes general health monitoring using fitness wearables and simple symptom diaries
- Communicate: allows 2-way communication among citizens, patients, or health care professionals

Evidence tier A: system impact

- System service: digital health interventions with no measurable patient outcomes but which provide services to the health and social care system

The abovementioned evidence tiers serve as concrete examples in digital health for the EPHF. For instance, level A refers to the vertical EPHF *health care* (digital public health tools in this field could be electronic health records). In contrast, the 3 functions in evidence tier B belong to the horizontal public health function *social participation and health communication*. Recent examples for tier B are proximity-tracing apps (sometimes called contact-tracing apps), which various countries use in epidemic or pandemic outbreaks such as SARS-CoV-2. Such apps usually display level 2 functions (ie, informing, simple monitoring, and communication), which mirrors the underlying EPHF, including disease prevention and health information systems as underlying EPHF. The first 3 functions within evidence tier C serve as a digital example for the vertical functions of *health promotion* as well as *disease prevention* (such as mobile apps on prescription [16]). Finally, the last 3 functions in tier C, although focusing more on the medical and individual level than the other tiers and functions, can be seen as a part of *health promotion* and *disease prevention*. Unlike the first 3 functions in tier C, which focus more on the primary prevention area, the last 3 functions are more closely linked to secondary and tertiary prevention. Specifically, the functions of DHIs in tier C include the early diagnosis of specific conditions and rehabilitation and healing, which improves the user's health (for instance, a national telemedicine service [19]).

As seen, there is an interrelation between the NICE framework and EPHF, supporting the argument that digital public health interventions can address EPHF. The critical part here is that the NICE framework, unlike the WHO EPHF, provides a structure for the degree of complexity (ie, level of interaction) based on the user's risk. Following the understanding of complex and multicomponent interventions that act and interact on different levels, benefits, and acceptance of digital public

XSL•FO
RenderX

health interventions, depending on the users of such interventions and their specific perspectives. Any digital public health intervention can ultimately fail if the population does not accept or use it. Thus, it is essential to involve target groups in the development of these interventions. We propose a participatory and user-centered approach for intervention development as the third cornerstone of digital public health interventions.

User-Centered Approach in Intervention Development

Hochmuth et al [25] advocated that complex and multicomponent interventions require a user-oriented intervention design because of the varying intricacies of such interventions. This intricacy can be based on the following:

- Interactions between technological components (eg, sensors for data acquisition)
- Different requirements for users in the implementation of the intervention (eg, knowledge of data security)
- Involvement of other groups or organizational levels (eg, patients or researchers)
- Degree of adaptation or flexibility of the intervention (eg, further agile development through software updates) [3,25].

To follow a user-centered approach, developers must integrate the users (ie, the target group) in the development process. A way of structuring the involvement of users is the *Steps of Participation Approach* suggested by Wright [6]. This model describes the user's nonparticipation and involvement in the research process. It further differentiates among 9 stages, ranging from instrumentalization to self-organization. The 9 stages provide a hierarchical order not only for participation but also for the nonparticipation of target groups in the development of public health interventions. Although it includes 9 stages, only the last 4 include real participation, according to Wright [6], as the first 2 have no target group members involved in the development process. Steps 3 to 5 are the precursors for participation. As stated by this approach, one can only speak of participation in only those areas where people have the power to participate in the decision-making processes [26]. The 9 significant steps based on Wright [6] are described in the following sections. As shown in Table 1, the difference among the 3 groups of nonparticipation is that the first 2 steps completely exclude the target group. Although grades 3 to 5 recognize the target group as advisers, they do not include them in the decision-making process, which occurs in steps 6 to 9 (Table 1). The chance to successfully roll out and implement a digital public health intervention increases as the development process includes the target group [27]. Therefore, a user-oriented way of developing digital public health interventions to increase acceptance and use of such interventions should be a goal of digital public health.

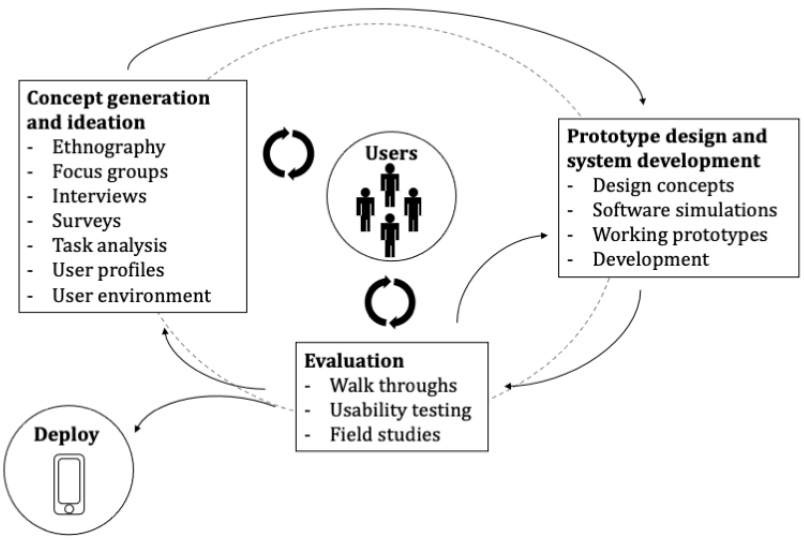
A way of including target groups in the development of new digital public health interventions could be to apply user-centered approaches to the development process [28]. The aim here is to look at issues from various stakeholder perspectives and create new ideas in an interdisciplinary team to solve potential problems and challenges throughout the development of a digital public health intervention. Ideally, this approach also includes the target group (eg, for an app) to increase acceptance and use. Generally, participatory development processes are iterative and may be designed in various forms depending on the goal. In principle, the following four steps can shape the process: (1) concept generation and ideation; (2) prototype design and system development; (3) Evaluation; and (4) deployment, including various feedback loops (Figure 3). After an initial analysis of the user's needs, the developers collect the criteria for functions and design. Then, they convert these recommendations into the functional specifications of a user-centered design. Using walkthroughs and usability testing, prototypes are tested and perfected before deployment, which helps to expose latent practical and interface design weaknesses. The developing team can achieve this by analyzing remotely collected data using automatic data transmission or video use. As usability testing is a pillar of the best practices for medical system architecture [28], production teams should also apply this to digital public health interventions' development.

XSL•FO
RenderX

Table 1. The 9 different steps for including participant perspectives [6].

Participation and steps	Descriptions
No participation	
Step 1: instrumentalization	• 1.1 The interests of the target group are not necessary • 1.2 Production team makes decisions outside the target group • 1.3 The interests of the decision-makers are the focus of attention • 1.4 Target group members as decoration
Step 2: instruction	• 2.1 The situation of the target group is perceived • 2.2 The problem is defined exclusively from the perspective of the decision-makers (professionals) • 2.3 The opinion of the target group is not considered • 2.4 Communication is direct
Precursors for participation	
Step 3: information	• 3.1 The decision-makers tell the target group what problems the group has and what help they need • 3.2 Recommendation of various courses of action • 3.3 Explanation and justification of the procedure of the decision-makers • 3.4 The point of view of the target group is considered to increase the acceptance of the messages
Step 4: consultation	• 4.1 The decision-makers are interested in the view of the target group. • 4.2 The members of the target group are listened to
Step 5: involvement	• 5.1 The decision-makers are advised by (selected persons from) the target group
Participation	
Step 6: co-determination	• 6.1 The decision-makers consult with the target group • 6.2 Negotiations between target group representatives and decision-makers • 6.3 The target group members have a say
Step 7: partial transfer of decision-making authority	• 7.1 A right of participation in the decision-making process • 7.2 Decision-making authority is limited to certain aspects
Step 8: decision-making power	• 8.1 The target group itself determines all essential aspects • 8.2 Partnership-based cooperation between all parties involved • 8.3 Accompaniment or support of others
Step 9: self-organization	• 9.1 The responsibility for a measure or a project is entirely in the hands of the target group

Figure 3. Schematic representation of the user-centered design process [28].



XSL•FO
RenderX

Discussion

Principal Findings

The aim of this viewpoint paper was to define digital public health interventions and provide an exemplary classification framework for digital public health interventions. Such an approach may help identify core areas of digital public health interventions, which in turn might be helpful during the development and evaluation phases of digital public health interventions. We argue that it is crucial to examine digital public health interventions from 3 different perspectives. The first one should be the WHO framework for EPHF [4]. This is important as it provides an overview of what kind of activities, which strengthen and maintain health at the population level, belong to public health as a discipline and, therefore, what a public health intervention may be. The second perspective focuses on the digital aspects of an intervention. A suitable framework is the *Evidence Standards Framework for Digital Health Technologies* by NICE, as it classifies digital interventions based on their functions and defines corresponding evidence standards [5]. Both frameworks combined enable us to categorize digital public health interventions according to the area of public health and the level of interaction between the user and the digital tool. The last perspective focuses on user involvement in the development of such interventions, as proposed by Wright [6]. This is of great importance, as studies suggest that the acceptance of target users increases with more involvement in the process of development, testing, and implementation. Therefore, acknowledging the 9 levels of user participation (and focusing on levels 6 to 9) may enable developers to create even more significant and meaningful digital public health interventions for their target group.

Our current approach relates to a single and, to the best of our knowledge, the only definition of digital public health. As it is natural for such definitions to evolve over time as the field evolves, our suggestion for a definition of digital public health interventions might also evolve, as one cannot talk about a definition for digital public health interventions without defining the borders of digital public health. Although the suggested EPHF in this perspective piece refers to a summary of the WHO, some readers of this paper might find it hard to apply it to their specific context. This may very much depend on the health care system in which the digital public health intervention is developed. Therefore, the EPHF listed in this study should not be seen as a final list of public health functions or classification frameworks but rather as examples of core functions and goals. Similarly, the NICE framework might not be applied directly in other countries with different health systems and contexts; however, it might provide a helpful starting point for identifying relevant frameworks for such systems or developing their own frameworks that focus on interaction and functional classifications. Possible steps for participation to include user perspectives and methods (eg, user stories) might differ depending on the format and content of a specific intervention. For example, one cannot expect an app that facilitates communication between physicians and patients to unfold its full potential when the development team does not consider both perspectives regarding design, functions, and content [29,30]. Some effects might be more visible on a public health scale than others, depending on the population's size and the health system for which the intervention was initially developed. More importantly, digital public health interventions should display their effectiveness beyond the laboratory in the real world. They should do so by providing study results with high internal validity and results with high external validity. This well-known approach within empirical social research ensures that measurable effects transpire from the laboratory to the real world.

The following example aims to display the connections among the 3 analyzed frameworks and models. Since the beginning of the SARS-CoV-2 pandemic in early 2020, various countries have developed contact-tracing apps for monitoring and surveillance [31]. The primary function of such apps is to notify users after contact with someone who was (later) tested positive for the SARS-CoV-2 virus [32]. As previously mentioned, contact-tracing apps serve the horizontal EPHF of health information systems. Most countries set the bars for data security in contact-tracing apps high to improve users' trust. Conversely, a high level of data protection prevents the collection and analysis of epidemiologically relevant data, making it more difficult to assess the effectiveness from a public health perspective. When developing a contact-tracing app, it is necessary to weigh the protection of privacy and the potential public health benefits against each other [33]. This constraint of data availability for public health (research) limits contact-tracing apps to evidence tiers 1 and 2 within the NICE framework for DHIs. Although simple monitoring (as level 2 demands) is possible, the apps do not aim to calculate the diagnoses needed for tier 3 (instead, recommendations such as different colors for warning levels in the German app).

As previously mentioned, participation in the development process is key to a successful intervention. Germany introduced the *Corona Warn App* in June 2020. The code was published as an open-source project in May 2020 on the GitHub coding platform [34]. According to the developers, this approach should allow interested target group users to assess the code for themselves and add suggestions to improve the app [34]. It is also possible to claim an interest in working on a specific part of the app's code, which suggests a high level of involvement. Target group users are not just listened to but can also actively participate in the development process. However, although this approach offers a high level of participation for some members of the target group with a background or interest in coding, this approach excludes most other users because of their missing knowledge in information and communication technology. Furthermore, no clear information on the extent to which user groups were included in the actual development process is available. Despite the limited involvement of the target group, the app was downloaded 28.3 million times with 472.960 positive tests shared within the app by June 11, 2021, suggesting at least some success [35].

Conclusions

This study aimed to provide the first definition and classification framework for digital public health interventions. Here, we suggest that digital public health, as a complex intervention,

XSL•FO
RenderX

should be viewed from different perspectives. As such, our proposed classification is a combination of the elements of already existing models, specifically, the EPHF by the WHO, which provided us with an overview of the necessary core functions of public health that might also be addressed by digital means. The NICE framework gave us an overview of different areas for digital technologies and potential evaluation requirements. Both models together form a framework for describing digital public health interventions. However, without the inclusion of target groups in user-centered processes during the development, these interventions may lack efficiency and the acceptance of potential users. Therefore, we propose an established user-centered design process to be included in the development of digital public health interventions. Nevertheless, users of our definition and framework must check the validity of our criteria within their setting (eg, population structure, understanding of public health, and health care system). Taking the different strains of research that together might provide a better understanding of the term *digital public health intervention*, the first definition might be as follows:

> *A Digital Public Health Intervention addresses at least one essential Public Health function through digital means. Applying a framework for functional classification and stratification categorizes its interaction level with the user. The developmental process of a digital public health intervention includes the user perspective by applying participatory methods to support its effectiveness and implementation with the goal to achieve a population health impact.*

The first step toward a potential intervention classification framework was developed based on this definition and its underlying frameworks (Multimedia Appendix 1). The aim of this framework is three-fold: (1) support the future reporting of digital public health intervention functions and effectiveness by providing a framework for classification, (2) identify future requirements (eg, for evaluation) of such interventions, and (3) support the implementation processes of digital public health interventions by linking implementation needs and characteristics with the classification framework (ie, a digital public health intervention addressing active monitoring in health care with high levels of user involvement might have other implementation needs than a digital public health intervention that addresses simple monitoring in health care with low levels of user involvement) [36]. We view a combination of all 3 models as a chance to set up a first definition and classification for digital public health interventions and hope that our approach will encourage the uptake and further development of our idea.

Acknowledgments

The authors gratefully acknowledge Leibniz Science Campus Bremen Digital Public Health (lsc-diph.de) support, jointly funded by the Leibniz Association (W4/2018), the Federal State of Bremen, and the Leibniz Institute for Prevention Research and Epidemiology–BIPS.

Authors' Contributions

JW provided the idea for the manuscript, defined the criteria for digital public health interventions, prepared the first definition for such interventions, came up with the first draft of the manuscript, and revised the final draft of the manuscript. JW and LM contributed equally to the manuscript and further developed the draft for the final paper. TJ provided critical feedback and contributed to reviewing and editing the final draft.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Digital public health intervention classification framework.
[DOCX File , 80 KB - jmir_v24i6e31921_app1.docx]

References

1. 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]
2. Zeeb H, Pigeot I, Schüz B. [Digital public health-rapid technological progress, but many open public health questions]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2020 Feb;63(2):135-136 [FREE Full text] [doi: 10.1007/s00103-020-03092-0] [Medline: 31960072]
3. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
4. Essential Public Health Functions, Health Systems and Health Security Developing Conceptual Clarity and a WHO Roadmap for Action. Geneva: World Health Organization; 2018.

XSL•FO
RenderX

5. Evidence standards framework for digital health technologies. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/evidence-standards-framework/digital-evidence-standards-framework.pdf [accessed 2022-02-17]
6. Wright MT. [Participatory health research: origins and current trends]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2021 Feb;64(2):140-145 [FREE Full text] [doi: 10.1007/s00103-020-03264-y] [Medline: 33336312]
7. Nievas SB, García DS, Fernández AA, Bonillo PA, Parrón CT. eHealth: advantages, disadvantages and guiding principles for the future. JMIR 2019 (forthcoming). [doi: 10.2196/preprints.15366]
8. 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]
9. Shaw T, McGregor D, Brunner M, Keep M, Janssen A, Barnet S. What is eHealth (6)? Development of a conceptual model for eHealth: qualitative study with key informants. J Med Internet Res 2017 Oct 24;19(10):e324 [FREE Full text] [doi: 10.2196/jmir.8106] [Medline: 29066429]
10. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
11. Iacono T, Stagg K, Pearce N, Hulme Chambers A. A scoping review of Australian allied health research in eHealth. BMC Health Serv Res 2016 Oct 04;16(1):543 [FREE Full text] [doi: 10.1186/s12913-016-1791-x] [Medline: 27716325]
12. Pagliari C, Sloan D, Gregor P, Sullivan F, Detmer D, Kahan JP, et al. What is eHealth (4): a scoping exercise to map the field. J Med Internet Res 2005 Mar 31;7(1):e9 [FREE Full text] [doi: 10.2196/jmir.7.1.e9] [Medline: 15829481]
13. Davis TL, DiClemente R, Prietula M. Taking mHealth forward: examining the core characteristics. JMIR Mhealth Uhealth 2016 Aug 10;4(3):e97 [FREE Full text] [doi: 10.2196/mhealth.5659] [Medline: 27511612]
14. Fatehi F, Samadbeik M, Kazemi A. What is digital health? Review of definitions. Stud Health Technol Inform 2020 Nov 23;275:67-71. [doi: 10.3233/SHTI200696] [Medline: 33227742]
15. Zeeb H, Pigeot I, Schüz B, Leibniz-WissenschaftsCampus Digital Public Health Bremen. [Digital public health-an overview]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2020 Feb;63(2):137-144. [doi: 10.1007/s00103-019-03078-7] [Medline: 31919531]
16. Digital health applications (DiGA). Federal Institute for Drugs and Medical Devices. URL: https://www.bfarm.de/EN/MedicalDevices/DiGA/_node.html [accessed 2022-02-17]
17. Davalbhakta S, Advani S, Kumar S, Agarwal V, Bhoyar S, Fedirko E, et al. A systematic review of smartphone applications available for corona virus disease 2019 (COVID19) and the assessment of their quality using the mobile application rating scale (MARS). J Med Syst 2020 Aug 10;44(9):164 [FREE Full text] [doi: 10.1007/s10916-020-01633-3] [Medline: 32779002]
18. Murray CJ, Alamro NM, Hwang H, Lee U. Digital public health and COVID-19. Lancet Public Health 2020 Sep;5(9):e469-e470 [FREE Full text] [doi: 10.1016/S2468-2667(20)30187-0] [Medline: 32791051]
19. Bradford NK, Caffery LJ, Smith AC. Telehealth services in rural and remote Australia: a systematic review of models of care and factors influencing success and sustainability. Rural Remote Health 2016;16(4):4268 [FREE Full text] [Medline: 27817199]
20. Edwards EA, Lumsden J, Rivas C, Steed L, Edwards LA, Thiyagarajan A, et al. Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. BMJ Open 2016 Oct 04;6(10):e012447 [FREE Full text] [doi: 10.1136/bmjopen-2016-012447] [Medline: 27707829]
21. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
22. Kloc K, Rémuzat C, François C, Toumi M. PNS235 assessment of digital health technologies - comparison of evidence frameworks of nice and has. Value Health 2019 Nov;22(3):S801. [doi: 10.1016/j.jval.2019.09.2135]
23. Forsyth JR, Chase H, Roberts NW, Armitage LC, Farmer AJ. Application of the national institute for health and care excellence evidence standards framework for digital health technologies in assessing mobile-delivered technologies for the self-management of type 2 diabetes mellitus: scoping review. JMIR Diabetes 2021 Feb 16;6(1):e23687 [FREE Full text] [doi: 10.2196/23687] [Medline: 33591278]
24. Nwe K, Larsen ME, Nelissen N, Wong DC. Medical mobile app classification using the national institute for health and care excellence evidence standards framework for digital health technologies: interrater reliability study. J Med Internet Res 2020 Jun 05;22(6):e17457 [FREE Full text] [doi: 10.2196/17457] [Medline: 32501271]
25. Hochmuth A, Exner A, Dockweiler C. [Implementation and participatory design of digital health interventions]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2020 Feb;63(2):145-152. [doi: 10.1007/s00103-019-03079-6] [Medline: 31938837]
26. Bergold J, Thomas S. Participatory research methods: a methodological approach in motion. Forum Qual Soc Res 2012;13(1) [FREE Full text]
27. Vastine A, Gittelsohn J, Ethelbah B, Anliker J, Caballero B. Formative research and stakeholder participation in intervention development. Am J Health Behav 2005;29(1):57-69. [doi: 10.5993/ajhb.29.1.5] [Medline: 15604050]
28. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]

XSL•FO
RenderX

29. Lippke S, Wienert J, Keller FM, Derksen C, Welp A, Kötting L, et al. Communication and patient safety in gynecology and obstetrics - study protocol of an intervention study. BMC Health Serv Res 2019 Nov 28;19(1):908 [FREE Full text] [doi: 10.1186/s12913-019-4579-y] [Medline: 31779620]
30. Wienert J. Understanding health information technologies as complex interventions with the need for thorough implementation and monitoring to sustain patient safety. Front ICT 2019 May 17;6. [doi: 10.3389/fict.2019.00009]
31. A flood of coronavirus apps are tracking us. MIT Technology Review. URL: https://www.technologyreview.com/2020/05/07/1000961/launching-mittr-covid-tracing-tracker/ [accessed 2022-02-17]
32. Effective configurations of a digital contact tracing app: a report to NHSX. CDN. URL: https://cdn.theconversation.com/static_files/files/1009/Report_-_Effective_App_Configurations.pdf?1587531217 [accessed 2022-02-17]
33. Jahnel T, Gerhardus A, Wienert J. Digitales contact tracing: dilemma zwischen Datenschutz und Public Health Nutzenbewertung. Datenschutz Datensich 2020 Nov 05;44(12):786-790. [doi: 10.1007/s11623-020-1367-0]
34. Corona Warn-App hompage. Corona Warn-App. URL: https://www.coronawarn.app/en/community/ [accessed 2022-02-17]
35. Key figures for the Corona-Warn-App. Robert Koch-Institut. URL: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/WarnApp/Archiv_Kennzahlen/Kennzahlen_29012021.pdf [accessed 2022-02-17]
36. Wienert J, Zeeb H. Implementing health apps for digital public health - an implementation science approach adopting the consolidated framework for implementation research. Front Public Health 2021;9:610237 [FREE Full text] [doi: 10.3389/fpubh.2021.610237] [Medline: 34026702]

Abbreviations

DHI: digital health intervention
EPHF: Essential Public Health Function
mHealth: mobile health
NICE: National Institute for Health and Care Excellence
WHO: World Health Organization

Edited by A Mavragani; submitted 09.07.21; peer-reviewed by L Gupta, M Prietula, A Janssen; comments to author 21.12.21; revised version received 04.01.22; accepted 04.01.22; published 28.06.22.

Please cite as:
Wienert J, Jahnel T, Maaß L
What are Digital Public Health Interventions? First Steps Toward a Definition and an Intervention Classification Framework
J Med Internet Res 2022;24(6):e31921
URL: https://www.jmir.org/2022/6/e31921
doi:10.2196/31921
PMID:35763320

©Julian Wienert, Tina Jahnel, Laura Maaß. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.06.2022. 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

Indicated Web-Based Prevention for Women With Anorexia Nervosa Symptoms: Randomized Controlled Efficacy Trial

Corinna Jacobi[1], Prof Dr; Bianka Vollert[1], Dipl-Psych; Kristian Hütter[1], Dipl-Psych; Paula von Bloh[1], Dipl-Psych; Nadine Eiterich[1], Dipl-Psych; Dennis Görlich[2], Dr rer nat; C Barr Taylor[3], MD

[1]Department of Clinical Psychology and Psychotherapy, Faculty of Psychology, Technische Universität Dresden, Dresden, Germany

[2]Institute of Biostatistics and Clinical Research, Westfälische Wilhelms-Universität Münster, Münster, Germany

[3]Department of Psychiatry, Stanford University School of Medicine, Stanford, CA, United States

Corresponding Author:
Corinna Jacobi, Prof Dr
Department of Clinical Psychology and Psychotherapy
Faculty of Psychology
Technische Universität Dresden
Chemnitzer Strasse 46
Dresden, 01187
Germany
Phone: 49 351 46338576
Fax: 49 351 463 463372
Email: corinna.jacobi@tu-dresden.de

Abstract

Background: Although preventive interventions for eating disorders in general have shown promise, interventions specifically targeting individuals at risk for anorexia nervosa (AN) are lacking.

Objective: The aim of this study was to determine the efficacy of a guided, indicated web-based prevention program for women at risk for AN.

Methods: We conducted a randomized controlled efficacy trial for women at risk for AN. Assessments were carried out at baseline (before the intervention), after the intervention (10 weeks after baseline), and at 6- and 12-month follow-ups (FUs). A total of 168 women with low body weight (17.5 kg/m^2≤BMI≤19 kg/m^2) and high weight concerns or with normal body weight (19 kg/m^2<BMI≤25 kg/m^2), high weight concerns, and high restrained eating were recruited from 3 German universities as well as on the web and randomized to Student Bodies-AN (SB-AN; intervention group [IG]) or a wait-list control group (CG). The exclusion criteria were current Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition–based full-syndrome eating disorders and serious medical or mental problems. The interventions were a cognitive-behavioral guided web-based prevention program (SB-AN) over 10 weeks (IG) and a wait-list CG. The primary outcomes were clinically significant changes in disordered eating attitudes and behaviors and change in BMI at 12-month FU in the group of participants who were underweight. The secondary outcomes were new onset of eating disorders, symptoms of disordered eating, and associated psychopathology.

Results: Data were available for 81.5% (137/168) of the women after the intervention and for 69% (116/168) of the women at 12-month FU. At 12-month FU, the IG participants showed larger decreases in Eating Disorder Examination total scores (38/48, 79% vs 33/58, 57%) than the CG participants and the IG participants who were underweight also showed larger clinically relevant increases in BMI (15/31, 49% vs 10/32, 32%) than the CG participants, but these differences were not significant. In addition, after the intervention and at 12-month FU, we found a significant increase in continuously measured BMI for the participants who were underweight and significant improvements in disordered eating attitudes and behaviors (eg, restrained eating as well as weight and shape concerns). At all time points, the rates of new-onset eating disorder cases were (nonsignificantly) lower in the IG than in the CG and the reductions in Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition–based eating disorder syndromes were (nonsignificantly) higher in the IG than in the CG.

Conclusions: SB-AN is the first preventive intervention shown to significantly reduce specific risk factors for, and symptoms of, AN and shows promise for reducing full-syndrome AN onset.

Trial Registration: ISRCTN Registry ISRCTN70380261; https://www.isrctn.com/ISRCTN70380261

XSL•FO
RenderX

(J Med Internet Res 2022;24(6):e35947) doi:10.2196/35947

KEYWORDS

anorexia nervosa; internet; indicated prevention

Introduction

Background

Anorexia nervosa (AN) is a serious condition, often accompanied by severe medical complications and high psychiatric comorbidity [1]. Mortality rates for AN are higher than for any other psychiatric disorder [2,3]. Evidence from controlled treatment trials for AN is limited compared with trials for other eating disorders (EDs), with no specific treatment for older adolescents or adults demonstrating clear superiority over nonspecific treatment [4,5]. Studies addressing long-term outcomes of AN have also demonstrated a rather poor outcome for at least a third to half of the patients [3,6]. Finally, patients with AN also have significantly impaired health-related quality of life and AN is associated with increased health care use and health care costs [7-9]. Given the seriousness and often chronic course of the disorder, early preventive interventions are of crucial importance. These interventions should target modifiable potent risk factors to reduce the onset of the disorder and mitigate core symptoms of the disorder before the onset, thus lowering risk for AN onset. However, although a number of longitudinally assessed risk factors for EDs in general have been confirmed, knowledge regarding specific risk factors for AN is still very scarce [10,11].

Several previous reviews and meta-analyses have examined the efficacy of universal targeted or indicated prevention programs for EDs delivered face-to-face [12-16] or over the internet [12,17-19]. Overall, these reviews found evidence that preventive interventions can reduce potent risk factors for, and symptoms of, EDs, with mostly small to moderate effect sizes. A few individual studies [20-23] also found evidence that preventive interventions can reduce new onset of (mostly) bulimia nervosa (BN) or binge eating–type EDs. However, because specific risk factors for specific ED diagnoses have not been replicated, participants in studies with targeted or indicated programs are usually selected based on general modifiable potent risk factors for ED, such as weight concern, shape concern, or body dissatisfaction. These interventions are not specifically directed at individuals at risk for specific ED diagnoses such as AN. Only recently, in an amalgam of 3 previous prevention trials, Stice et al [24] identified some risk factors with unique predictive effects for ED diagnoses. In the study, low BMI and dieting were found to specifically predict onset of subthreshold or threshold AN. However, based on (21 out of 26) prevention trials included in 2 meta-analyses, the mean BMI of young adult participants was 23.3 (SD 2.8; range 21.6-24.8) kg/m^2 [15] and 23.5 (SD 0.9; range 21.9-25.5) kg/m^2 [12]; none of the studies had included lower body weight to determine risk status as the selection criterion. Consequently, adult participants with a lower BMI (ie, BMI<21 kg/m^2) who may be specifically at risk for AN were not included in these programs.

The question of which variables might moderate intervention effects for specific symptoms or diagnoses of EDs has also hardly been addressed by meta-analyses [19] and individual studies. Of the few studies, 1 found the largest effects of a web-based prevention program on onset of subclinical BN and binge ED (BED) for participants with higher BMI and higher levels of compensatory behaviors at baseline [22]. A second study found lowest intervention effects on abstinence of binge eating, compensatory behaviors, and restrictive eating after the intervention for individuals with purely restrictive eating at baseline [25].

Prior Work

As part of a pilot study, we specifically designed a web-based indicated preventive intervention (*Student Bodies-AN* [SB-AN]) for this risk group and assessed its feasibility, acceptance, and effectiveness in a pilot study with 36 women, including those with low BMI (<19 kg/m^2) and higher restrained eating. Overall, the pilot study showed that recruitment of participants at risk for AN with low body weight and high restrained eating is feasible and shows promise. We found significant pre–post reductions in common risk factors for EDs (eg, weight concern) with medium to large effects, as well as specific effects for the underweight subgroup in terms of reductions in restrained eating and an increase in BMI [26].

Goal of This Study

The major objective of this study was to determine the efficacy of this web-based intervention for women at risk for AN in reducing core risk factors; early symptoms; or syndrome progression of pre-existing, or onset of newly emerging subclinical syndromes of AN compared with a wait-list control group (CG). We hypothesized that the intervention group (IG) participants would show greater improvements in attitudes and symptoms that are more specific for AN, that is, low BMI, and in general ED risk factors such as weight concern and shape concern. In addition, we expected that the participating women would show significantly fewer subclinical ED syndromes at 12-month follow-up (FU).

Methods

Study Design and Participants

We conducted a randomized controlled efficacy trial in women at risk for AN. Participants were screened, recruited, and assessed between September 2013 and November 2015 through different faculties at 3 German universities (Dresden, Leipzig, and Halle) and other educational institutions in Dresden through announcements in local media, health insurance membership magazines, and social media (eg, Facebook) as well as through flyer distribution. To ensure a high-enough rate of women with subclinical AN, the study was also announced to women seeking information or help at a secondary advisory center for EDs (ANAD eV, Munich).

XSL•FO
RenderX

We included women aged >18 years with high weight concerns (Weight Concerns Scale [WCS] score>42) and lower body weight (17.5 kg/m^2≤BMI≤21 kg/m^2), or with normal body weight (21 kg/m^2<BMI≤25 kg/m^2), high weight concerns, and high restrained eating (Eating Disorder Examination [EDE] Questionnaire Restraint score≥2.6, which was >1 SD above the mean of healthy controls [23]). The exclusion criteria were a current Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV)–based full-syndrome AN, BN, or BED; serious medical or mental problems such as current substance abuse, acute or chronic organic or schizophrenic psychosis, and severe suicidal ideation or behavior; and no internet access.

After the screening, the assessment points were as follows: before the intervention (baseline), midintervention point, after the intervention (10 weeks after baseline), and 6- and 12-month FUs. Quality control methods comprised case report forms, independent data management, on-site monitoring, and documentation of adverse and severe adverse events. The Koordinierungszentrum für Klinische Studien (KKS; Coordination Center for Clinical Trials), Dresden, was responsible for setting up a database according to International Council for Harmonisation Good Clinical Practice requirements, using the software MACRO 4.0 (Microsoft). To ensure data quality, validity and consistency checks were programmed for data entry and regularly checked by a research assistant.

Ethics Approval

The study was approved by the local ethics committee of Technische Universität Dresden, Germany (EK264082012). Written informed consent was obtained from all participants. The study was conducted according to the Declaration of Helsinki and Good Clinical Practice principles.

Patient Involvement

Patients with a full-syndrome ED at baseline were excluded from the study. Study participants were not involved in the research question, design of the study, development of outcome measures, or recruitment. However, participant feedback on the intervention included in the pilot study led to some content-related and technical revisions of the intervention (eg, improvement in the dashboard function and the technical usability of the platform, revision of instructions for the symptom checklist and self-monitoring diary, and inclusion of a booster session 2 months after the end of the intervention).

Randomization and Masking

Concealed randomization was carried out centrally in a ratio of 1:1 by an independent clinical trials center (KKS) after participants had been enrolled in the study and had given informed consent. The randomization was stratified by weight group (underweight, 17.5 kg/m^2≤BMI≤19 kg/m^2; low weight, 19 kg/m^2<BMI≤21 kg/m^2; normal weight, BMI>21 kg/m^2) at baseline. A block randomization with random block sizes was used. The KKS also carried out data monitoring and statistical analyses for the main outcomes. The assessors (Anne Buchholz, Sarah Bunzel, Silke Elsäßer, Melanie Hassler, Sabrina Irrgang, Gerda Keil, Francie Kriegel, Franziska Miksch, Annegret Neubauer, Angelika Schuster, Juliane Thieme, Pia Trübenbach, Anna Wagner, and Monique Zobel) who carried out the postintervention and FU assessments were blind to intervention allocation and were not involved in either the moderation of the intervention or the final data analyses.

Procedures

Participants were recruited through lectures and seminars from different departments with a high proportion of women (eg, psychology or social sciences) or a large number of students in general (eg, business studies) of 3 German universities (Dresden, Halle, and Leipzig). All female students were invited to participate in a study to improve body image and asked to fill out a short screening questionnaire (either a paper-and-pencil version or a web-based version). In addition, the web-based version of this questionnaire was advertised through posters, university mailing lists, flyers, websites of ED associations, local and nationwide media (eg, Facebook), and health insurance companies. Women who screened positive were subsequently invited to a face-to-face or telephone interview where the study was described in detail and informed consent was obtained. Thereafter, the EDE interview [27,28] was administered to assess a current or past ED and participants received log-in data to access the password-protected web-based platform to fill out baseline self-report questionnaires. If participants met the criteria for a current full-syndrome ED, the research team provided treatment recommendations. Postintervention as well as 6- and 12-month FU assessments also included EDE interviews and self-report questionnaires provided through the web-based platform that hosted the intervention.

Participants were provided individual feedback on current ED risk factors (EDE scores, BMI, and ED symptoms) at baseline and on change in the risk factors at postintervention and FU assessments. At the completion of each interview, participants received €20 (US $21.7).

The Intervention (SB-AN)

We designed the intervention SB-AN based upon existing targeted web-based cognitive-behavioral versions of the intervention *Student Bodies* [22,25,29], expanding its duration from 8 to 10 weekly sessions. The core goals of these programs are to reduce weight concern as well as shape concern, enhance body image, promote healthy weight regulation, and increase knowledge of the risks associated with EDs and specific ED symptoms; for example, binge eating and compensatory behaviors. The program is supplemented by a web-based asynchronous moderated discussion group. Other elements include a personal journal and a body image journal.

For this study, we made contextual adaptations according to the special needs of the groups with higher restrained eating or lower weight. In anticipation of the noted ambivalence to change in this population, we added elements of motivational interviewing [30] to the first sessions (eg, pros and cons of low weight and restrained eating). We also expanded the psychoeducational content on EDs to increase participants' awareness of their current eating and exercise behavior as well as body image compared with patients with other EDs. Furthermore, compared with programs addressing non-AN EDs,

XSL•FO
RenderX

the program focused more specifically on restrictive eating. Other topics addressed were media literacy, coping with negative emotions, improving social skills, and healthy eating and exercise. The symptom checklist that was integrated into the web-based program was expanded to include frequencies of core ED symptoms (body weight, restrained eating, meals per day, missed meals, reduced meals, avoided foods, objective and subjective episodes of binge eating, episodes of vomiting, laxative abuse, abuse of diuretics or appetite suppressants to control weight, and driven exercise). To normalize their eating behavior and reduce ED symptoms (eg, restrained eating, binge eating, and purging), participants were prompted weekly to fill out the symptom checklist and given individual weekly feedback by the program moderators on their entries in the symptom checklist and other interactive program elements (ie, personal and self-monitoring logs and contributions to the web-based discussion group). The program was moderated by psychology (master's degree or diploma level) graduate students in training for behavior therapy who were supervised by a licensed clinical psychologist (CJ). The feedback was intended to foster reflection on, and change in, dysfunctional eating and weight-related thoughts and behaviors. Each session of the program took 45 to 90 minutes to be completed. The program's home page also provided short résumés of the program moderators once participants had logged in to facilitate the credibility of the intervention.

Wait-list CG

Given that SB-AN is the first prevention program specifically targeting women at risk of AN, no alternative interventions (*treatment as usual*) exist. A wait-list CG therefore seemed to be the first-choice control condition and ethically justifiable to determine the efficacy of the intervention. Participants assigned to the CG were assessed at all interview assessment points and offered to participate in the program after completion of the 12-month FU.

Outcome Measures

Outcomes were selected to reflect core features of the included risk group and based on preliminary effects found in the pilot study [26]. The primary outcomes were rates (in percentages) of participants with a decrease in EDE interview total score [31] below a score of 1.87 between before the intervention and 12-month FU (reflecting a clinically significant change) and rates (in percentages) of participants who were underweight with a BMI increase of at least 0.8 kg/m^2 between screening and 12-month FU.

The secondary outcomes were continuously measured BMI of participant who were underweight, disordered eating attitudes and behaviors, numbers of subjective and objective binge eating episodes (for the binge eating subgroup), and rates of participants fulfilling criteria for onset of a full-syndrome and subclinical ED.

Disordered eating attitudes and behaviors were assessed by the WCS [32], the EDE (total score; subscales: Weight Concern, Shape Concern, Eating Concern, and Restraint; and numbers of objective and subjective binge eating episodes [27]), and the Eating Disorder Inventory-2 (EDI-2) subscales Drive for Thinness and Body Dissatisfaction [33]. For all measures, good psychometric properties have been reported for both the original and the German-validated versions [28,31,32,34-37]. Additional measures covered associated psychopathology such as general psychopathology (Brief Symptom Inventory [38]) and depression (Beck Depression Inventory [39]), as well as a knowledge test concerning program content. Good psychometric properties have been reported for all these measures for both the original versions and the German adaptations [40,41]. To assess clinical impairment due to ED psychopathology we used our own translation of the Clinical Impairment Assessment [42,43].

After the study had begun, DSM, Fifth Edition (DSM-5) [44] was published, which included slight changes to some of the diagnostic criteria. The new classification system loosens the criteria for some EDs, resulting in individuals who were not diagnosed as being ED cases with DSM-IV now becoming ED cases with DSM-5. We decided to adopt DSM-5 criteria for all baseline and FU assessments. This resulted in 9 individuals becoming full-syndrome AN cases at baseline, which allowed us to determine the treatment effects of the intervention in reducing symptoms in this group. To determine the preventive effect of the intervention, we excluded all DSM-5 cases at baseline and only examined individuals who became cases according to the new criteria.

In addition, participants who met all criteria in accordance with the following definitions were considered to be cases of subclinical ED (Eating Disorder Not Otherwise Specified and Other Specified Feeding or Eating Disorder [44]) at baseline or subsequent assessment points: subclinical AN: (1) 18.5 kg/m^2>BMI<19.2 kg/m^2), (2) fear of weight gain in the past 3 months (DSM-5 criterion B), and (3) either undue influence of body weight or shape on self-esteem or feeling fat on more than half of the days in the past 3 months (DSM-5 criterion C) or (1) BMI<18.5 kg/m^2 and (2) DSM-5 criterion B or C. Subclinical BN: All DSM-5 criteria for BN are met except undue influence of body weight and shape on self-esteem (DSM-5 criterion D). Subclinical BED: All DSM-5 criteria are met except marked distress regarding binge eating (DSM-5 criterion C).

All interviews were conducted by assessors not involved in intervention moderation and data analyses. They completed a 2-day workshop during which they were trained on the EDE interview assessments, on the use of the database for the assessments, and on providing feedback regarding ED psychopathology to participants. Feedback to participants was recorded, and the assessors were supervised by graduate students regarding the quality of the feedback provided. Over the course of the trial, interviewer trainings were repeated for new interviewers.

BMI was measured using a portable stadiometer measuring height to the nearest millimeter and a digital scale measuring weight to the nearest 0.1 kg. In case of telephone interviews, BMI was calculated based on self-reported height and weight (obtained during the interview). All primary and secondary outcomes were assessed at all 4 assessment points. After the intervention as well as at 6- and 12-month FUs we also assessed the use of any additional inpatient, outpatient, or day-patient

XSL•FO
RenderX

treatment that patients had received since the start of their participation in the SB-AN study.

Sample Size

We based sample size calculations on the assumption that 50% of the IG and 20% of the CG would show a decrease in EDE total score below the critical value of 1.87. Applying the Fisher exact test based on a power of at least 80%, the required sample size for the analysis would be 44 participants per group or a total of 88 participants.

For the group of participants who were underweight, we assumed that a rate of 50% of the participants in the IG succeeding in increasing their BMI between before the intervention and 12-month FU by at least 0.8 kg/m^2 compared with 5% of the CG participants represented a clinically significant difference. To detect this difference, 34 (17 in each group) participants would be required in the group of participants who were underweight (BMI between 17.5 kg/m^2 and 19 kg/m^2; based on the Fisher exact test with α=.05 and power of 80%).

Assuming similar rates of returned screens (74%), women who screen positive (13.9% of the screened women), and eligible women (31.9% of the screened positives) as in the pilot study, we would have to administer 8273 screening questionnaires to obtain 6122 (74%) returned screens, of which 851 (13.9%; n=448, 7.32%, needed) would be women who screen positive. Of these 448 women, 143 (31.92%) would be eligible.

Conservatively estimating an attrition rate of 45% until 12-month FU (15% of the participants after the intervention and another 15% at each of the FU assessments), 88 participants would provide sufficient data for the analyses.

Statistical Analyses

Study data are described using absolute and relative frequencies for categorical variables. Continuous outcomes are described using means and SDs. All analyses of primary and secondary outcomes were conducted as intention-to-treat analyses. We analyzed the study data according to the study protocol and as outlined in the study registry.

Primary Outcomes

Originally, we planned to use the Fisher exact test to compare the 12-month FU rates of participants in the IG and CG fulfilling the primary outcome variables (percentage of participants who were underweight with a BMI increase of at least 0.8 kg/m^2 and percentage of participants with EDE total score of no more than 1.87 with baseline score above 1.87) at a significance level of *P*=.05 for the primary analysis. In this analysis, missing data were imputed using the last observation carried forward (LOCF) method.

As LOCF makes assumptions that can yield biased results and is no longer considered the best method to impute missing data, we also used the more robust method of generalized linear mixed models (GLMMs) [45]. Here, the full available data from each participant at all assessments were considered in the analyses using the maximum likelihood method. We entered fixed effects for randomized group, time, and the group×time interaction. Random intercepts, allowing us to model repeated measurements and to account for heterogeneity in outcomes across individuals, were fitted. To model longitudinal data appropriately, we used a covariance matrix with a first-order autoregressive structure. The primary outcome variables were modeled with a binomial distribution and a logit link. To test differences in binary outcomes between the IG and the CG, Wald test *P* values were calculated.

Secondary Outcomes

GLMMs were used to analyze continuous secondary outcomes as well. Data from each participant at all assessments (baseline, midintervention point [except for EDE scales, BMI, and binge eating episodes], after the intervention, and FUs) were considered in the analyses. We entered fixed effects for randomized group, time, and the group×time interaction and calculated suitable contrasts to test the changes from baseline to all subsequent time points between the groups. Random intercepts were fitted, and we used a covariance matrix with a first-order autoregressive structure. For skewed outcome variables, we used the best fitting distribution (eg, γ) for the model. The canonical link function was chosen for all models.

Changes in dichotomous secondary outcomes (diagnoses) between the groups were analyzed using Fisher exact tests for each assessment point separately.

We also analyzed the onset of a new DSM-5–based diagnosis by means of a time-to-event analysis. Observation times were determined as days between baseline and date of onset or last observation. Participants without an onset were censored in the analysis at their last observation date. Cumulative incidence curves were calculated as 1 minus the Kaplan-Meier estimate. Numbers at risk are given along with the incidence curves. Groups were compared using the log-rank test.

Effect Sizes

We constructed Cohen *d*–like effect sizes from the GLMMs using the known relation between *t* statistics and effect sizes [46]. The resulting effect size can be interpreted similar to common Cohen *d* because it quantifies the effect as standardized estimated mean differences. For binary outcomes, odds ratios were reported as effect size. In addition, we calculated the number needed to treat (NNT) or number needed to harm based on the absolute risk reduction (ARR) [47]. For the time-to-event analysis, hazard ratios and 95% CIs were calculated using a Cox proportional hazards model.

No multiple-testing correction was applied for the analyses of the secondary outcomes because these results are considered explorative. Analyses were performed using SPSS software (version 24.0; IBM Corp) and, for the GLMMs, SAS software (version 9.4 TS1M3, SAS/STAT 14.1; SAS Institute Inc).

Results

Recruitment

Between September 2013 and July 2014, a total of 4646 women were screened for inclusion (n=3741, 80.5%, based on paper-pencil screening and n=905, 19.5%, based on web-based screening); 333 (7.17%) were invited to the preintervention

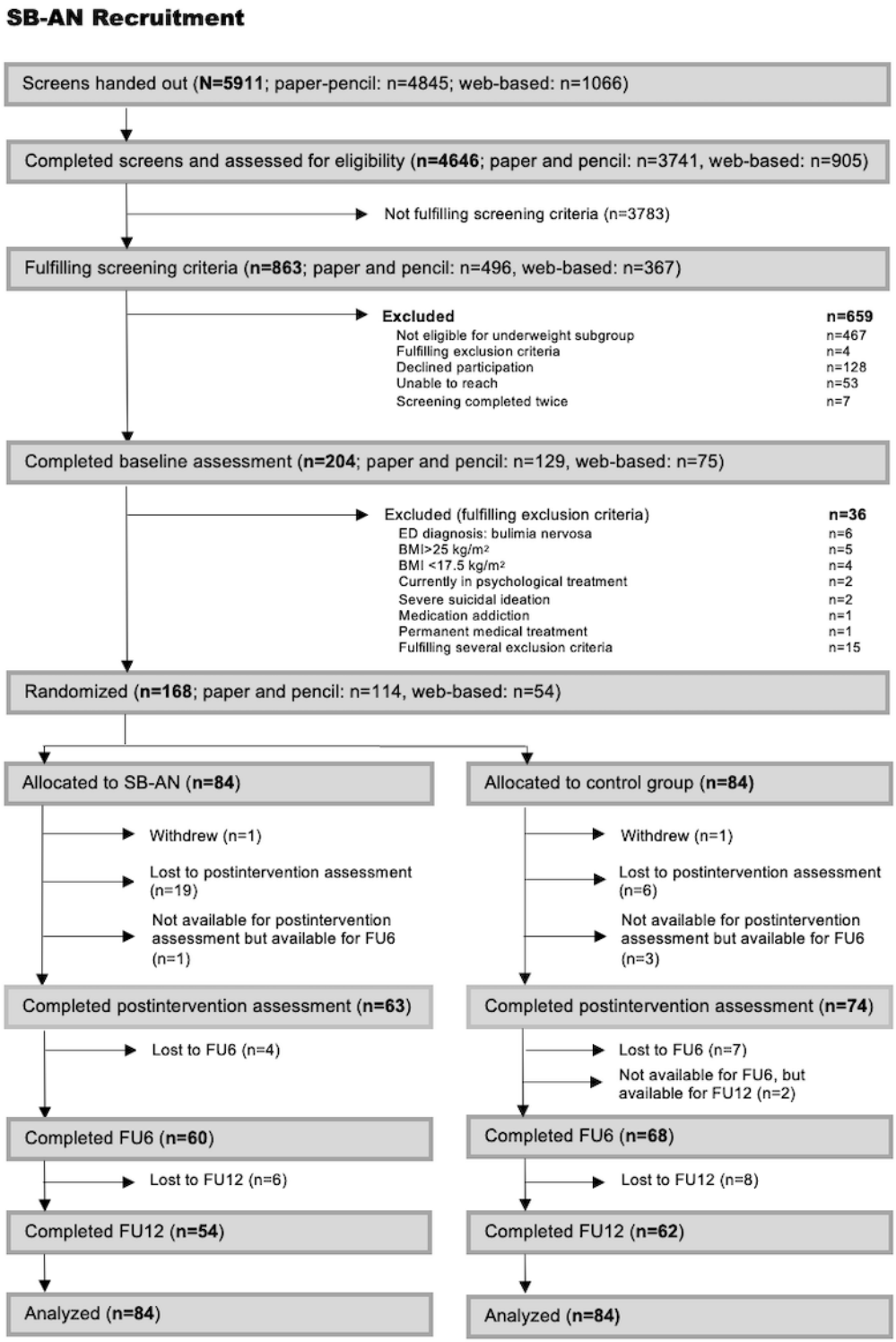
XSL•FO
RenderX

interviews and 168 (3.62%) were randomized to the SB-AN or wait-list control condition. As the recruitment of participants who were underweight took longer in this trial than in the pilot study, we continued our recruitment beyond the originally calculated sample size to ensure a large enough subgroup of participants who were underweight for the main analyses. This resulted in an overall almost doubled sample size of randomized participants. At the end of the intervention period, of the 168 participants, 137 (81.5%) had completed assessments, and at 12-month FU, 64% (54/84) of the participants in the IG and 74% (62/84) of the participants in the CG had completed assessments, resulting in an overall dropout rate of 31% (52/168; 30/84, 36%, and 22/84, 26%, respectively; *P*=.24; Figure 1). Overall, most (114/168, 67.9%) of the EDE assessments of the randomized participants were carried out in person.

The participating women were on average aged 23.3 (SD 3.77) years. Most (140/168, 83.3%) were students from the eastern parts of Germany (Saxony). The average BMI of the sample was 20.08 (SD 1.72) kg/m^2, with 37.5% (63/168) in the lower or underweight BMI range (17.5 kg/m^2≤BMI≤19 kg/m^2). All women showed on average high restrained eating (based on the EDE Restraint subscale) and high weight concerns (based on the WCS). In addition, 12.5% (21/168) of the women reported objective binge eating episodes and vomiting to control weight up to 20 times in the past 4 weeks whereas 11.3% (19/168) engaged in abuse of laxatives or diuretics to control weight up to 32 times in the past 4 weeks. On the basis of the DSM-5 criteria [45] that were published over the course of the study, 5.4% (9/168) of the women (3/84, 4%, in the IG and 6/84, 7%, in the CG) met the criteria for full-syndrome AN. In addition, 7.1% (12/168) of the women (7/84, 8%, in the IG and 5/84, 6%, in the CG) met the criteria for subthreshold AN, whereas 1.2% (2/168) of the women (1/84, 1.2%, in the IG and 1/84, 1.2%, in the CG) met the criteria for subthreshold BN. Table 1 summarizes baseline sociodemographic characteristics and Table 2 shows the baseline clinical scores of all participants.

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow of participants. ED: eating disorder; FU6: 6-month follow-up; FU12: 12-month follow-up; SB-AN: Student Bodies-Anorexia Nervosa.



XSL•FO
RenderX

Table 1. Baseline sociodemographic and clinical characteristics of participants (N=168).

	All participants	Intervention group (n=84)	Control group (n=84)
Age (years), mean (SD)	23.23 (3.77)	22.93 (3.56)	23.53 (3.97)
Education level, n (%)			
University degree	48 (28.5)	22 (26.2)	26 (30.9)
Professional qualification	7 (4.2)	4 (4.8)	3 (3.6)
High school diploma	108 (64.3)	56 (66.7)	52 (61.9)
Secondary school certificate	5 (3)	2 (2.4)	3 (3.6)
Occupation, n (%)			
Employee	20 (11.9)	8 (9.5)	12 (14.3)
Student	140 (83.3)	73 (86.9)	67 (79.8)
Apprentice	2 (1.2)	1 (1.2)	1 (1.2)
Other	6 (3.6)	2 (2.4)	4 (4.8)

XSL•FO
RenderX

Table 2. Baseline clinical characteristics of participants (N=168).

	All participants		Intervention group (n=84)		Control group (n=84)	
	Value, n (%)	Value, mean (SD)	Value, n (%)	Value, mean (SD)	Value, n (%)	Value, mean (SD)
BMI (T0[a])	168 (100)	20.08 (1.72)	84 (100)	20.14 (1.76)	84 (100)	20.02 (1.69)
BMI (T1[b])	168 (100)	20.46 (1.80)	84 (100)	20.51 (1.82)	84 (100)	20.41 (1.78)
BMI (UW[c]-T0)	63 (37.5)	18.40 (0.44)	31 (36.9)	18.43 (0.45)	32 (38.1)	18.38 (0.43)
BMI (UW-T1)	41 (24.4)	18.39 (0.37)	20 (23.8)	18.43 (0.39)	21 (25)	18.35 (0.35)
Binges, objective	14 (8.3)	5.86 (5.22)	6 (7.1)	5.50 (2.59)	8 (9.5)	6.13 (6.75)
Binges, subjective	43 (25.6)	7.16 (5.38)	21 (25)	7.43 (5.54)	22 (26.2)	6.91 (5.34)
Binges, objective+subjective	54 (32.1)	7.22 (5.43)	26 (31)	7.27 (5.39)	28 (33)	7.18 (5.58)
Purging	23 (13.7)	6.78 (6.54)	10 (11.9)	9.60 (9.11)	13 (15.5)	4.62 (2.14)
ED[d] diagnosis	23 (13.7)	N/A[e]	11 (13.1)	N/A	12 (14.3)	N/A
EDE[f] total	168 (100)	2.35 (1.07)	84 (100)	2.21 (1.00)	84 (100)	2.49 (1.13)
EDE RS[g]	168 (100)	2.54 (1.30)	84 (100)	2.46 (1.29)	84 (100)	2.62 (1.32)
EDE EC[h]	168 (100)	1.35 (1.11)	84 (100)	1.16 (1.01)	84 (100)	1.53 (1.17)
EDE SC[i]	168 (100)	3.02 (1.28)	84 (100)	2.90 (1.23)	84 (100)	3.15 (1.33)
EDE WC[j]	168 (100)	2.48 (1.37)	84 (100)	2.30 (1.34)	84 (100)	2.67 (1.39)
EDI-2[k] BD[l]	168 (100)	38.29 (9.57)	84 (100)	36.98 (9.87)	84 (100)	39.61 (9.14)
EDI-2 BUL[m]	168 (100)	13.11 (5.69)	84 (100)	12.76 (5.06)	84 (100)	13.45 (6.27)
EDI-2 DFT[n]	168 (100)	27.99 (7.86)	84 (100)	26.92 (7.69)	84 (100)	29.06 (7.94)
WCS[o]	168 (100)	59.50 (17.82)	84 (100)	55.85 (16.18)	84 (100)	63.16 (18.71)
BDI[p]	168 (100)	12.50 (8.54)	84 (100)	11.74 (8.62)	84 (100)	13.26 (8.44)
BSI[q]	168 (100)	0.71 (0.55)	84 (100)	0.65 (0.56)	84 (100)	0.78 (0.53)
CIA[r] total score	168 (100)	12.61 (9.77)	84 (100)	10.98 (8.76)	84 (100)	14.25 (10.48)
Knowledge test	168 (100)	18.07 (2.61)	84 (100)	18.11 (2.57)	84 (100)	18.02 (2.67)

[a]T0: screening.
[b]T1: baseline.
[c]UW: underweight.
[d]ED: eating disorder.
[e]N/A: not applicable.
[f]EDE: Eating Disorder Examination.
[g]RS: Restraint.
[h]EC: Eating Concern.
[i]SC: Shape Concern.
[j]WC: Weight Concern.
[k]EDI-2: Eating Disorder Inventory-2.
[l]BD: Body Dissatisfaction.
[m]BUL: bulimia nervosa.
[n]DFT: Drive for Thinness.
[o]WCS: Weight Concerns Scale.
[p]BDI: Beck Depression Inventory.
[q]BSI: Brief Symptom Inventory.
[r]CIA: Clinical Impairment Assessment.

XSL•FO
RenderX

Effects of the Intervention on Primary and Secondary Outcomes

The analysis of the primary outcomes in the intention-to-treat sample using the Fisher exact test and the LOCF method revealed no significant differences between the IG and the CG at 12-month FU (EDE criterion, *P*=.99; weight gain criterion, *P*=.99). On the basis of the mixed model analyses of the primary outcomes, 79% (38/48) of the IG participants and 57% (33/58) of the CG participants with baseline EDE total scores above 1.87 showed a decrease in EDE total scores below 1.87 at 12-month FU, but this difference was again not significant (*P*=.19). In addition, 49% (15/31) of the IG participants who were underweight and 32% (10/32) of the CG participants who were underweight showed a BMI increase of at least 0.8 points. This difference was also not significant (*P*=.59; Table 3).

On the basis of mixed model analyses of the secondary outcomes, there was a significant group×time interaction in continuously measured BMI in the underweight subgroup between screening and 12-month FU. Furthermore, we found a significant group×time interaction for EDE Restraint, EDI Drive for Thinness, and EDI Body Dissatisfaction for all participants and for subjective and objective binge eating episodes for the subgroup of participants with symptoms of binge eating at 12-month FU. After the intervention, a significant group×time interaction was found for the EDE total score, EDE Restraint, and EDE Shape Concern; for EDI Drive for Thinness and EDI Body Dissatisfaction; for Weight Concerns; for the Beck Depression Inventory and the Clinical Impairment Assessment total scores; and for the knowledge test for all participants. For all secondary outcomes, the IG participants showed larger reductions (ie, more positive effects) than the CG participants, with effect sizes ranging from medium to large (Multimedia Appendix 1).

Table 3. Primary outcomes.

Outcome	Analysis	IG[a], n (%)	CG[b], n (%)	*P* value[c]	OR[d] (95% CI)[e]
BMI increase of at least 0.8 kg/m^2 in participants who were underweight (IG, N=31; CG, N=32)	LOCF[f]	10 (32)	11 (34)	.99	0.91 (0.32-2.59)
BMI increase of at least 0.8 kg/m^2 in participants who were underweight (IG, N=31; CG, N=32)	GLMM[g]	15 (49)	10 (32)	.59	2.04 (0.15-28.31)
EDE[h] total score below 1.87 (participants: IG, N=48; CG, N=58)	LOCF	24 (50)	28 (48)	.99	1.07 (0.50-2.30)
EDE total score below 1.87 (participants: IG, N=48; CG, N=58)	GLMM	38 (79)	33 (57)	.19	2.87 (0.60-13.67)

[a]IG: intervention group.

[b]CG: control group.

[c]*P* values correspond to the Fisher exact test for the last observation carried forward imputation and Wald tests for the generalized linear mixed model.

[d]OR: odds ratio.

[e]Odds ratios and 95% CIs were calculated in a logistic regression model.

[f]LOCF: last observation carried forward.

[g]GLMM: generalized linear mixed model for a binary outcome with logit link estimated with the unimputed data using a fixed effects model with γ=group, time, group×time, and a random effect for the repeated measurements. Response rates are marginal estimates shown as percentages.

[h]EDE: Eating Disorder Examination.

Effects of the Intervention on ED Cases

Prevention Effects

To assess the prevention effects of the intervention we compared all available IG and CG participants without any full-syndrome or subthreshold EDs at baseline (145/168, 86.3%) with respect to the newly emerging DSM-5 diagnoses at subsequent assessment points. After the intervention, data from 82.1% (119/145) of the participants were available; at 6-month FU, data from 80% (116/145) were available; and at 12-month FU, data from 72.4% (105/145) were available (Table 4).

After the intervention, among the 56 IG participants, 3 (5%) new-onset subclinical ED cases (n=2, 67%, subclinical AN and n=1, 33%, subclinical BN) emerged. Among the 63 CG participants, 5 (8%) new-onset cases (n=4, 80%, subclinical AN and n=1, 20%, subclinical BN) emerged (Fisher exact test, *P*=.72). At 6-month FU, among the 53 IG participants, 1 (2%) subclinical case of AN was observed, and among the 59 CG participants, 4 (7%) new-onset cases (n=1, 25%, full-syndrome case of AN and n=3, 75%, subclinical cases of AN) were observed (Fisher exact test, *P*=.37). Finally, at 12-month FU, among the IG participants (n=48), no new-onset cases emerged, whereas among the 55 CG participants, 3 (5%) new-onset cases (n=2, 67%, subclinical cases of AN and n=1, 33%, subclinical case of BN) were diagnosed (Fisher exact test, *P*=.13).

XSL•FO
RenderX

Table 4. Treatment and prevention effects.

Effect	After the intervention				FU6[a]				FU12[b]			
	IG[c], n/N (%)	CG[d], n/N (%)	*P* value[e]	OR[f] (95% CI)	IG, n/N (%)	CG, n/N (%)	*P* value	OR (95% CI)	IG, n/N (%)	CG, n/N (%)	*P* value	OR (95% CI)
Treatment	3/7 (43)	6/10 (60)	.64	0.50 (0.07-3.55)	1/7 (14)	4/9 (44)	.31	0.21 (0.02-2.52)	0/5 (0)	4/6 (67)	.06[g]	0.0 (0.0-1.1)[g]
Prevention	3/56 (5)	5/63 (8)	.72	0.66 (0.15-2.90)	1/53 (2)	4/59 (7)	.37	0.26 (0.03-2.46)	0/48 (0)	3/55 (5)	.13[g]	0.0 (0.0-1.9)[g]

[a]FU6: 6-month follow-up.

[b]FU12: 12-month follow-up.

[c]IG: intervention group.

[d]CG: control group.

[e]All *P* values are from the Fisher exact test.

[f]OR: odds ratio.

[g]The *P* value and the odds ratio of the 12-month FU treatment and prevention effects were estimated with a 0.5 correction of the underlying frequency table to reach a more stable estimate of the odds ratio in this extreme case of results.

Time to First Diagnosis

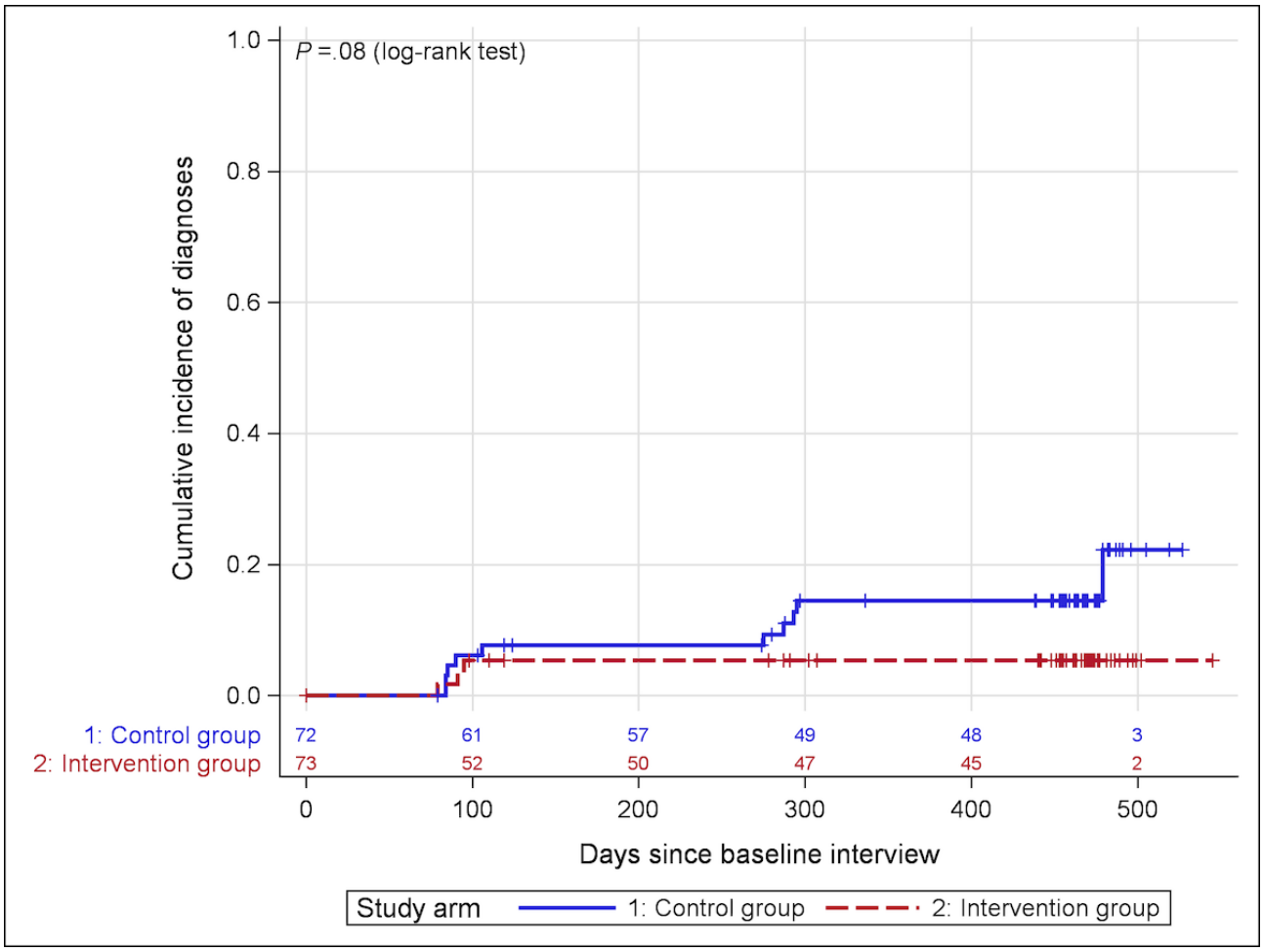
We also analyzed time to first onset of a newly emerging DSM-5–based full-syndrome or subclinical ED diagnosis, that is, days between baseline and the first FU assessment point where a participant met any of these diagnoses. Participants without any diagnosis at any assessment point were included as censored cases in the analysis with their time from baseline until the last available interview.

In the IG, new diagnoses only emerged around the intervention. In the CG, new diagnoses also occurred at 6-month FU and 12-month FU. The estimated 1-year cumulative incidence was 14.6% in the CG and 5.4% in the IG (*P*=.08; hazard ratio 0.335, 95% CI 0.092-1.216; Figure 2).

We calculated the ARR and NNT from the estimated survival curves. The ARR was 0.091 (95% CI 0.07-0.11), which translates into an NNT for the benefit of 11 participants to prevent at least one onset within 12 months after the program ends. The 95% CI of the NNT was 8.97-14.14.

XSL•FO
RenderX

Figure 2. Cumulative incidence curves for new-onset Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, diagnoses in the intervention and control groups. The respective numbers of participants at risk are provided below the incidence curves.



Treatment Effects

At baseline, of the 168 participants, 23 (13.7%; IG: n=11, 48%; CG: n=12, 52%) met DSM-5 criteria for full-syndrome and subclinical ED (n=9, 39%, cases of AN; n=12, 52%, cases of subclinical AN; and n=2, 9%, cases of subclinical BN).

After the intervention, of the 168 participants, 17 (10.1%) were available. Of the 10 CG participants, 6 (60%) still met the DSM-5 criteria, whereas of the 7 IG participants, 3 (43%) met the criteria (Fisher exact test, *P*=.64).

At 6-month FU, of the 168 participants, 16 (9.5%) were available. Of the 9 CG participants, 4 (57%) still met the DSM-5 criteria, whereas of the 7 IG participants, 1 (14%) met the criteria (Fisher exact test, *P*=.31). At 12-month FU, none of the IG participants (n=5) met the DSM-5 criteria, whereas of the 6 CG participants, 4 (66%) still met the criteria for full-syndrome and subclinical ED (n=1, 25%, case of AN; n=2, 50%, cases of subclinical AN; and n=1, 25%, case of BN; Fisher exact test, *P*=.06; Table 4).

We also analyzed changes in DSM-5 status for each pairwise transition between 2 assessment points within each study arm using McNemar tests (Table 5). In the IG, a significant proportion of the participants improved over time comparing baseline with 12-month FU and after the intervention with the FU assessment points. For the CG, such a trend could not be shown.

XSL•FO
RenderX

Table 5. Pairwise differences on eating disorder diagnoses. Control group results are displayed in the upper right triangle, intervention group results in the lower left triangle.

	Baseline		After the intervention		6-month FU[a]		12-month FU	
	Chi-square (*df*)	*P* value	Chi-square (*df*)	*P* value	Chi-square (*df*)	*P* value	Chi-square (*df*)	*P* value
Baseline	N/A[b]	N/A	0.1 (1)[c]	.74[c]	0.1 (1)[c]	.74[c]	0.2 (1)[c]	.65[c]
After the intervention	0.1 (1)[d]	.71[d]	N/A	N/A	0.1 (1)[c]	.74[c]	1.0 (1)[c]	.32[c]
6-month FU	3.6 (1)[d]	.06[d]	4.0 (1)[d]	.05[d]	N/A	N/A	0 (1)[c]	.99[c]
12-month FU	5.0 (1)[d]	.03[d]	5.0 (1)[d]	.03[d]	1.0 (1)[d]	.32[c]	N/A	N/A

[a]FU: follow-up.

[b]N/A: not applicable.

[c]For control group.

[d]For intervention group.

Program Adherence

Of the 84 IG participants, 11 (13%) never logged on to the program. Of the remaining 73 women, 53 (72%) opened at least half of the sessions and 47 (64%) accessed at least half of the intervention content. On average, all intervention participants, including those who never logged on to the program, opened 55.4% (SD 40.7; median 62.4%, IQR 87.3%) of the program pages and accessed 6.6 (SD 4.0; median 9%, IQR 8%) of the 10 sessions. Active participants who logged on to the program at least once on average opened 63.8% (SD 37.0%; median 83.3%, IQR 77.39%) of the program pages and accessed 7.6 (SD 3.3; median 10, IQR 6) of the 10 sessions.

Treatment Seeking

After the intervention, of the 84 CG participants, 1 (1%) reported having resumed treatment after study start. At 6- and 12-month FUs, of the 168 participants, 3 (1.8%) women (n=1, 33%, in the IG and n=2, 67%, in the CG) reported having resumed outpatient treatment for an ED.

Discussion

Principal Findings

This is the first study to evaluate the efficacy of an indicated preventive web-based intervention (SB-AN) for young women at risk for AN in reducing risk factors and symptoms of AN as well as syndrome progression of pre-existing, and onset of newly emerging subclinical syndromes of, AN compared with a wait-list CG. The intervention was specifically developed to target early symptoms and potential risk factors for AN that distinguishes SB-AN from other preventive interventions for EDs.

For our primary outcomes we found that the proportion of participants showing reductions in EDE interview total scores at 12-month FU below a score of 1.87 was 22% (79% vs 57%) larger in the IG than in the CG, but this difference was not significant. In addition, larger proportions of IG participants who were underweight showed a BMI increase of at least 0.8 points (49% vs 32%) compared with CG participants who were underweight, but this difference was again not significant. However, medium to large effects of the intervention were seen in several of the secondary outcomes of ED pathology: after the intervention, there were larger reductions in disordered eating attitudes such as restrained eating, shape concern, drive for thinness, body dissatisfaction, weight concern, depression, and clinical impairment in the IG than in the CG. At 12-month FU, the IG still showed larger reductions in restrained eating, drive for thinness, and body dissatisfaction than the CG. The intervention also proved effective in reducing symptoms of disordered eating: subjective and objective binge eating episodes for the subgroup of participants with symptoms of binge eating were significantly lower at 12-month FU, and continuously measured BMI of the subgroup of participants who were underweight was significantly larger in the IG than in the CG. Although fewer IG participants developed DSM-5 new-onset full-syndrome and subclinical EDs than the CG participants, this difference was not significant, probably because of the small numbers of overall cases. In addition, there was a trend for IG participants to develop new-onset cases only in the period around the intervention and not thereafter, whereas CG participants developed new-onset cases over the whole course of the study. Finally, there was also a trend regarding a treatment effect of the intervention, that is, a reduction in DSM-5–based ED diagnoses between baseline and 12-month FU. Fewer IG participants also resumed treatment for their ED during that time. Finally, the NNT at 12-month FU also indicates a benefit of the intervention.

Strengths and Weaknesses of the Study

Participants in this trial were specifically selected because they were at risk for AN based on either low or lower BMI and clinically elevated restrained eating scores. Stratification for weight group resulted in a mean BMI markedly lower than in other prevention trials for ED [12], with 37.5% (63/168) of the participants falling in the lower-weight to underweight range. The study sample was rather large, and the 12-month FU allowed for assessing the sustainability effects of the intervention. ED diagnoses and AN symptomatology were obtained using a well-validated clinical interview. We used central randomization, conducted the analyses blinded for IG, and controlled for missing data by multiple imputation in the analyses. The intervention itself is easily accessible for

XSL•FO
RenderX

participants and likely to be more cost-effective than face-to-face interventions.

However, there were also several limitations. The reach of the intervention [48] was limited: because of our relatively strict inclusion criteria, only 3.62% (168/4646) of the participants who had filled out screens and 19.5% (168/863) of those who had screened positive could be included. Of the 168 participants, 141 (83.9%) were students; hence, the generalizability of the results for more diverse populations remains unclear. Participants could not be blinded as in most other psychological interventions. Although adherence was comparable with other studies that included targeted preventive interventions for ED [49], 13% (11/84) of the randomized participants never logged on to the intervention, and on average, participants used only half to two-thirds of the program. One could assume that higher adherence may have more pronounced intervention effects; however, it may be worthwhile to test the effect of an abbreviated version of the intervention as part of future research.

Attrition was substantial, with rates of 36% (30/84) in the IG and 26% (22/84) in the CG, although again not unusually high compared with other preventive web-based interventions [49]. The selection of dichotomous primary outcomes (EDE total score and BMI differences) was based on uncontrolled effect sizes found in the pilot study [26], which had a much smaller sample size. These uncontrolled effect sizes may not be representative for a larger sample and may have overestimated the true effects. Women in the CG also improved over time on many outcomes, which may indicate that the extensive interviews with feedback on ED symptomatology itself also yielded effects. To control for these potential assessment effects, we would have needed to include a third assessment-only group. Nevertheless, effect sizes for secondary outcomes were all in the medium to large range. Health economic outcomes were not included in the study.

Comparison With Other Studies

To our knowledge, this is the first study to specifically target young women at risk for AN based on risk factors and early symptoms that may be uniquely predictive for AN onset. Most previous trials that included targeted preventive interventions used weight concern or body dissatisfaction as selection criteria. We found higher intervention effects after the intervention and at FU on ED risk factors (weight concern, shape concern, drive for thinness, dieting, and body dissatisfaction) and bulimic symptomatology compared with previous prevention trials [13,17-19]. The study is unique in promoting weight gain, at least in the lower-weight group. Many prevention studies have either not shown changes in weight [23,29] or targeted weight loss as an outcome [50,51]. This study is also unique in suggesting a preventive effect of late-onset ED cases in women at risk for AN. Only very few face-to-face and web-based ED prevention programs have significantly reduced ED onset [20-23,29,52]. With the exception of 1 case of AN in the study by Taylor et al [23], cases in these studies were subthreshold BN, BED, ED not otherwise specified or subthreshold ED, and full-syndrome BN and BED.

Implications

The results from this study suggest that the guided web-based intervention SB-AN is the first indicated prevention program to significantly reduce risk factors and symptom progression of AN symptoms such as restrained eating and low body weight. The intervention also shows promise for late onset of newly emerging full-syndrome and subclinical AN syndromes. Compared with studies reporting on ED onset, this study clearly succeeded in recruiting women with low or lower BMI and at higher risk for AN than participants of previous prevention trials. Consequently, the study yielded the highest rates of newly emerging subclinical and clinical AN. Although full-syndrome DSM-IV ED cases, including AN, were excluded from the study because of the revision of the weight criterion with the introduction of DSM-5, a relatively high rate (21/168, 12.5%) of women fulfilling criteria for full-syndrome and subclinical AN were unintentionally included in the study. For these women, the intervention proved beneficial in reducing symptom progression. The intervention could therefore be recommended as a specific intervention for women with low weight and high restrained eating to clinicians and health care providers who often underestimate these risks. However, the recruitment process also demonstrates that women at risk for AN are not easily enrollable in an intervention targeting low body weight, resulting in a considerable proportion (128/863, 14.8%) of participants screening positive who declined participation [53].

Conclusions

Although SB-AN overall proved moderately effective for women at risk for AN, future studies should try to improve the reach and uptake of the intervention, that is, examine its effects in more diverse populations; try to further increase motivation to change, especially in participants who are underweight; or examine whether reach can be increased by use of a mobile-based version of the intervention. Moderator variables—part of a separate analysis—might also shed more light on how to better tailor the intervention to increase reach and effectiveness for specific subgroups of participants. The consideration of health economic outcomes could demonstrate further benefits of the intervention compared with the costs for treating medical complications and treatment of symptoms of AN and subclinical AN.

Acknowledgments

This study was funded by the Else Kröner-Fresenius-Stiftung (2012_A01). The sponsor of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. The authors thank all women who participated in this study for their contributions. The authors also thank Professor Dr Konrad Reschke and Professor Dr Bernd Leplow for their support during the recruitment and assessment process in Leipzig and Halle (data monitoring: PvB, Sarah Bunzel, KKS; data management: Maik Hahmann; assessors: Anne Buchholz, Sarah Bunzel, Silke Elsäßer, Melanie Hassler, Sabrina Irrgang,

XSL•FO
RenderX

Gerda Keil, Francie Kriegel, Franziska Miksch, Annegret Neubauer, Angelika Schuster, Juliane Thieme, Pia Trübenbach, BV, Anna Wagner, Monique Zobel; student research assistants: Anne Buchholz, Sarah Bunzel, Lisa-Sophie Kant, Swantje Petersen, Ann-Kathrin Reinhardt, Charlotte Schnapka, Annette Spitzner). The corresponding author affirms that the 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 originally planned (and, if relevant, registered) have been explained.

Data Availability

Participant level data are available on request from the corresponding author. Participants' consent on data sharing was not obtained, but the presented data are anonymized and risk of identification is low.

Authors' Contributions

CJ and CBT designed and planned the study. CJ and KH wrote the first draft of the manuscript with support from BV and DG. BV was responsible for study conduct and coordination, supported by PvB and NE. DG and KH conducted the data analyses and, together with CJ, BV, and BT, were responsible for data interpretation. PvB, NE, and BV provided the intervention and were responsible for the assessments. All authors had access to the data, and CJ had the final responsibility for the decision to submit the paper for publication. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Secondary outcomes.
[DOCX File , 18 KB - jmir_v24i6e35947_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 408 KB - jmir_v24i6e35947_app2.pdf]

References

1. Zipfel S, Giel KE, Bulik CM, Hay P, Schmidt U. Anorexia nervosa: aetiology, assessment, and treatment. Lancet Psychiatry 2015 Dec;2(12):1099-1111. [doi: 10.1016/S2215-0366(15)00356-9] [Medline: 26514083]
2. Arcelus J, Mitchell AJ, Wales J, Nielsen S. Mortality rates in patients with anorexia nervosa and other eating disorders. A meta-analysis of 36 studies. Arch Gen Psychiatry 2011 Jul;68(7):724-731. [doi: 10.1001/archgenpsychiatry.2011.74] [Medline: 21727255]
3. Zipfel S, Löwe B, Reas DL, Deter HC, Herzog W. Long-term prognosis in anorexia nervosa: lessons from a 21-year follow-up study. Lancet 2000 Feb 26;355(9205):721-722. [doi: 10.1016/S0140-6736(99)05363-5] [Medline: 10703806]
4. Hay PJ, Claudino AM, Touyz S, Abd Elbaky G. Individual psychological therapy in the outpatient treatment of adults with anorexia nervosa. Cochrane Database Syst Rev 2015 Jul 27;2015(7):CD003909 [FREE Full text] [doi: 10.1002/14651858.CD003909.pub2] [Medline: 26212713]
5. National Guideline Alliance (UK). Eating Disorders: Recognition and Treatment. London, UK: National Institute for Health and Care Excellence; May 2017.
6. Steinhausen HC. The outcome of anorexia nervosa in the 20th century. Am J Psychiatry 2002 Aug;159(8):1284-1293. [doi: 10.1176/appi.ajp.159.8.1284] [Medline: 12153817]
7. Ágh T, Kovács G, Supina D, Pawaskar M, Herman BK, Vokó Z, et al. A systematic review of the health-related quality of life and economic burdens of anorexia nervosa, bulimia nervosa, and binge eating disorder. Eat Weight Disord 2016 Sep;21(3):353-364 [FREE Full text] [doi: 10.1007/s40519-016-0264-x] [Medline: 26942768]
8. Agras WS. The consequences and costs of the eating disorders. Psychiatr Clin North Am 2001 Jun;24(2):371-379. [doi: 10.1016/s0193-953x(05)70232-x] [Medline: 11416936]
9. Krauth C, Buser K, Vogel H. How high are the costs of eating disorders - anorexia nervosa and bulimia nervosa - for German society? Eur J Health Econ 2002;3(4):244-250. [doi: 10.1007/s10198-002-0137-2] [Medline: 15609150]
10. Jacobi C, Fittig E, Hütter K. Psychosocial risk factors for eating disorders. In: Agras WS, Robinson A, editors. The Oxford Handbook of Eating Disorders. 2nd ed. Oxford, UK: Oxford University Press; 2018:106-125.
11. Jacobi C, Hayward C, de Zwaan M, Kraemer HC, Agras WS. Coming to terms with risk factors for eating disorders: application of risk terminology and suggestions for a general taxonomy. Psychol Bull 2004 Jan;130(1):19-65. [doi: 10.1037/0033-2909.130.1.19] [Medline: 14717649]

XSL•FO
RenderX

12. Harrer M, Adam SH, Messner EM, Baumeister H, Cuijpers P, Bruffaerts R, et al. Prevention of eating disorders at universities: a systematic review and meta-analysis. Int J Eat Disord 2020 Jun;53(6):813-833. [doi: 10.1002/eat.23224] [Medline: 31943298]
13. Le LK, Barendregt JJ, Hay P, Mihalopoulos C. Prevention of eating disorders: a systematic review and meta-analysis. Clin Psychol Rev 2017 Apr;53:46-58. [doi: 10.1016/j.cpr.2017.02.001] [Medline: 28214633]
14. Pratt BM, Woolfenden SR. Interventions for preventing eating disorders in children and adolescents. Cochrane Database Syst Rev 2002;2002(2):CD002891 [FREE Full text] [doi: 10.1002/14651858.CD002891] [Medline: 12076457]
15. Stice E, Shaw H, Marti CN. A meta-analytic review of eating disorder prevention programs: encouraging findings. Annu Rev Clin Psychol 2007;3:207-231. [doi: 10.1146/annurev.clinpsy.3.022806.091447] [Medline: 17716054]
16. Watson HJ, Joyce T, French E, Willan V, Kane RT, Tanner-Smith EE, et al. Prevention of eating disorders: a systematic review of randomized, controlled trials. Int J Eat Disord 2016 Sep;49(9):833-862. [doi: 10.1002/eat.22577] [Medline: 27425572]
17. Beintner I, Jacobi C, Taylor CB. Effects of an Internet-based prevention programme for eating disorders in the USA and Germany--a meta-analytic review. Eur Eat Disord Rev 2012 Jan;20(1):1-8. [doi: 10.1002/erv.1130] [Medline: 21796737]
18. Loucas CE, Fairburn CG, Whittington C, Pennant ME, Stockton S, Kendall T. E-therapy in the treatment and prevention of eating disorders: a systematic review and meta-analysis. Behav Res Ther 2014 Dec;63:122-131 [FREE Full text] [doi: 10.1016/j.brat.2014.09.011] [Medline: 25461787]
19. Melioli T, Bauer S, Franko DL, Moessner M, Ozer F, Chabrol H, et al. Reducing eating disorder symptoms and risk factors using the internet: a meta-analytic review. Int J Eat Disord 2016 Jan;49(1):19-31. [doi: 10.1002/eat.22477] [Medline: 26607683]
20. Martinsen M, Bahr R, Børresen R, Holme I, Pensgaard AM, Sundgot-Borgen J. Preventing eating disorders among young elite athletes: a randomized controlled trial. Med Sci Sports Exerc 2014 Mar;46(3):435-447. [doi: 10.1249/MSS.0b013e3182a702fc] [Medline: 24549033]
21. Stice E, Marti CN, Spoor S, Presnell K, Shaw H. Dissonance and healthy weight eating disorder prevention programs: long-term effects from a randomized efficacy trial. J Consult Clin Psychol 2008 Apr;76(2):329-340 [FREE Full text] [doi: 10.1037/0022-006X.76.2.329] [Medline: 18377128]
22. Taylor CB, Bryson S, Luce KH, Cunning D, Doyle AC, Abascal LB, et al. Prevention of eating disorders in at-risk college-age women. Arch Gen Psychiatry 2006 Aug;63(8):881-888 [FREE Full text] [doi: 10.1001/archpsyc.63.8.881] [Medline: 16894064]
23. Taylor CB, Kass AE, Trockel M, Cunning D, Weisman H, Bailey J, et al. Reducing eating disorder onset in a very high risk sample with significant comorbid depression: a randomized controlled trial. J Consult Clin Psychol 2016 May;84(5):402-414 [FREE Full text] [doi: 10.1037/ccp0000077] [Medline: 26795936]
24. Stice E, Gau JM, Rohde P, Shaw H. Risk factors that predict future onset of each DSM-5 eating disorder: predictive specificity in high-risk adolescent females. J Abnorm Psychol 2017 Jan;126(1):38-51 [FREE Full text] [doi: 10.1037/abn0000219] [Medline: 27709979]
25. Völker U, Jacobi C, Trockel MT, Taylor CB. Moderators and mediators of outcome in Internet-based indicated prevention for eating disorders. Behav Res Ther 2014 Dec;63:114-121. [doi: 10.1016/j.brat.2014.09.008] [Medline: 25461786]
26. Ohlmer R, Jacobi C, Taylor CB. Preventing symptom progression in women at risk for AN: results of a pilot study. Eur Eat Disord Rev 2013 Jul;21(4):323-329. [doi: 10.1002/erv.2225] [Medline: 23520152]
27. Fairburn CG, Cooper Z. The eating disorder examination. In: Fairburn CG, Wilson GT, editors. Binge Eating: Nature, Assessment, and Treatment. 12th ed. New York, NY, USA: The Guilford Press; 1993:317-360.
28. Hilbert A, Tuschen-Caffier B. Eating disorder Examination - Questionnaire: Deutschsprachige Übersetzung. Münster: Verlag für Psychotherapie. 2006. URL: https://www.dgvt-verlag.de/e-books/2_Hilbert_Tuschen-Caffier_EDE-Q_2016.pdf [accessed 2022-04-28]
29. Jacobi C, Völker U, Trockel MT, Taylor CB. Effects of an Internet-based intervention for subthreshold eating disorders: a randomized controlled trial. Behav Res Ther 2012 Feb;50(2):93-99. [doi: 10.1016/j.brat.2011.09.013] [Medline: 22137366]
30. Miller WR, Rollnick S. Motivierende Gesprächsführung: Motivational Interviewing: 3. Auflage des Standardwerks in Deutsch. Freiburg im Breisgau, Germany: Lambertus-Verlag; 2015.
31. Hilbert A, Tuschen-Caffier B, Ohms M. Eating disorder examination: Deutschsprachige version des strukturierten essstörungsinterviews. Diagnostica 2004 Apr;50(2):98-106. [doi: 10.1026/0012-1924.50.2.98]
32. Killen JD, Taylor CB, Hayward C, Wilson DM, Haydel KF, Hammer LD, et al. Pursuit of thinness and onset of eating disorder symptoms in a community sample of adolescent girls: a three-year prospective analysis. Int J Eat Disord 1994 Nov;16(3):227-238. [doi: 10.1002/1098-108x(199411)16:3<227::aid-eat2260160303>3.0.co;2-l] [Medline: 7833956]
33. Garner DM. Eating disorder inventory-2 : professional manual. Odessa, FL, USA: Psychological Assessment Resources; 1991.
34. Garner DM, Olmstead MP, Polivy J. Development and validation of a multidimensional eating disorder inventory for anorexia nervosa and bulimia. Int J Eat Disord 1983;2(2):15-34. [doi: 10.1002/1098-108x(198321)2:2<15::aid-eat2260020203>3.0.co;2-6]

XSL•FO
RenderX

35. Grund K. Validierung der Weight Concerns Scale zur Erfassung von Essstörungen. Trier, Germany: Universität Trier; 2003.
36. Mond JM, Hay PJ, Rodgers B, Owen C, Beumont PJ. Temporal stability of the Eating Disorder Examination Questionnaire. Int J Eat Disord 2004 Sep;36(2):195-203. [doi: 10.1002/eat.20017] [Medline: 15282689]
37. Paul T, Thiel A. Eating Disorder Inventory-2 (EDI-2): Deutsche Version. Göttingen, Germany: Hogrefe Verlag; 2005.
38. Derogatis LR. The Brief Symptom Inventory (BSI). Administration, Scoring, and Procedures Manual-II. Minneapolis, MN, USA: National Computer Systems; 1992.
39. Hautzinger M, Keller F, Kühner C, Beck A. Beck Depressions-Inventar BDI II. Revision, 2. Auflage. Frankfurt, Germany: Pearson Assessment; 2009.
40. Franke GH. Brief Symptom Inventory von L.R. Derogatis (Kurzform der SCL-90-R)-Deutsche Version. Gottingen, Germany: Beltz Test; 2000.
41. Beck AT, Steer RA, Hautzinger M. Beck-Depressions-Inventar (BDI) Testhandbuch. Bern, Switzerland: Huber; 1995.
42. Bohn K, Doll HA, Cooper Z, O'Connor M, Palmer RL, Fairburn CG. The measurement of impairment due to eating disorder psychopathology. Behav Res Ther 2008 Oct;46(10):1105-1110 [FREE Full text] [doi: 10.1016/j.brat.2008.06.012] [Medline: 18710699]
43. Reas DL, Rø O, Kapstad H, Lask B. Psychometric properties of the clinical impairment assessment: norms for young adult women. Int J Eat Disord 2010 Jan;43(1):72-76. [doi: 10.1002/eat.20653] [Medline: 19260038]
44. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5). 5th ed. Washington, DC, USA: American Psychiatric Association; 2013.
45. Wu L. Mixed Effects Models for Complex Data. Boca Raton, FL, USA: CRC Press; 2009.
46. Rosnow RL, Rosenthal R. If you're looking at the cell means, you're not looking at only the interaction (unless all main effects are zero). Psychol Bull 1991;110(3):574-576. [doi: 10.1037/0033-2909.110.3.574]
47. Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. BMJ 1999 Dec 04;319(7223):1492-1495 [FREE Full text] [doi: 10.1136/bmj.319.7223.1492] [Medline: 10582940]
48. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]
49. Beintner I, Jacobi C, Taylor CB. Participant adherence to the Internet-based prevention program StudentBodies™ for eating disorders — a review. Internet Interv 2014 Mar;1(1):26-32. [doi: 10.1016/j.invent.2014.03.001]
50. Doyle AC, Goldschmidt A, Huang C, Winzelberg AJ, Taylor CB, Wilfley DE. Reduction of overweight and eating disorder symptoms via the Internet in adolescents: a randomized controlled trial. J Adolesc Health 2008 Aug;43(2):172-179 [FREE Full text] [doi: 10.1016/j.jadohealth.2008.01.011] [Medline: 18639791]
51. Jones M, Luce KH, Osborne MI, Taylor K, Cunning D, Doyle AC, et al. Randomized, controlled trial of an Internet-facilitated intervention for reducing binge eating and overweight in adolescents. Pediatrics 2008 Mar;121(3):453-462. [doi: 10.1542/peds.2007-1173] [Medline: 18310192]
52. Stice E, Rohde P, Shaw H, Marti CN. Efficacy trial of a selective prevention program targeting both eating disorders and obesity among female college students: 1- and 2-year follow-up effects. J Consult Clin Psychol 2013 Feb;81(1):183-189 [FREE Full text] [doi: 10.1037/a0031235] [Medline: 23231574]
53. Vollert B, von Bloh P, Eiterich N, Beintner I, Hütter K, Taylor CB, et al. Recruiting participants to an Internet-based eating disorder prevention trial: impact of the recruitment strategy on symptom severity and program utilization. Int J Eat Disord 2020 May;53(5):476-484. [doi: 10.1002/eat.23250] [Medline: 32112593]

Abbreviations

AN: anorexia nervosa
ARR: absolute risk reduction
BED: binge eating disorder
BN: bulimia nervosa
CG: control group
DSM: Diagnostic and Statistical Manual of Mental Disorders
ED: eating disorder
EDI: Eating Disorder Inventory
FU: follow-up
GLMM: generalized linear mixed model
IG: intervention group
KKS: Koordinierungszentrum für Klinische Studien
LOCF: last observation carried forward
NNT: number needed to treat
SB-AN: Student Bodies-Anorexia Nervosa
WCS: Weight Concerns Scale

XSL•FO
RenderX

Edited by A Mavragani; submitted 23.12.21; peer-reviewed by K Mathiasen, D Levine; comments to author 30.01.22; revised version received 03.02.22; accepted 04.02.22; published 02.06.22.

Please cite as:
Jacobi C, Vollert B, Hütter K, von Bloh P, Eiterich N, Görlich D, Taylor CB
Indicated Web-Based Prevention for Women With Anorexia Nervosa Symptoms: Randomized Controlled Efficacy Trial
J Med Internet Res 2022;24(6):e35947
URL: https://www.jmir.org/2022/6/e35947
doi:10.2196/35947
PMID:35653174

©Corinna Jacobi, Bianka Vollert, Kristian Hütter, Paula von Bloh, Nadine Eiterich, Dennis Görlich, C Barr Taylor. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 02.06.2022. 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

Evaluation of the Web-Based OutsidePlay-ECE Intervention to Influence Early Childhood Educators' Attitudes and Supportive Behaviors Toward Outdoor Play: Randomized Controlled Trial

Mariana Brussoni[1,2,3,4,5*], PhD; Christina S Han[1,4*], MA; Yingyi Lin[6*], MSc; John Jacob[1], MBA; Fritha Munday[5], MPH; Megan Zeni[7], MEd; Melanie Walters[8], MA; Eva Oberle[3,5], PhD

[1]Department of Pediatrics, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada
[2]British Columbia Injury Research & Prevention Unit, British Columbia Children's Hospital Research Institute, Vancouver, BC, Canada
[3]Human Early Learning Partnership, School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada
[4]BC Children's Hospital Research Institute, Vancouver, BC, Canada
[5]School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada
[6]Spatial Sciences Institute, University of Southern California, Los Angeles, CA, United States
[7]Faculty of Education, University of British Columbia, Vancouver, BC, Canada
[8]Child Care Services, University of British Columbia, Vancouver, BC, Canada
[*]these authors contributed equally

Corresponding Author:
Mariana Brussoni, PhD
Department of Pediatrics
Faculty of Medicine
University of British Columbia
4480 Oak St
Vancouver, BC, V6H 0B3
Canada
Phone: 1 604 875 3712
Email: mbrussoni@bcchr.ubc.ca

Abstract

Background: Outdoor play is critical to children's healthy development and well-being. Early learning and childcare centers (ELCCs) are important venues for increasing children's outdoor play opportunities, and early childhood educators' (ECE) perception of outdoor play can be a major barrier to outdoor play. The OutsidePlay-ECE risk-reframing intervention is a fully automated and open access web-based intervention to reframe ECEs' perceptions of the importance of outdoor play and risk in play and to promote a change in their practice in supporting it in ELCC settings. We grounded the intervention in social cognitive theory and behavior change techniques.

Objective: The aim of this study is to evaluate the effectiveness of the OutsidePlay-ECE web-based risk-reframing intervention.

Methods: We conducted a single-blind randomized controlled trial in Canada between December 2020 and June 2021 to test the OutsidePlay-ECE risk-reframing intervention for ECEs. We recruited participants using social media and mass emails through our partner and professional networks. We invited ECEs and administrators working in an ELCC, who can speak, read, and understand English. We randomized consented participants to the intervention or control condition. The participants allocated to the intervention condition received a link to the OutsidePlay-ECE intervention. Participants allocated to the control condition read the Position Statement on Active Outdoor Play, a 4-page document on research and recommendations for action in addressing barriers to outdoor play. The primary outcome was a change in tolerance of risk in play. The secondary outcome was goal attainment. We collected data on the web via REDCap (Vanderbilt University) at baseline and 1 week and 3 months after intervention.

Results: A total of 563 participants completed the baseline survey, which assessed their demographics and tolerance of risk in play. They were then randomized: 281 (49.9%) to the intervention and 282 (50.1%) to the control condition. Of these, 136 (48.4%) and 220 (78%) participants completed the baseline requirements for the intervention and control conditions, respectively. At 1 week after intervention, 126 (44.8%) and 209 (74.1%) participants completed follow-up assessments, respectively, and at 3

XSL•FO
RenderX

months after intervention, 119 (42.3%) and 195 (69.1%) participants completed the assessments, respectively. Compared with participants in the control condition, participants in the intervention group had significantly higher tolerance of risk in play at 1 week (β=.320; *P*=.001) and 3 months after intervention (β=.251; *P*=.009). Intention-to-treat analyses replicated these findings (β=.335; *P*<.001 and β=.271; *P*=.004, respectively). No significant intervention effect was found for goal attainment outcomes (odds ratio 1.124, 95% CI 0.335-3.774; *P*=.85).

Conclusions: The results of this randomized controlled trial demonstrated that the OutsidePlay-ECE intervention was effective and had a sustained effect in increasing ECEs' and administrators' tolerance of risk in play. It was not effective in increasing goal attainment.

Trial Registration: ClinicalTrials.gov NCT04624932; https://clinicaltrials.gov/ct2/show/NCT04624932

International Registered Report Identifier (IRRID): RR2-10.2196/31041

(J Med Internet Res 2022;24(6):e36826) doi:10.2196/36826

KEYWORDS

risk perception; risky play; risk reframing; early childhood education; tolerance of risk in play; child care; outside play; preschool

Introduction

Background

Children's opportunities for outdoor play have been decreasing over successive generations in many developed countries [1,2]. This decline is concerning because outdoor play is integral to children's physical and mental health [3,4]. The literature consistently illustrates that children who engage in outdoor play more often demonstrate increased physical activity [5-8], which has subsequent benefits for their physical health (eg, lower blood pressure, lower BMI, lower obesity, and healthy development of bone mineral density) [9-11]. In addition, outdoor play increases children's social competency and self-esteem [12,13].

Over the past few decades, several factors have been proposed to explain the overall decline in children's outdoor play. Increasingly, risk-averse cultural norms have resulted in ubiquitous adult supervision and playground equipment that offer little challenge [1,14-16]. This intersects with 21st century parenting norms. Influenced by anxieties about children's educational attainment and safety, risk-averse caregiving prioritizes children's achievement at the expense of play and encourages the heavy surveillance of children's activities [14,16-19].

To address these concerns and influence parents' perception of the importance of outdoor play, reduce their fears regarding the risks taken in play, and help them develop supportive parenting behaviors toward outdoor play, we built the OutsidePlay web-based risk-reframing intervention [20], where we first developed a module for parents (OutsidePlay-Parents). We found that this intervention was effective in changing the tolerance of risk in play among mothers of children aged 6 to 10 years [21]. Building on this work, we developed a new module for early childhood educators (ECEs; OutsidePlay-ECE) on the OutsidePlay intervention. Similar to the OutsidePlay-Parents module, the new module for the ECE uses social cognitive theory and behavior change techniques to address ECEs' attitudes and behaviors toward outdoor play and its inherent risks [22,23]. For example, self-reflection questions highlighted incompatible beliefs to help participants think differently about assumed barriers (eg, the benefit participants obtained from the risks they took in play as a child vs their focus on limiting risk for the children in their care). A full description of the intervention mapping approach we followed in its development has been previously published, and details of the intervention components are provided in the OutsidePlay-ECE Intervention section in Methods [24]. The OutsidePlay-ECE intervention is a fully automated and open access web-based intervention. The intervention mobilizes evidence-based behavior change techniques underpinned in social cognitive theory to change ECE's perception of outdoor play and practices and to facilitate behavior change by setting attainable goals in support of children's outdoor play in early learning childcare center (ELCC) settings [24].

For many children, most of their waking hours are spent in an ELCC, which can be an invaluable opportunity to provide them with high-quality opportunities for outdoor play, particularly for children who may have limited access to outdoor play in their home environments [25,26]. Unfortunately, this opportunity has not been fully leveraged because of various limiting factors. For instance, amid societal risk aversion trends, ECEs face many actual and perceived barriers that are primarily linked to safety concerns [27,28]. Canadian ELCCs require a license to operate and need to follow their provincial or territorial childcare licensing guidelines, which are often interpreted by ECEs in restrictive ways [27,28]. Furthermore, these barriers intersect with ECEs' cultural backgrounds and the level of confidence, knowledge, and experience in promoting and accommodating children's outdoor play, as well as support received (perceived and actual) from their colleagues and the ELCC administration [27-29].

Objectives

The aim of this study was to report the results of a randomized controlled trial (RCT) evaluating the effectiveness of the OutsidePlay-ECE intervention in increasing ECEs' and ELCC administrators' tolerance of risk in play and attain a behavior change goal related to providing outdoor play opportunities for children in their ELCC. Given the positive RCT results that we obtained on the effectiveness of the OutsidePlay-Parent intervention previously developed for parents [20], we hypothesized that participants completing the intervention for ECEs would have significantly greater increase in tolerance of

XSL•FO
RenderX

risk in play than those in the control condition at 1 week and 3 months after intervention. We also hypothesized that a greater proportion of ECEs in the intervention condition would attain their behavior change goal than those in the control condition at 1 week and 3 months after intervention.

Methods

Study Design

We used a single-blind (researchers and outcome assessors), 2-parallel condition RCT. We conducted this study between December 1, 2020, and June 30, 2021, in Canada and collected measures at baseline and at 1 week and 3 months after intervention. Our primary outcome was the change in tolerance of risk in play at either follow-up time point. The secondary outcome was the participants' goal attainment at either follow-up time point. The details on the intervention's theoretical framework, development, content, and the RCT study protocol can be found in the study by Brussoni et al [24]. We registered our RCT with the US National Institutes of Health Protocol Registration and Results System (ClinicalTrials.gov NCT04624932). We followed the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) guidelines for reporting this trial [30] (Multimedia Appendix 1).

Ethics Approval and Privacy

The RCT was approved by the University of British Columbia and Children's and Women's Health Center of British Columbia Research Ethics Board (H19-03644). We conducted the RCT (including the intervention) entirely on the web; thus, there was no human involvement, except when participants had inquiries and reported technical issues via email. The sole identifiable information collected was the participants' email addresses, which were required for sending allocated study material, follow-up measures, and reminders. We did not export the participants' email addresses for data analysis. We assigned each participant a study number that did not include identifiable personal information.

Participant Recruitment

We recruited participants between December 1, 2020, and March 15, 2021, via social media posts on Facebook, Facebook advertisements, Twitter, and Instagram. We also circulated mass recruitment emails through partners and professional networks. Potential participants completed a web-based survey in REDCap (Research Electronic Data Capture; Vanderbilt University), an electronic data capture tool hosted by and stored in the British Columbia Children's Hospital Research Institute server [31]. We included a complete description and procedure for the study in the web-based survey. This allowed participants to self-assess their eligibility and to consent on the web, with the capability of downloading the consent form, if they decided to participate in our study.

We temporarily halted participant recruitment and participation from December 18, 2020, to January 4, 2021, to accommodate the Christmas and New Year holidays, during which most ELCCs were closed. We made this decision to secure more valid participant responses to the goal attainment question in the 1 week after intervention follow-up time point, asking "Did you accomplish your goal?" which concerned their behavior in promoting children's outdoor play, specifically in their ELCC. We posted a message on the REDCap enrollment survey informing participants of this interruption and the date that RCT recruitment would resume.

Eligibility and Exclusion Criteria

Eligible participants were adult ECEs and ELCC administrators currently working in Canada, who could speak, read, and understand English. Given that ECEs work closely with, and are influenced by, ELCC administrators, we included both ECEs and ELCC administrators in this RCT. We did not have any exclusion criteria. We included participants deemed eligible according to the aforementioned eligibility criteria. As the RCT was conducted entirely on the web, computer and internet literacy was an implicit de facto eligibility criterion. Eligible and interested participants provided consent by downloading the consent form for review and selecting a checkbox to participate. We then invited the enrolled participants to complete the baseline survey, which included sociodemographic questions and a questionnaire that assessed participant tolerance of risk in play, and enter their email address.

Randomization and Blinding

We automatically randomized the enrolled participants who completed the baseline survey in REDCap to 1 of the 2 conditions: intervention and control. The participants had an equal (50%) likelihood of being assigned to each condition. We generated the randomization schedule beforehand by the Sealed Envelope service (Sealed Envelope Ltd) using randomized permuted blocks of sizes 4, 6, and 8. We then transferred the list to REDCap. We concealed allocation to the researchers during participant assignment and data analysis. We sent participants a unique link to their materials upon completion of the baseline survey and randomization. The nature of the intervention did not permit blinding of the participants. They may have intuited which condition they were allocated to, based on the details of the 2 conditions provided in the consent form: intervention (eg, web-based intervention) and control (ie, a PDF document). In addition, there has always been a risk that 2 or more participants from the same ELCC participated in the study and have become exposed to a condition different from theirs by communicating with their peers. We believe this scenario would be unlikely given that we recruited participants across Canada, and as such, we did not implement any precaution.

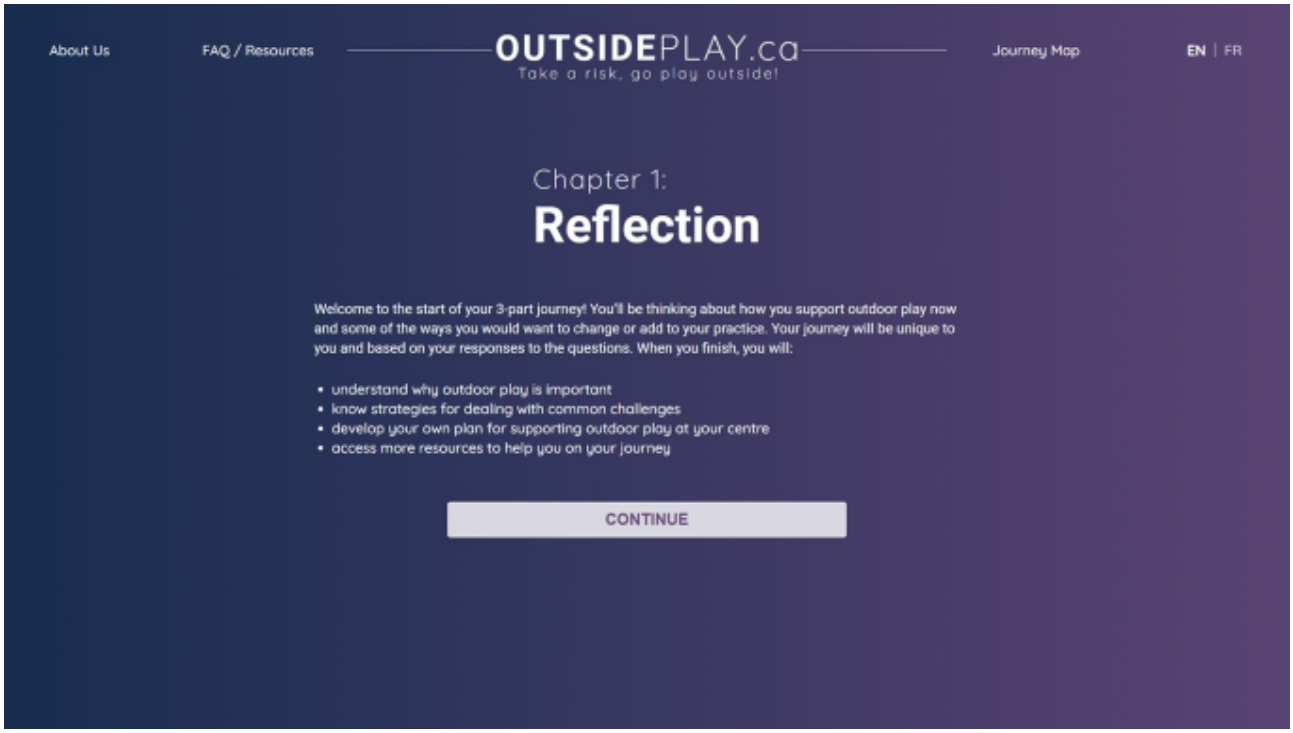
OutsidePlay-ECE Intervention

The goal of the OutsidePlay-ECE intervention is to reframe ECEs' perception of the importance of outdoor play and its inherent risks and promote a change in their practice in supporting children's outdoor play in ELCC settings. We designed OutsidePlay-ECE for ECE as a fully automated web-based risk-reframing intervention. It consists of 3 chapters, which are guided by the animated character of the first author (MB) and include self-reflection questions. We (the study authors) developed the OutsidePlay-ECE intervention following the intervention mapping process [32]. Social cognitive theory [33] provides a theoretical basis for the selection of behavior

XSL•FO
RenderX

change techniques [22] adopted in the intervention. According to social cognitive theory, individuals are motivated to change behavior when their self-efficacy is high (eg, "I am capable of supporting more opportunities for outdoor play for children at my center"), are dissatisfied with their current state (eg, "I do not provide children at my center with enough opportunities for outdoor play"), and believe that changing their behavior will lead to the preferred outcome (eg, "Outdoor play will benefit children") [32]. The OutsidePlay-ECE landing page is shown in Figure 1, and the complete screenshots of the OutsidePlay-ECE intervention are shown in Multimedia Appendix 2.

Figure 1. Screenshots of the OutsidePlay-ECE intervention landing page.



Our protocol paper [20] provides a full description of OutsidePlay-ECE intervention and details regarding its development. In brief, this intervention consists of 3 chapters, guiding participants through a journey. Chapter 1 prepares participants to position themselves on why they want to promote children's outdoor play at their center by reflecting on their own childhood play and considering how children play in their center. Chapter 2 includes a series of videos presenting common challenging scenarios: (1) communicating with parents and caregivers, (2) rough-and-tumble play, (3) play at heights, (4) conflict resolution, (5) play with loose parts, and (6) play at speed. These scenarios evolved based on the 6 categories of risky play [34] and were then selected based on ECE feedback. For each scenario, the ECE must decide regarding what they allow children to do. For example, in the rough-and-tumble play scenario, 2 children start sword fighting with sticks, and the ECE is asked to decide whether to stop the children or talk to them about consent and safety. On the basis of the ECE's choice, the video continues to show the outcome of the ECE's decision. Each scenario includes a summary video by an experienced ECE, highlighting the main take-home messages of that scenario. The objective of Chapter 2 is to reflect on the barriers and challenges ECEs often encounter at their center while accommodating or promoting children's outdoor play and provide them with clear actions to address them. The last chapter summarizes the learning in the previous 2 chapters and invites participants to think of an achievable goal and create a plan to accomplish it. Their journey, including their goals and plans, can be downloaded or sent to their email.

The OutsidePlay-ECE intervention focused on social cognitive theory constructs: outcome expectation; knowledge, barriers and opportunities; observational learning; self-efficacy; behavioral skills; and intentions [20]. The behavior change techniques related to information about consequences (health, social, and emotional), social comparison, framing and reframing and incompatible beliefs, problem solving, instructions on how to perform the behavior, demonstration of the behavior, social comparison, comparative imagining of future outcomes and goal setting (behavior and outcome), problem solving, action planning, and credible sources [20]. Details regarding the constructs and their associated behavior change techniques are outlined in our protocol paper [20].

We soft launched the OutsidePlay-ECE intervention on December 1, 2020, for the RCT and froze the content during the RCT (ie, we did not make any changes) and analysis. This means that we released the intervention only for RCT purposes with no publicity push before its full launch to the public. The participants were allocated to the intervention condition with a link to the OutsidePlay-ECE intervention. It could be completed in up to 100 minutes, depending on the participants' movements through each chapter. Participants could also return to the intervention at their convenience, picking up from where they left off previously, provided that they did not delete their browser cache and http cookies. REDCap sent out a maximum

XSL•FO
RenderX

of 3 automated reminders at 24, 48, and 60 hours after completion of the baseline survey and at the 1 week and 3 months after intervention follow-ups.

Comparison Condition

We asked participants in the control condition to review a PDF of the Position Statement on Active Outdoor Play, a 4-page document with information on research and recommendations for action in addressing barriers to outdoor play [4,35]. We estimated that the participants took 15 to 20 minutes to read through the document. We did not send out automated reminders to participants in the control condition because once they opened the Position Statement on Active Outdoor Play, and closed their survey at that point, we considered that they completed the baseline requirement. However, we reminded them up to 3 times (24, 48, and 72 hours), if they did not finish the survey measures at any follow-up time point (ie, 1 week and 3 months after intervention).

Outcome Measures

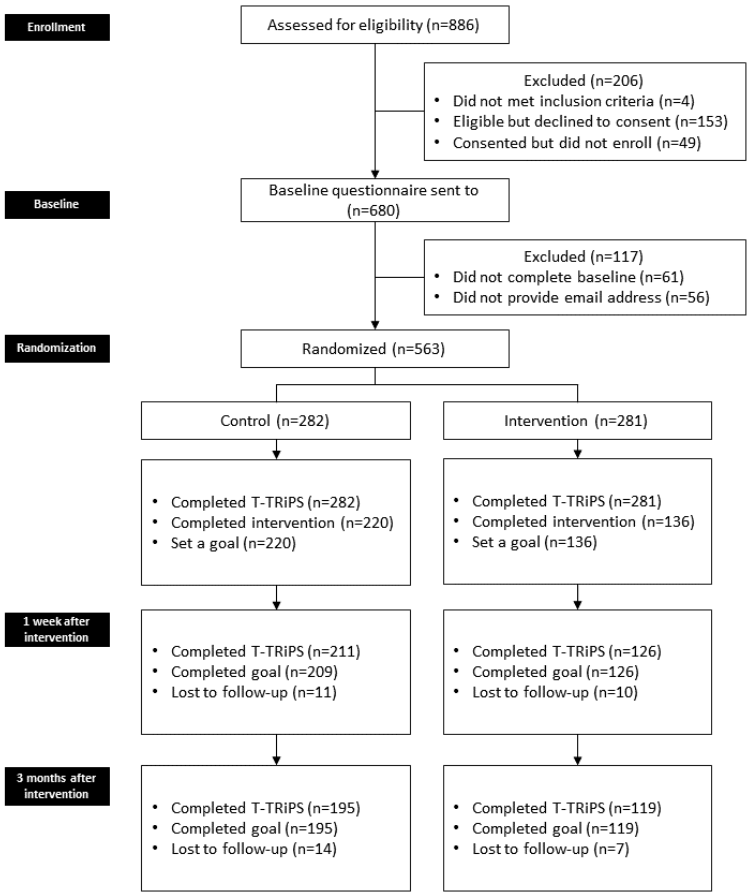
Our primary outcome measure was change in the total score on the Teacher Tolerance of Risk in Play Scale (T-TRiPS), a validated, reliable 26-item measure with dichotomous yes or no responses on items that reflect the 6 categories of risky play (great heights, high speed, dangerous tools, dangerous elements, rough-and-tumble, and disappear or get lost) [36] proposed by Sandseter [34]. The T-TRiPS is a modified version of the Tolerance of Risk in Play Scale for parents [37], which measures teachers' perceptions of risk. We assessed the psychometric properties of T-TRiPS in our sample using Rasch analysis, which considers the respondent's ability to choose a *correct* (in the case of this study, the *yes* response) item and the difficulty of each item [38]. Rasch analysis converts categorical responses to interval data. This analysis was conducted using the mirt package in R software (R Core Team, version 4.0.0) [39,40]. Rasch analysis of the baseline data (563 participants completed T-TRiPS; Figure 2) resulted in dropping 1 item ("Do you wait to see how well the children in your center manage challenges before getting involved?") owing to local dependence, such that this item was highly correlated with several other items on the T-TRiPS. The remaining 25 items resulted in the following model fit: root mean square error of approximation=0.060 (90% CI 0.056-0.065), standardized root mean square residual=0.101, Tucker-Lewis index=0.899, comparative fit index=0.899, and empirical reliability=0.851. θ standardized scores from the Rasch analysis of the final 25-item T-TRiPS ranged from −3.839 to 3.847, with a mean of −0.000 (SD 1.224). A higher standardized score indicated a higher tolerance of risk in play.

Our secondary outcome measure was self-reported behavior change, measured by participants' self-reported progress in attaining the goal they set for themselves. At each follow-up time point, participants were reminded of their goal and asked, "Did you accomplish your goal?" with dichotomous yes or no responses.

We assessed the primary outcome measure at baseline and at 1 week and 3 months after intervention. We only assessed the secondary outcome measure at 1 week and 3 months after intervention because at baseline, they could not have accomplished a goal they had not yet set. We paid the participants US $25 via electronic transfer upon completing the baseline questionnaire and allocated intervention. We then paid them US $16 at each of the follow-up time points at 1 week and 3 months after intervention. In addition, we issued participants in the intervention condition a professional development certificate for 100 minutes upon completion of the OutsidePlay-ECE intervention.

XSL•FO
RenderX

Figure 2. Study flow diagram.



Statistical Analyses

We conducted all statistical analyses in Stata (StataCorp, version 15) [41].

Power

For a sample size of 206 ECEs and ELCC administrators in total, a linear mixed model examining the impact of intervention relative to control, including an interaction with time, was calculated to have 80% power at a *P*=.05 level of significance to detect a difference of 0.75 between the intervention and control conditions on the T-TRiPS when the SD is 1.82, and the correlation between repeated observations is 0.75. From our previous work [20,21], we anticipated requiring 324 complete baseline requirements among ECEs and ELCC administrators who would then be randomized into the 2 conditions. We assumed a 75% retention rate at the 1 week after intervention follow-up time point (n=242) and an 85% retention rate at our 3 months after intervention follow-up time point, which would result in a final sample of 206 ECEs, corresponding to 103 in each condition.

Descriptive Analysis

To compare the raw outcome differences between conditions at each time point, for continuous outcomes (TRiPS scores), we used 1-way ANOVA or Kruskal-Wallis *H* test (if variance is not equal between conditions). For categorical outcomes (goal attainment), we used the chi-square test. Significance level was set at *P*=.05.

Effect of the Intervention

We concluded that linear and generalized linear mixed effects models with random intercepts and unstructured covariance were fit to analyze the effects of the intervention on T-TRiPS scores and goal attainment, respectively. In other words, we used the mixed effects regression analysis to examine (1) whether T-TRiPS scores changed between 1 week and 3 months after intervention and (2) whether these changes were greater in the intervention condition (ie, the OutsidePlay-ECE intervention) than in the control condition. We used intent-to-treat analysis of T-TRiPS scores that used the last observation carried forward as the method of imputation, because these participants only completed the baseline survey and did not complete the intervention, it is reasonable to expect their T-TRiPS scores to remain the same throughout the study period. Unstandardized (ie, raw) β coefficients were reported, which were interpreted as the change in T-TRiPS scores when comparing the intervention condition with the control condition at baseline.

Similar to the T-TRiPS analysis, we conducted a generalized mixed effects regression analysis to examine the effect of the intervention on goal attainment, when comparing the control condition at the 1 week after intervention follow-up time point with the intervention condition at the 3 months after intervention follow-up time point. An intention-to-treat analysis of goal attainment was not performed because of the absence of baseline data. To establish a goal, participants had to complete either interventions (eg, the OutsidePlay-ECE intervention or the

XSL•FO
RenderX

Position Statement on Active Outdoor Play). Consequently, there was no basis for imputing the values of goal attainment. We calculated odds ratios, which were interpreted as the odds of attaining goals for the intervention condition at the 3 months after intervention follow-up time point, divided by the odds of the control condition at 1 week after intervention (relative effect size). We also calculated the absolute effect size, that is, risk differences and the probability of attaining a goal in the intervention group minus the probability in the control condition.

Results

Overview

Figure 2 shows the flow diagram of the study. A total of 563 ECEs were randomly allocated to 1 of the 2 intervention conditions using REDCap. Of these, 420 (74.6%) completed the baseline requirement, which included completion of the baseline survey and the intervention and setting up their goals. Although randomization produced roughly equal numbers of participants allocated to each condition, the intervention condition experienced the most dropouts (145/281, 51.6%) at the time of baseline when completing the intervention as compared to dropouts in the control condition (62/282, 21.9%). The intervention condition involved more time commitment from the participants, as completing the OutsidePlay-ECE intervention typically took up to 100 minutes compared with 15 to 20 minutes for the control condition. However, of the participants completing the intervention, we only lost 5.0% (11/220) and 6.6% (14/209) of the participants to follow-up at 1 week and 3 months after intervention respectively, versus 7.3% (10/136) and 5.5% (7/126) of the participants in the control condition. We confirmed fidelity to the intervention through a review of participants' responses within each chapter of the intervention.

The University of British Columbia/Children's and Women's Health Center of British Columbia Research Ethics Board categorized the intervention as low risk and not associated with any harm. No privacy breaches or technical problems affected the participants. Although we tried to accommodate participants' varying internet bandwidths by automatically adjusting the media resolution (eg, high or low), this did not resolve issues caused by some users accessing the intervention from old or incompatible devices.

Sample Characteristics

We included baseline sociodemographic data from 563 ECEs and ELCC administrators who were randomized to a condition in our analyses (baseline characteristics between the 2 conditions are presented in Table 1). We did not observe any statistically significant differences between the baseline conditions with regard to all sociodemographic characteristics. We also compared sociodemographic data between the control and intervention conditions at the 1 week and 3 months after intervention follow-up time points, and no statistically significant differences were found.

We compared the sociodemographic characteristics among those who completed the baseline survey and between those who were randomized (N=563) and those who were not randomized (N=56), as these participants did not provide email address to proceed with the study and found that those who were randomized were more likely to be female (544/563, 96.6% vs 51/56, 91.1%, respectively; *P*=.04), less likely to be ECEs (392/563, 69.6% vs 48/56, 85.7% respectively; *P*=.04), more likely to be ELCC administrators (165/563, 29.3% vs 8/56, 14.3% respectively; *P*=.04), worked for a longer time in the field (mean 10.13, SD 9.33 vs mean 7.27, SD 7.64 years, respectively; *P*=.03), were more likely to be from British Columbia (258/560, 46.1% vs 15/53, 28.3% respectively; *P*=.007), less likely to be from Ontario (130/560, 23.2% vs 23/53, 43.4% respectively; *P*=.007), worked at a center with fewer children (for the number of children between 1 and 24: 186/557, 33.4% vs 8/51, 15.7% respectively; *P*=.03 and for the number of children ≥49: 217/557, 39% vs 26/51, 51% respectively; *P*=.03), and fewer staff (for the number of staff between 1 and 5: 202/551, 36.7% vs 10/52, 19.2% respectively; *P*=.006 and for the number of staff ≥13: 173/551, 31.4% vs 27/52, 51.9% respectively; *P*=.006). We did not find any statistical differences in other sociodemographic characteristics.

XSL•FO
RenderX

Table 1. Baseline characteristics between the 2 intervention conditions.

Characteristics of participants who completed the baseline survey	Control (n=282)	Intervention (n=281)	Total (N=563)
Sex (N=563), n (%)			
Male	8 (2.8)	8 (2.9)	16 (2.8)
Female	272 (96.4)	272 (96.8)	544 (96.6)
Other	2 (0.7)	0 (0)	2 (0.4)
Prefer not to answer	0 (0)	1 (0.4)	1 (0.2)
Age (years; N=563), n (%)			
19 to 24	26 (9.3)	33 (11.8)	59 (10.5)
25 to 30	72 (25.6)	55 (19.6)	127 (22.6)
31 to 40	71 (25.3)	86 (30.7)	157 (28)
41 to 50	66 (23.5)	64 (22.9)	130 (23.2)
51 to 60	32 (11.4)	36 (12.8)	61 (12.1)
61 to 70	13 (4.6)	6 (2.1)	19 (3.4)
≥71	1 (0.4)	0 (0)	1 (0.2)
Prefer not to answer	1 (0.4)	1 (0.4)	2 (0.4)
Language (N=563), n (%)			
English	263 (93.3)	261 (92.9)	524 (93.1)
Other[a]	19 (6.7)	20 (7.1)	39 (6.9)
Role (N=563), n (%)			
ECE[b]	203 (72)	189 (67.3)	392 (69.6)
ECE administrator	75 (26.6)	90 (32)	165 (29.3)
Other[c]	4 (1.4)	2 (0.7)	6 (1.1)
Working in the field (years; N=530), mean (SD)	10.22 (9.51)	10.05 (9.16)	10.14 (9.33)
Working at the current center (years; N=530), mean (SD)	6.09 (7.41)	5.18 (5.98)	5.63 (6.74)
Province of employment (N=560), n (%)			
Alberta	6 (2.1)	7 (2.5)	13 (2.3)
British Columbia	129 (45.9)	129 (46.2)	258 (46.1)
Manitoba	3 (1.1)	2 (0.7)	5 (0.9)
New Brunswick	52 (18.5)	44 (15.8)	96 (17.1)
Newfoundland and Labrador	8 (2.9)	12 (4.3)	20 (3.6)
Nova Scotia	13 (4.6)	9 (3.2)	22 (3.9)
Ontario	64 (22.8)	66 (23.7)	130 (23.2)
Prince Edward Island	0 (0)	2 (0.7)	2 (0.4)
Quebec	2 (0.7)	4 (1.4)	6 (1.1)
Saskatchewan	4 (1.4)	4 (1.4)	8 (1.4)
Whether the center is licensed (N=544), n (%)			
Yes	266 (97.4)	264 (97.4)	530 (97.4)
No	7 (2.6)	7 (2.6)	14 (2.6)
Children at the center (N=557), n (%)			
Small: 1 to 24	91 (32.7)	95 (34.1)	186 (33.4)
Medium: 25 to 48	76 (27.3)	78 (28.0)	154 (27.7)
Large: ≥49	111 (39.9)	106 (38.0)	217 (39.0)

XSL•FO
RenderX

Characteristics of participants who completed the baseline survey	Control (n=282)	Intervention (n=281)	Total (N=563)
Staff at the center (N=551), n (%)			
Small: 1 to 5	100 (36.2)	102 (37.1)	202 (36.7)
Medium: 6 to 12	85 (30.8)	91 (33.1)	176 (31.9)
Large: ≥13	91 (33.0)	82 (29.8)	173 (31.4)
Whether the center has a designated outdoor space for children (N=557), n (%)			
Yes	270 (97.1)	270 (96.8)	540 (97)
No	8 (2.9)	9 (3.2)	17 (3)
Quality of the center's outdoor space for children (N=539), n (%)			
Very good	96 (35.6)	88 (32.7)	184 (34.1)
Good	95 (35.2)	110 (40.9)	205 (38.0)
Acceptable	65 (24.1)	60 (22.3)	125 (23.2)
Poor	12 (4.4)	11 (4.1)	23 (4.3)
Very poor	2 (0.7)	0 (0)	3 (0.4)
Time children spent playing outdoors at the center (hours; N=556), mean (SD)	2.01 (1.15)	2.14 (1.13)	2.08 (1.14)
Feeling supported by colleagues in general (N=559), n (%)			
Yes	253 (89.7)	250 (90.3)	503 (90.0)
No	11 (3.9)	14 (5.1)	25 (4.5)
Feeling partially supported	8 (2.8)	4 (1.4)	12 (2.1)
N/A[d]	10 (3.5)	9 (3.2)	19 (3.4)

[a]Arabic (n=3), Cantonese (n=3), Chinese (n=2), Croatian (n=2), Gujarati (n=2), Hindi (n=1), Hungarian (n=1), Korean (n=6), Mandarin (n=1), Minnan (a Chinese dialect; n=1), Nepali (n=1), Punjabi (n=4), Serbian (n=1), Sinhala (n=1), Sinhalese (n=1), Slovak (n=2), Spanish (n=2), Tagalog (n=2), Tamil (n=2), Turkish (n=1), and Dutch (n=1).

[b]ECE: early childhood educator.

[c]Includes childcare provider consultant (n=1), child and youth care (n=1), no ECE (n=2), classroom teacher (n=1), and instructor at college (n=1).

[d]N/A: not applicable; for example, the participant is the only staff member.

Primary Outcome: T-TRiPS

Table 2 presents the description of T-TRiPS scores by intervention conditions and time points, without accounting for time effects, the interaction of intervention by time effects. We did not find any statistical differences in T-TRiPS scores between different conditions at baseline among those who reported T-TRiPS scores or among those who completed the intervention at baseline. At both the 1 week and 3 months after intervention follow-up time points, T-TRiPS scores were significantly higher in the intervention condition than in the control condition (1 week: mean −0.156, SD 1.304, for the control condition and mean 0.262, SD 1.117, for the intervention condition, *P*=.003; and 3 months: mean −0.118, SD 1.400, for the control condition and mean 0.200, SD 1.211, for the intervention condition, *P*=.04).

Table 3 describes the findings of the mixed effects regression analysis, considering intervention effects, time effects, and iteration of intervention by time effects. Participants who completed the intervention condition had significantly higher T-TRiPS scores than those who completed the control condition at 1 week (0.320, 95% CI 0.135-0.505; *P*=.001) and 3 months (0.251, 95% CI 0.062-0.440; *P*=.009) after intervention, indicating sustained change. Results of the intention-to-treat analyses for the effects of the intervention on T-TRiPS scores largely replicated the aforementioned analyses, indicating that ECEs and ELCC administrators in the intervention condition were significantly more likely to increase their T-TRiPS scores at 1 week (0.335, 95% CI 0.156-0.514; *P*<.001) and 3 months (0.271, 95% CI 0.088-0.454; *P*<.004) after intervention compared with those in the control condition.

Table 2. Description of the Teacher Tolerance of Risk in Play Scale scores by intervention conditions and time points.

Evaluation period	Sample size, N	Control, mean (SD)	Intervention, mean (SD)	*P* value for 1-way ANOVA
Baseline	563	0.040 (1.207)	−0.040 (1.243)	.44
Completed intervention	356	0.017 (1.211)	0.123 (1.196)	.86
1 week after intervention	337	−0.156 (1.304)	0.262 (1.117)	.003
3 months after intervention	314	−0.118 (1.400)	0.200 (1.211)	.04

XSL•FO
RenderX

Table 3. Mixed effects regression analysis for the Teacher Tolerance of Risk in Play Scale (T-TRiPS) θ score.

Regression and condition comparisons	Coefficients (95% CI)	*P* value for coefficients	*P* value for joint test of main effects
Raw T-TRiPS θ scores (N=356, for those who were randomized to a condition, completed baseline T-TRiPS measure, and completed the intervention)[a]			
Intervention effects: intervention versus control	0.100 (−0.169 to 0.369)	.47	.02
Time effects			.99
1 week versus baseline	−0.154 (−0.267 to −0.041)	.007	
3 months versus baseline	−0.124 (−0.240 to −0.008)	.04	
Intervention by time effects			.002
Intervention versus control by 1 week versus baseline	0.320 (0.135 to 0.505)	.001	
Intervention versus control by 3 months versus baseline	0.251 (0.062 to 0.440)	.009	
Intention-to-treat analysis (imputed T-TRiPS θ scores; N=563, for those who were randomized to a condition and completed baseline T-TRiPS measure)			
Intervention effects: intervention versus control	0.019 (−0.217 to 0.254)	.88	.054
Time effects			.96
1 week versus baseline	−0.156 (−0.268 to −0.044)	.006	
3 months versus baseline	−0.126 (−0.241 to −0.011)	.03	
Intervention by time effects			<.001
Intervention versus control by 1 week versus baseline	0.335 (0.156 to 0.514)	<.001	
Intervention versus control by 3 months versus baseline	0.271 (0.088 to 0.454)	.004	

[a]Italicization denotes two separate sets of analysis.

Secondary Outcome: Goal Attainment

We asked participants to think of one tangible and achievable goal and created a feasible plan to accomplish it. Participant goals varied widely from "I could add more building tools" to "Bring this topic and learning opportunity up with my colleagues at our daily check-in." Table 4 presents the description of goal attainment by condition and time point, without accounting for time effects and the interaction between time and intervention effects. We did not find any statistically significant differences in the secondary outcome, goal attainment, between the 2 conditions at either 1 week after intervention (141/209, 67.5% for the control condition and 94/126, 74.6% for the intervention condition; *P*=.17) or 3 months after intervention (163/195, 83.6% for the control condition and 106/119, 89.1% for the intervention condition; *P*=.18).

Table 5 presents the results of the generalized mixed effects regression analysis. There was no statistical difference in goal attainment between participants in the intervention and control condition at 3 months after intervention compared with 1 week after intervention (odds ratio 1.124, 95% CI 0.335-3.774; *P*=.85).

Table 4. Goal attainment by intervention condition and time point.

Evaluation period and goal attainment	Control, n (%)	Intervention, n (%)	Sample size, N	*P* value for chi-square
1 week after intervention			335	.17
Yes	141 (67.5)	94 (74.6)		
No	68 (32.5)	32 (25.4)		
3 months after intervention			314	.18
Yes	163 (83.6)	106 (89.1)		
No	32 (26.4)	13 (10.9)		

XSL•FO
RenderX

Table 5. Results of the mixed effects regression analysis for goal attainment by intervention condition and time.

Regression and condition comparisons[a]	Relative effect size		Absolute effect sizes	
	Odds ratios (95% CI)	*P* value	Risk differences (95% CI)	*P* value
Intervention effects: intervention versus control	2.046 (0.740 to 5.655)	.17	0.071 (−0.019 to 0.162)	.12
Time effects: 3 months versus 1 week	5.749 (2.664 to 12.407)	<.001	0.158 (0.098 to 0.218)	<.001
Intervention by time effects: intervention versus control by 3 months versus 1 week	1.124 (0.335 to 3.774)	.85	−0.018 (−0.115 to 0.079)	.72

[a]N=335, who were randomized to a condition and set a goal at baseline and completed goal attainment measures at a follow-up time point.

Discussion

Principal Findings

Our RCT tested the effectiveness of the web-based OutsidePlay-ECE intervention in changing ECEs' and ELCC administrators' tolerance of risk in play and the attainment of their personalized goals for change to support children's outdoor play within the ELCC. The RCT results partially support our hypotheses. ECEs and ELCC administrators receiving the intervention reported significantly higher increases in their tolerance of risk in play at 1 week after intervention than participants in the control condition. These differences remained significant at 3 months after intervention. There were no significant differences in goal attainment. These results are consistent with the findings of a previous RCT testing the OutsidePlay-Parent intervention, which also found significantly greater increases in tolerance of risk in play for intervention versus control participants at 1 week and 3 months after intervention [20].

There are several possibilities for the lack of a significant effect of the intervention on the secondary outcome, goal attainment. Findings reported for the OutsidePlay-Parent intervention [21] were similar, and it was postulated that the null finding may have resulted from the participants' goals being too ambitious or not sufficiently actionable. To limit participants from developing overly ambitious or less actionable goals, a short clip of video in the last chapter was included to encourage participants to consider one thing that they can do to support children's outdoor play: "It shouldn't be something too big or complicated. Make sure it is concrete and achievable—that you don't feel overwhelmed by it." In addition, we provided some basic actionable goals and stressed that this was meant to be the first step in a journey toward change. We recognized that addressing the many barriers and challenges to outdoor play in ELCC environments would require complex intervention at multiple organizational levels (eg, the individual, relationship with colleagues, and licensing regulations) [42,43]. The OutsidePlay-ECE intervention was designed to open their minds to a different way of thinking and approach, which would be the first step in the process. Although participants may have thought that they were setting a manageable goal, they may have subsequently realized that it was more ambitious than anticipated and evaluated their actions as insufficient. Furthermore, it is possible that although the intervention successfully opened their minds to a new way of thinking, as evidenced by the increase in tolerance of risk in play, it did not sufficiently influence their self-efficacy in implementing the change, even a small one, and that a more intensive and complex intervention is required to shift behavior.

As ECE attitudes toward outdoor play and risk-taking in play have a major impact on children's outdoor play in ELCCs [27], avenues for shifting attitudes are necessary to foster changes in outdoor play provision. The OutsidePlay-ECE intervention was efficient and effective in changing ECEs and ELCC administrators' tolerance to the risk of children's outdoor play. Given the ease of distribution, no cost to users, and low resource requirement for ongoing maintenance of the web-based tool, the OutsidePlay-ECE intervention can be easily deployed in these efforts. This is particularly relevant during the COVID-19 pandemic, when increased outdoor time is recommended as a major strategy for reducing transmission [44]. Outdoor play is a means for improving children's mental health and coping strategies [45,46], and a web-based tool is the most feasible way to deliver such an intervention. In conclusion, the findings support the use of OutsidePlay-ECE as an intervention to improve outdoor play in ELCCs.

Limitations

Our study had several limitations. First, although social cognitive theory and respective behavior change techniques have been mapped [20], we did not test the hypothesized relationships between the theory constructs and outcomes. This limits our ability to better understand why we observed sustained changes in T-TRiPS and why we did not observe such changes in goal attainment. Given the significant findings on T-TRiPS, future research can explore the potential causal pathways leading to intervention effectiveness. Second, participant attrition was greater in the intervention condition than in the control condition. We did not find any sociodemographic differences between the 2 conditions among participants who remained in the study. However, attrition is a concern for eHealth interventions [47], and the field would benefit from further research on the factors that influence participant retention. Third, we conducted a study during the COVID-19 pandemic, which included initial ELCC closures. The data collection period occurred after most ELCCs resumed operations. However, practices remained in flux as Canadian ELCCs received rapidly evolving provincial and federal guidance regarding recommended procedures, whereas understanding and implementation of the guidance was challenging and varied between centers. Although this may have provided novel insights into ECEs' and ELCC administrators' perceptions of children's outdoor play in the specific context of the pandemic, the findings

XSL•FO
RenderX

may have differed in other conditions. Fourth, data collection spanned winter and spring, and it is unclear whether seasonal changes influence risk tolerance. However, few of the questions within the T-TRiPS would be expected to differ with seasons, as they assess more general risk attitudes (eg, "Do you wait to see how well the children in your centre manage challenges before getting involved?"). Finally, we did not implement a system to monitor dwell time on the Position Statement on Active Outdoor Play, which is the document shared with participants in the control condition. Therefore, we did not know whether the participants read the document or how long they took to do so. In future research, this information could offer opportunities for further analyses.

Conclusions

ELCCs are important settings for influencing early childhood, and these childcare experiences can impact lifelong health, development, and well-being trajectories [48]. High-quality early childhood education can mitigate the effects of early adversity and reduce inequities in more disadvantaged children [49]. Research is growing on the importance of outdoor play to children's physical, social, and cognitive development; risk perception; and mental health [14,16-19], and it is necessary to ensure children's regular and repeated access to outdoor play opportunities, particularly in ELCCs. To facilitate this, ECEs need to understand the essence and benefits of risky outdoor play for children and how best to provide and accommodate it. Our RCT results demonstrated that the OutsidePlay-ECE web-based intervention is effective in increasing ECEs' and ELCC administrators' risk tolerance in children's outdoor play. As an easily accessible and free resource, the OutsidePlay-ECE has great potential to support early childhood education practices. For example, it can be integrated into ECE professional development, provided as a standalone ECE program, and revisited over time to help ECEs deepen their understanding and expand their practice related to outdoor play provision.

Acknowledgments

This RCT study was funded by the Lawson Foundation (grant GRT 2019-79). The Lawson Foundation was not involved in any aspect of the study design or the writing of the manuscript. MB was supported by salary awards from the British Columbia Children's Hospital Research Institute. The authors are also grateful to the Government of Canada—Early Learning and Child Care—under the Social Development Partnership program, Project 016554719 for funding the development of the OutsidePlay-ECE intervention. The authors thank the Digital Lab at the University of British Columbia Department of Pediatrics for providing partnership and technical support in the development and programming of the OutsidePlay intervention.

Data Availability

All data, password-protected and stored in the secure network at the British Columbia Children's Hospital Research Institute, will be available from MB upon reasonable request within 5 years of the completion of the study.

Authors' Contributions

MB conceived of the study. CSH, FM, MZ, MW, TC, and EO assisted MB with the development of the OutsidePlay-ECE intervention content. MB and JJ led the development of the intervention design, with contribution from CSH. YL performed the statistical analysis. CSH and YL wrote the first full draft of this manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare that they are both the developers and evaluators of the intervention.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 278 KB - jmir_v24i6e36826_app1.pdf]

Multimedia Appendix 2
Complete screenshots of the OutsidePlay-ECE intervention.
[PDF File (Adobe PDF File), 51324 KB - jmir_v24i6e36826_app2.pdf]

References

1. Holt NL, Neely KC, Spence JC, Carson V, Pynn SR, Boyd KA, et al. An intergenerational study of perceptions of changes in active free play among families from rural areas of Western Canada. BMC Public Health 2016 Aug 19;16(1):829 [FREE Full text] [doi: 10.1186/s12889-016-3490-2] [Medline: 27538781]
2. Kyttä M, Hirvonen J, Rudner J, Pirjola I, Laatikainen T. The last free-range children? Children's independent mobility in Finland in the 1990s and 2010s. J Transport Geography 2015 Jul;47:1-12. [doi: 10.1016/j.jtrangeo.2015.07.004]

XSL•FO
RenderX

3. Brussoni M, Gibbons R, Gray C, Ishikawa T, Sandseter EB, Bienenstock A, et al. What is the relationship between risky outdoor play and health in children? a systematic review. Int J Environ Res Public Health 2015 Jun 08;12(6):6423-6454 [FREE Full text] [doi: 10.3390/ijerph120606423] [Medline: 26062038]
4. Tremblay MS, Gray C, Babcock S, Barnes J, Bradstreet CC, Carr D, et al. Position statement on active outdoor play. Int J Environ Res Public Health 2015 Jun 08;12(6):6475-6505 [FREE Full text] [doi: 10.3390/ijerph120606475] [Medline: 26062040]
5. Schoeppe S, Duncan MJ, Badland HM, Oliver M, Browne M. Associations between children's independent mobility and physical activity. BMC Public Health 2014 Jan 29;14:91 [FREE Full text] [doi: 10.1186/1471-2458-14-91] [Medline: 24476363]
6. Floyd MF, Bocarro JN, Smith WR, Baran PK, Moore RC, Cosco NG, et al. Park-based physical activity among children and adolescents. Am J Prev Med 2011 Sep;41(3):258-265. [doi: 10.1016/j.amepre.2011.04.013] [Medline: 21855739]
7. Gray C, Gibbons R, Larouche R, Sandseter EB, Bienenstock A, Brussoni M, et al. What is the relationship between outdoor time and physical activity, sedentary behaviour, and physical fitness in children? a systematic review. Int J Environ Res Public Health 2015 Jun 08;12(6):6455-6474 [FREE Full text] [doi: 10.3390/ijerph120606455] [Medline: 26062039]
8. Gordon-Larsen P, Nelson MC, Page P, Popkin BM. Inequality in the built environment underlies key health disparities in physical activity and obesity. Pediatrics 2006 Feb;117(2):417-424. [doi: 10.1542/peds.2005-0058] [Medline: 16452361]
9. Velde GT, Plasqui G, Willeboordse M, Winkens B, Vreugdenhil A. Associations between physical activity, sedentary time and cardiovascular risk factors among Dutch children. PLoS One 2021;16(8):e0256448 [FREE Full text] [doi: 10.1371/journal.pone.0256448] [Medline: 34449807]
10. Janssen I, Leblanc AG. Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. Int J Behav Nutr Phys Act 2010 May 11;7:40 [FREE Full text] [doi: 10.1186/1479-5868-7-40] [Medline: 20459784]
11. Duncan MJ, Clarke ND, Birch SL, Tallis J, Hankey J, Bryant E, et al. The effect of green exercise on blood pressure, heart rate and mood state in primary school children. Int J Environ Res Public Health 2014 Apr 02;11(4):3678-3688 [FREE Full text] [doi: 10.3390/ijerph110403678] [Medline: 24699030]
12. Prezza M, Pilloni S, Morabito C, Sersante C, Alparone FR, Giuliani MV. The influence of psychosocial and environmental factors on children's independent mobility and relationship to peer frequentation. J Commun Appl Soc Psychol 2001 Nov;11(6):435-450. [doi: 10.1002/casp.643]
13. Tremblay MS, LeBlanc AG, Kho ME, Saunders TJ, Larouche R, Colley RC, et al. Systematic review of sedentary behaviour and health indicators in school-aged children and youth. Int J Behav Nutr Phys Act 2011 Sep 21;8:98 [FREE Full text] [doi: 10.1186/1479-5868-8-98] [Medline: 21936895]
14. Why parenting matters for children in the 21st century. OECD iLibrary. URL: https://www.oecd-ilibrary.org/education/why-parenting-matters-for-children-in-the-21st-century_129a1a59-en [accessed 2022-05-25]
15. Ball DJ, Brussoni M, Gill TR, Harbottle H, Spiegal B. Avoiding a dystopian future for children's play. Int J Play 2019 Apr 04;8(1):3-10. [doi: 10.1080/21594937.2019.1582844]
16. Brussoni M, Olsen LL, Pike I, Sleet DA. Risky play and children's safety: balancing priorities for optimal child development. Int J Environ Res Public Health 2012 Aug 30;9(9):3134-3148 [FREE Full text] [doi: 10.3390/ijerph9093134] [Medline: 23202675]
17. Doepke M, Sorrenti G, Zilibotti F. The economics of parenting. Annu Rev Econ 2019 Aug 02;11(1):55-84. [doi: 10.1146/annurev-economics-080218-030156]
18. Janssen I. Hyper-parenting is negatively associated with physical activity among 7-12 year olds. Prev Med 2015 Apr;73:55-59 [FREE Full text] [doi: 10.1016/j.ypmed.2015.01.015] [Medline: 25634332]
19. Schiffrin HH, Godfrey H, Liss M, Erchull MJ. Intensive parenting: does it have the desired impact on child outcomes? J Child Fam Stud 2014 Aug 20;24(8):2322-2331. [doi: 10.1007/s10826-014-0035-0]
20. Brussoni M, Han CS, Lin Y, Jacob J, Pike I, Bundy A, et al. A web-based and in-person risk reframing intervention to influence mothers' tolerance for, and parenting practices associated with, children's outdoor risky play: randomized controlled trial. J Med Internet Res 2021 Apr 27;23(4):e24861 [FREE Full text] [doi: 10.2196/24861] [Medline: 33904820]
21. Brussoni M, Ishikawa T, Han C, Pike I, Bundy A, Faulkner G, et al. Go Play Outside! Effects of a risk-reframing tool on mothers' tolerance for, and parenting practices associated with, children's risky play: study protocol for a randomized controlled trial. Trials 2018 Mar 7;19(1):e24861 [FREE Full text] [doi: 10.1186/s13063-018-2552-4]
22. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
23. Wheeler A, Fowler J, Hattingh L. Using an intervention mapping framework to develop an online mental health continuing education program for pharmacy staff. J Contin Educ Health Prof 2013;33(4):258-266. [doi: 10.1002/chp.21198] [Medline: 24347104]
24. Brussoni M, Han C, Jacob J, Munday F, Zeni M, Walters M, et al. A web-based risk-reframing intervention to influence early childhood educators' attitudes and supportive behaviors toward outdoor play: protocol for the OutsidePlay study

XSL•FO
RenderX

randomized controlled trial. JMIR Res Protoc 2021 Nov 18;10(11):e31041 [FREE Full text] [doi: 10.2196/31041] [Medline: 34792479]
25. Sinha M. Child care in Canada. Statistics Canada. URL: https://www150.statcan.gc.ca/n1/pub/89-652-x/89-652-x2014005-eng.htm [accessed 2022-05-26]
26. Copeland KA, Khoury JC, Kalkwarf HJ. Child care center characteristics associated with preschoolers' physical activity. Am J Prev Med 2016 Apr;50(4):470-479 [FREE Full text] [doi: 10.1016/j.amepre.2015.08.028] [Medline: 26585052]
27. Bilton H. Values stop play? Teachers' attitudes to the early years outdoor environment. Early Child Development Care 2019 Sep 19;190(1):12-20. [doi: 10.1080/03004430.2019.1653548]
28. Ernst J. Early childhood educators' use of natural outdoor settings as learning environments: an exploratory study of beliefs, practices, and barriers. Environ Educ Res 2013 Sep 16;20(6):735-752. [doi: 10.1080/13504622.2013.833596]
29. New R, Mardell B, Robinson D. Early childhood education as risky business: going beyond what's safe to discovering what's possible. Early Child Res Pract 2005 Sep 1;7(2):1-21.
30. 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]
31. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
32. Bartholomew L, Markham C, Ruiter R, Fernandez M, Kok G, Parcel G. Planning Health Promotion Programs: An Intervention Mapping Approach, 4th Edition. San Francisco, CA, USA: John Wiley & Sons; 2016.
33. Bandura A. Social cognitive theory: an agentic perspective. Annu Rev Psychol 2001;52:1-26. [doi: 10.1146/annurev.psych.52.1.1] [Medline: 11148297]
34. Sandseter EB. Characteristics of risky play. J Adventure Educ Outdoor Learn 2009 Jun 24;9(1):3-21. [doi: 10.1080/14729670802702762]
35. The biggest risk is keeping kids indoors: 2015 ParticipACTION report card on physical activity for children and youth. ParticipACTION. 2015. URL: https://participaction.cdn.prismic.io/participaction%2F61cf55e8-c1c0-42c7-ba6b-1480fd2c29b9_participaction-2015-report-card-full.pdf [accessed 2022-05-26]
36. Ihrig KM. Teacher tolerance of risk in play scale (T-TRiPS): evaluating the psychometric properties of a new measure. Colorado State University. URL: https://mountainscholar.org/handle/10217/232469 [accessed 2022-06-02]
37. Hill A, Bundy AC. Reliability and validity of a new instrument to measure tolerance of everyday risk for children. Child Care Health Dev 2014 Jan;40(1):68-76. [doi: 10.1111/j.1365-2214.2012.01414.x] [Medline: 22846064]
38. Bond TG, Fox CM. Applying the Rasch Model: Fundamental Measurement in the Human Sciences, 2nd edition. Milton Park, Abingdon-on-Thames, Oxfordshire, England, UK: Routledge; 2007.
39. Chalmers RP. Mirt: a multidimensional item response theory package for the R environment. J Stat Soft 2012;48(6). [doi: 10.18637/jss.v048.i06]
40. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. URL: https://www.r-project.org/ [accessed 2022-06-03]
41. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC; 2017.
42. Cheng T, Brussoni M, Han C, Munday F, Zeni M. Perceived challenges of early childhood educators in promoting unstructured outdoor play: an ecological systems perspective. Early Years 2022 Feb 26:1-17 [FREE Full text] [doi: 10.1080/09575146.2022.2034140]
43. Messing S, Rütten A, Abu-Omar K, Ungerer-Röhrich U, Goodwin L, Burlacu I, et al. How can physical activity be promoted among children and adolescents? A systematic review of reviews across settings. Front Public Health 2019;7:55 [FREE Full text] [doi: 10.3389/fpubh.2019.00055] [Medline: 30941342]
44. Public health guidance for child care settings. BC Centre for Disease Control. URL: http://www.bccdc.ca/Health-Info-Site/Documents/COVID_public_guidance/Guidance_Child_Care.pdf [accessed 2020-09-26]
45. McCormick R. Does access to green space impact the mental well-being of children: a systematic review. J Pediatr Nurs 2017;37:3-7. [doi: 10.1016/j.pedn.2017.08.027] [Medline: 28882650]
46. Tillmann S, Tobin D, Avison W, Gilliland J. Mental health benefits of interactions with nature in children and teenagers: a systematic review. J Epidemiol Community Health 2018 Oct;72(10):958-966 [FREE Full text] [doi: 10.1136/jech-2018-210436] [Medline: 29950520]
47. Van der Mispel C, Poppe L, Crombez G, Verloigne M, De Bourdeaudhuij I. A self-regulation-based eHealth intervention to promote a healthy lifestyle: investigating user and website characteristics related to attrition. J Med Internet Res 2017 Jul 11;19(7):e241 [FREE Full text] [doi: 10.2196/jmir.7277] [Medline: 28698168]
48. Doyle O, Harmon CP, Heckman JJ, Tremblay RE. Investing in early human development: timing and economic efficiency. Econ Hum Biol 2009 Mar;7(1):1-6 [FREE Full text] [doi: 10.1016/j.ehb.2009.01.002] [Medline: 19213617]
49. Heckman JJ. The economics of inequality: the value of early childhood education. Am Educator 2011;35(1):31-35.

XSL•FO
RenderX

Abbreviations

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
ECE: early childhood educator
ELCC: early learning childcare center
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture
T-TRiPS: Teacher Tolerance of Risk in Play Scale

Edited by T Leung; submitted 27.01.22; peer-reviewed by S Wyver, P McCrorie, H Li; comments to author 19.04.22; revised version received 09.05.22; accepted 10.05.22; published 10.06.22.

Please cite as:
Brussoni M, Han CS, Lin Y, Jacob J, Munday F, Zeni M, Walters M, Oberle E
Evaluation of the Web-Based OutsidePlay-ECE Intervention to Influence Early Childhood Educators' Attitudes and Supportive Behaviors Toward Outdoor Play: Randomized Controlled Trial
J Med Internet Res 2022;24(6):e36826
URL: https://www.jmir.org/2022/6/e36826
doi:10.2196/36826
PMID:35687394

©Mariana Brussoni, Christina S Han, Yingyi Lin, John Jacob, Fritha Munday, Megan Zeni, Melanie Walters, Eva Oberle. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.06.2022. 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 Aim2Be mHealth Intervention for Children With Overweight or Obesity and Their Parents: Person-Centered Analyses to Uncover Digital Phenotypes

Olivia De-Jongh González[1], PhD; Claire N Tugault-Lafleur[2], PhD, RD; E Jean Buckler[3], PhD; Jill Hamilton[4], PhD; Josephine Ho[5], MD; Annick Buchholz[6], PhD; Katherine M Morrison[7], MD; Geoff DC Ball[8], PhD, RD; Louise C Mâsse[1], PhD

[1]School of Population and Public Health, University of British Columbia, BC Children's Hospital Research Institute, Vancouver, BC, Canada
[2]School of Nutrition Sciences, Faculty of Health Sciences, The University of Ottawa., Ottawa, ON, Canada
[3]School of Exercise Science, Physical and Health Education, University of Victoria., Victoria, BC, Canada
[4]Department of Paediatrics, Hospital for Sick Children, University of Toronto, Toronto, ON, Canada
[5]Cumming School of Medicine, Department of Pediatrics, University of Calgary, Calgary, AB, Canada
[6]Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON, Canada
[7]Department of Pediatrics, Center for Metabolism, Obesity and Diabetes Research, McMaster University, Hamilton, ON, Canada
[8]Department of Pediatrics, Faculty of Medicine & Dentistry, University of Alberta, Edmonton, AB, Canada

Corresponding Author:
Louise C Mâsse, PhD
School of Population and Public Health, University of British Columbia
BC Children's Hospital Research Institute
4480 Oak St.
Vancouver, BC, V6H 3V4
Canada
Phone: 1 6048752000 ext 5563
Email: lmasse@bcchr.ubc.ca

Abstract

Background: Despite the growing number of mobile health (mHealth) interventions targeting childhood obesity, few studies have characterized user typologies derived from individuals' patterns of interactions with specific app features (digital phenotypes).

Objective: This study aims to identify digital phenotypes among 214 parent-child dyads who used the Aim2Be mHealth app as part of a randomized controlled trial conducted between 2019 and 2020, and explores whether participants' characteristics and health outcomes differed across phenotypes.

Methods: Latent class analysis was used to identify distinct parent and child phenotypes based on their use of the app's behavioral, gamified, and social features over 3 months. Multinomial logistic regression models were used to assess whether the phenotypes differed by demographic characteristics. Covariate-adjusted mixed-effect models evaluated changes in BMI *z* scores (zBMI), diet, physical activity, and screen time across phenotypes.

Results: Among parents, 5 digital phenotypes were identified: *socially engaged* (35/214, 16.3%), *independently engaged* (18/214, 8.4%) (*socially* and *independently engaged* parents are those who used mainly the social or the behavioral features of the app, respectively), *fully engaged* (26/214, 12.1%), *partially engaged* (32/214, 15%), and *unengaged* (103/214, 48.1%) users. Married parents were more likely to be *fully engaged* than *independently engaged* (P=.02) or *unengaged* (P=.01) users. *Socially engaged* parents were older than *fully engaged* (P=.02) and *unengaged* (P=.01) parents. The latent class analysis revealed 4 phenotypes among children: *fully engaged* (32/214, 15%), *partially engaged* (61/214, 28.5%), *dabblers* (42/214, 19.6%), and *unengaged* (79/214, 36.9%) users. *Fully engaged* children were younger than *dabblers* (P=.04) and *unengaged* (P=.003) children. *Dabblers* lived in higher-income households than *fully* and *partially engaged* children (P=.03 and P=.047, respectively). *Fully engaged* children were more likely to have *fully engaged* (P<.001) and *partially engaged* (P<.001) parents than *unengaged* children. Compared with *unengaged* children, *fully* and *partially engaged* children had decreased total sugar (P=.006 and P=.004, respectively) and energy intake (P=.03 and P=.04, respectively) after 3 months of app use. *Partially engaged* children also had decreased sugary beverage intake compared with *unengaged* children (P=.03). Similarly, children with *fully engaged* parents had decreased

XSL•FO
RenderX

zBMI, whereas children with *unengaged* parents had increased zBMI over time (*P*=.005). Finally, children with *independently engaged* parents had decreased caloric intake, whereas children with *unengaged* parents had increased caloric intake over time (*P*=.02).

Conclusions: Full parent-child engagement is critical for the success of mHealth interventions. Further research is needed to understand program design elements that can affect participants' engagement in supporting behavior change.

Trial Registration: ClinicalTrials.gov NCT03651284; https://clinicaltrials.gov/ct2/show/NCT03651284

International Registered Report Identifier (IRRID): RR2-10.1186/s13063-020-4080-2

(J Med Internet Res 2022;24(6):e35285) doi:10.2196/35285

KEYWORDS

mobile health; mHealth; childhood obesity; digital phenotypes; latent class analysis

Introduction

Background

Childhood obesity remains a significant health problem in Canada [1]. Evidence shows that family-based multicomponent interventions that integrate self-regulatory strategies (ie, goal setting, graded tasks, and self-monitoring) and support changes at the familial and individual levels are necessary to significantly affect child weight outcomes (eg, BMI, waist to hip ratio, and total fat mass [2-9]). However, a 2018 meta-analysis [4] found that family-based multicomponent behavioral interventions had a small effect in reducing children's BMI in efficacy trials versus standard-of-care controls (β=−.16, 95% CI −0.24 to −0.07).

Mobile health (mHealth) interventions offer a promising adjunct or alternative to in-person treatments to support lifestyle behavior change [10,11]. Several reviews [12-15] and meta-analyses [16,17] have suggested that mHealth interventions offer multiple advantages to in-person interventions (eg, real-time data collection, intervention in natural environments, lower costs, health behavior tracking with feedback, and incorporation of gamified elements), which may appeal to children and youth [12]. Data on the efficacy of mHealth interventions for the prevention and management of childhood obesity are promising but limited as this is still a rapidly evolving field of research [12,14,18].

mHealth interventions for children living with obesity are most often evaluated using randomized controlled trials and, in some cases, evaluate the *dose* of the intervention received to provide a better understanding of their effects [19-24]. Dose-response analyses are often measured in terms of total minutes or percentage of content examined; however, this approach does not provide a nuanced picture of how users may benefit from different mHealth intervention components (ie, what design elements of the app may be more successful in engaging participants and promoting health behavior change) [25,26]. Studies examining how intervention exposure affects behavior change cannot solely focus on the quantity of the intervention received, but must also consider how participants engage with the *active ingredients* of the intervention—namely, the features that support behavior change.

mHealth interventions are particularly well-suited to examine in greater detail which components of the intervention participants engage with through app analytics data. Recently, there have been calls to develop analytical methods to process the vast amounts of data that are available when using mHealth technologies [27] and identify *digital phenotypes* (ie, user typologies derived from individuals' patterns of interactions with specific app features) [28,29]. Although digital phenotypes have been used in other areas of health research (eg, diabetes [30], sleep [31], mental health [32]) and dietary and physical activity behaviors in a nonclinical sample [33], little attention has been paid to the treatment of obesity in childhood. Some studies have investigated which app features participants use [23,33] and individual characteristics associated with partial or total use of an intervention [33-36]. However, most studies evaluated usability derived from self-reported measures (eg, asking participants about their preferences and use of app features), total app use, or the use of individual features instead of focusing on patterns of app use [23,34,37].

Objectives

To our knowledge, no study targeting childhood obesity has identified user typologies based on participants' engagement with objectively measured components of an mHealth intervention. To address this gap, this study aimed to (1) identify digital phenotypes of Canadian children with overweight or obesity and their parents who used an mHealth app (the Aim2Be app [25]) over a 3-month period, (2) explore whether participants' characteristics differed by digital phenotype, and (3) evaluate 3-month changes in children's BMI *z* scores (zBMI) and dietary, physical activity, and screen time behaviors across digital phenotypes.

Methods

Study Design

This study was a secondary analysis of data collected from a randomized controlled trial evaluating the efficacy of the Aim2Be app (version 2) to improve lifestyle behaviors and adiposity among children with overweight or obesity [25,38]. The trial was registered on ClinicalTrials.gov (NCT03651284) on August 29, 2018 [25]. The CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist [39] is available in Multimedia Appendix 1. Data analyzed in this study were collected from March 2019 to June 2020.

XSL•FO
RenderX

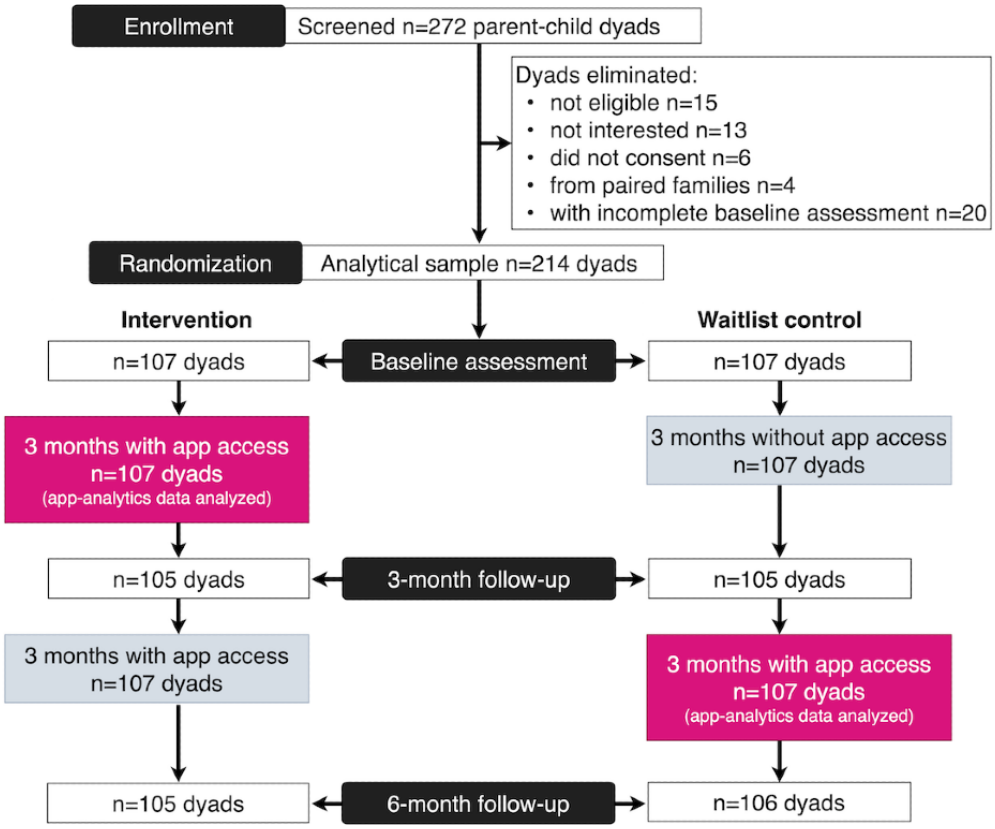
Ethics Approval

The evaluation protocol was approved by the Children's and Women's Research Ethics Board at the University of British Columbia (H16-03090/H17-02032), the Health Research Ethics Board at the University of Alberta (Pro00076869), the Hospital for Sick Children Research Ethics Board (REB1000059362), the Hamilton Integrated Research Ethics Board (project 4250), and the Children's Hospital of Eastern Ontario Research Ethics Board (18/01E). All the participants provided web-based consent before participating in the study.

Data Collection Protocol

The detailed study protocol has been published elsewhere [25]. The participating families (N=214) were recruited from 6 weight management clinic sites across Canada, as well as through Facebook. Children were eligible to participate if they were aged 10 to 17 years and were overweight or obese, as defined by the age- and sex-specific World Health Organization cutoffs [40]. After providing consent, eligible participants completed a web-based survey, three 24-hour dietary recalls, and received assessment tools for height (measuring tape) weight (digital scale), and physical activity (Fitbit Flex 2, Fitbit Inc) to complete baseline measurements. Participants completed follow-up assessments at the 3- and 6-month follow-ups. Families randomized to the experimental group (107/214, 50%) had access to the app after completing baseline measures. Waitlisted control families (107/214, 50%) were given access to the app after completing their assessment at the 3-month follow-up. This study combined data collected from baseline to 3 months in the intervention group, and from 3 to 6 months in the waitlisted control group (Figure 1). Randomization was successful, and participants' characteristics did not differ between the intervention and the waitlisted control group; however, our analyses were not based on the randomization group but dependent on users' engagement.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram depicting the study methodology and data analyzed.



The Aim2Be Intervention

The theoretical foundations of the Aim2Be app have been described elsewhere [25]. Briefly, the app, which was cocreated by Ayogo Health Inc [41] and the Childhood Obesity Foundation with expert input, aimed to promote healthy lifestyle behaviors among children and their parents by targeting dietary, physical activity, screen time, and sleep behaviors while emphasizing healthy body image, strong self-esteem, and living green [25]. The behavior change techniques incorporated in the app are grounded in social cognitive theory [42], the Player Experience and Need Satisfaction Model—an extension of the self-determination theory [43,44]—and the Agency, Challenge, Uncertainty, Discovery, and Outcomes framework [45]. The content within different features of the app was also informed by the Canadian 24-Hour Movement Guidelines [46] and the Canadian dietary guidelines in place at the time of the study (Canada's Food Guide 2007) [47].

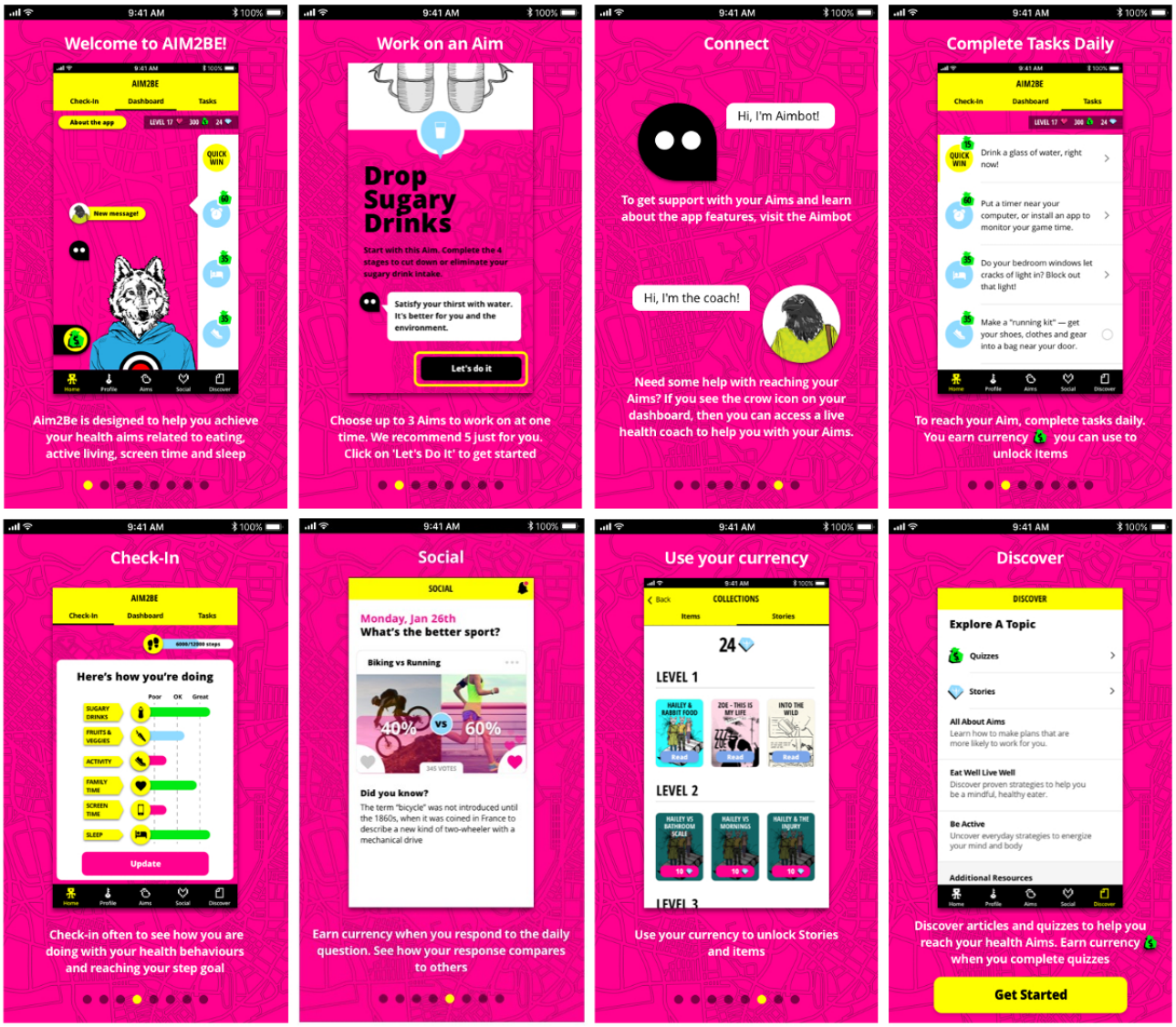
The Aim2Be app features fall under 3 broad domains: behavioral, gamified, and social. The behavioral domain draws on self-regulatory strategies such as goal setting, self-monitoring, and graded tasks to facilitate behavior change by strengthening self-regulatory skills [4,8]. The gamified domain focuses on increasing participants' enjoyment, engagement, and motivation through various gamification elements (eg, personalization, challenges, uncertainty). The

XSL•FO
RenderX

social domain facilitates peer support, behavior modeling, and interaction with other app users or with a coach through different interactive features (eg, answering poll questions, viewing poll results, posting on the social wall, and responding to others' posts). Social support is also provided to children through a companion app for parents, which aims to facilitate behavioral changes through a positive familial environment, reinforcement strategies, and environmental and stimulus control. Figures 2 and 3 illustrate screenshots of the child and parent interventions, respectively.

In addition to the parent companion app, 2 very similar versions of Aim2Be were developed for preteens (aged 10-13 years) and teenagers (aged 13-17 years), with 3 app features (ie, stages, posting on a social wall, and responding to others' posts) available only to teenagers. As this study combined data from both teenagers and preteens, features only available to both groups were included in the analyses.

Figure 2. Screenshot of the Aim2Be app for children.



XSL•FO
RenderX

Figure 3. Screenshot of the Aim2Be app for parents.

Measures

Use of Aim2Be App Features

App analytics (provided by Ayogo Health Inc [41]) were used to track the number of times children and parents used each Aim2Be app feature. The behavioral domain included the following five features: (1) *aims*, indicating the number of high-level goals chosen by users while indicating their perceived importance and potential obstacles; (2) *tasks*, indicating the number of activities users completed to accomplish their aims; (3) *check-ins*, indicating the number of times users self-monitored their progress regarding specific health behaviors, with short recommendations on how to improve their behaviors; (4) *articles read*, indicating the number of articles providing educational content read by the user; and (5) *articles reflected on*, indicating the number of written responses provided by the user after reading an article.

The gamified domain included the following four features: (1) *quick wins*, which involved simple tasks users completed to engage in a healthy behavior or explore a new feature of the app, and which allowed them to earn digital currency; (2) *stories*, which involved the number of times users read *choose your own adventure* stories (these stories used a user-guided fictional character involved in a series of decision-making processes); (3) *quizzes*, which involved short tests that the users answered, allowing them to earn digital currency if they selected the correct answer; and (4) *collections*, which involved digital items the user purchased with digital currency within the app. The parent app integrated only *quick wins* as the gamified domain.

The social domain included the following four features: (1) *answer poll*, which involved the number of 2-choice poll questions from the social poll responded to by the users, with feedback on the percentage of users who selected each option; (2) *digital coach*, which involved the number of chat sessions between the user and a digital coach with preprogrammed prompts, questions, and answers; (3) *live coach*, which involved the number of one-on-one messages the user sent to a live trained health coach; and (4) *posts*, which involved the number of times the user posted a message on a social wall, sharing thoughts, feelings, or experiences with others. By design, the live coach feature, analyzed as part of the social domain, was not made available to participants randomized to the waitlisted

XSL•FO
RenderX

control group; therefore, data for these participants were missing at random. In addition, only teens (119/214, 55.6%) had access to the posting feature. Consequently, this feature was not analyzed in the children's sample as the missing data were age-related.

Children's zBMI

Parents were mailed a digital scale (Active Era) and a measuring tape (HDX Corp) with instructions (using the Centers for Disease Control and Prevention home protocol [48]) to accurately measure their child's height and weight at home. This procedure has been validated to assess children's height and weight at home [49]. Children's height and weight were then used to compute *z*BMI using the World Health Organization Stata macro [40], where being overweight in childhood was defined as *z*BMI >1 SD and ≤2 SD, and obesity was defined as *z*BMI >2 SD.

Health Behaviors

Dietary Behaviors

Children's dietary behaviors were evaluated with the Waterloo Eating Behavior Questionnaire, a 24-hour web-based dietary recall (intraclass correlation coefficients ranging from 0.39 to 0.71 for energy, carbohydrates, sugar, fiber, and fats, validated against dietitian interviews) [50]. Children reported all foods and beverages consumed on the previous day. Standardized food group servings using the 2007 Canada's Food Guide classification framework were used to quantify the amount of food consumed (eg, number of servings of vegetables and fruits) [47]. A composite index of dietary quality (the Canadian Healthy Eating Index [51]) was used as a measure of overall adherence to the 2007 Canada's Food Guide. The index ranges from 0 to 100 points, where scores <50, 50 to 80, and >80 indicate *poor*, *requiring improvement*, and *good* dietary quality, respectively [51]. Parents' dietary behaviors were evaluated using 7 items adapted from the Canadian Community Health Surveys [52]. Parents reported their own consumption of vegetables and fruits (excluding fruit juices), 100% fruit juices, and sugar-sweetened beverages.

Physical Activity

Children's physical activity was evaluated using Fitbit Flex 2 (Fitbit Inc). Children wore the Fitbit for 7 to 14 days at baseline and at 3 and 6 months, and their daily step count was obtained by our team using Fitabase, a web-based platform designed for research using Fitbits [53]. Furthermore, we computed the children's average number of daily steps. In addition, children completed a web-based survey, which included 5 questions from the Physical Activity Questionnaire for Older Children [54]; a 7-day recall inquiring about the amount of physical activity with responses between *none* and *more than 2 hours*. The total score of the questionnaire was significantly related to moderate and vigorous physical activity using accelerometers (r=0.33) [55]. Parents' physical activity was evaluated using 7 items from the Physical Activity Questionnaire Short Form (repeatability reliability across different countries ranged from 0.32 to 0.88, with 75% of the correlation coefficients >0.65 and a pooled coefficient of 0.76 [56]). Participants were asked about the number of days and minutes spent sitting, walking, and engaging in vigorous and moderate physical activity over the past 7 days. The average daily time was calculated for sitting, walking, and moderate and vigorous physical activity.

Screen Time

Children's and parents' screen time was evaluated with 2 items adapted from the Sedentary Behavior Questionnaire for adults (intraclass correlation coefficient ranged from 0.51 to 0.93 [57]). Children and parents reported the time (minutes) spent watching television; playing computer or video games; using a computer, tablet, or mobile device outside of school or paid work; and talking or texting on a cell phone during their most recent week and weekend day. The average daily sedentary time was then calculated.

Statistical Approach

Latent class analysis (LCA) in MPlus version 8 (Muthen and Muthen [58]) was used to separately identify digital phenotypes among children and parents. There is no fixed minimum sample size for LCA as it depends on multiple factors (eg, number and quality of indicators, class differentiation, and relative samples in each class) [59]. Of relevance, previous Monte Carlo simulations [60] have found that an LCA with 100 participants can result in reliable solutions when conducted with robust indicators, providing support for conducting an LCA with 214 participants. The LCA identified digital phenotypes based on different use patterns for the various behavioral, social, and gamified app features, similar to a recent study profiling child users using an earlier version of Aim2Be [33]. As the distribution of use for each feature was skewed, an individual's use of each app feature was ranked as *no use* (a participant never used a given app feature), and among the remaining participants, *low use* (a participant's use of a feature was at or below the median use), or *high use* (a participant's use of a feature was above the median use). The robust maximum likelihood estimator with the expectation-maximization algorithm and 2000 random starts was used. The LCA used full information maximum likelihood to handle data missing at random in the *live coach* feature (no other variables included in the LCA had missing data). Various fit and relative indices were used to compare different k-class solutions to determine the best number of classes to be retained with the LCA [59]. We first evaluated the Bayesian information criterion, sample size–adjusted Bayesian information criterion, Akaike information criterion, consistent Akaike information criterion, and approximate weight of evidence. For these indices, both a lower value and a meaningful decrease after adding another class to the solution are desirable. Second, we compared neighboring solutions of k classes (eg, 3 vs 4 classes) with the relative indices of the Vuong-Lo-Mendell-Rubin likelihood ratio test, bootstrap likelihood ratio test, and Bayes factor. For the Vuong-Lo-Mendell-Rubin and bootstrap likelihood ratio test, a significant P value indicates a better fit of k classes than with the previous model (k-1 classes). For the Bayes factor, higher scores indicate stronger evidence supporting k classes than those supporting k+1 classes. Third, we estimated how each model was corrected by all models using the correct model probability index, where higher values are desirable. Other indicators of well-differentiated classes such as entropy and average posterior

XSL•FO
RenderX

probability were also evaluated, where desirable values were >0.8% and >70%, respectively. Finally, the *k*-class solution selected also considered practical and theoretical interpretability and the relative sample size of each class. Although some authors retained classes that encompassed at least 5% of the sample [59], the authors recognized the limitations of estimating classes with a low relative prevalence (1%-8%). This was accounted for when selecting the final solution.

Multiple multinomial logistic regression models were used to evaluate the associations between digital phenotypes included as the dependent variable and demographic factors (children's and parents' age and sex, parental educational attainment, marital status, household income, ethnicity, and recruitment site) as independent variables. Parental phenotypes were also added as predictors of children's phenotypes.

Mixed-effect models evaluated changes in health behaviors and *z*BMI across children's and parents' phenotypes. One model was run for each outcome (children's *z*BMI, children's and parents' diet, physical activity, and screen time). All models included an interaction term between time and phenotype and were adjusted for children's and parents' age and sex, parental educational attainment, marital status, household income, ethnicity, and recruitment site. Postestimation contrasts of marginal linear predictions tested overall group differences. For outcomes with borderline significance ($P<.10$) or significant ($P<.05$) overall group differences, we conducted pairwise comparisons and calculated the Cohen effect size as follows:

$$f^2 = (R^2_{AB} - R^2_A) / (1 - R^2_{AB})$$

Here, *B* is the predictor of interest (eg, interaction phenotype 1×time), *A* is the set of all other predictors (ie, demographics, time, and other phenotypes), R^2_{AB} is the proportion of variance that *A* and *B* together (ie, the full model) account for, and R^2_A is the proportion of variance the predictors explain in a reduced model, with all fixed effects from the full model, except for the effect of *B* and random effects constrained to be the same as those from the full model. Therefore, f^2 represents the proportion of variance uniquely accounted for by *B* [61,62].

All regression analyses were performed using Stata (version 15; StataCorp) [63]. The significance level was set at $P<.05$.

Results

Demographic Characteristics of the Participants

From the 214 parent-child dyads, the mean age of the children was 13 (SD 2.2) years, and the sample was evenly split among boys (104/214, 48.6%) and girls (110/214, 51.4%). Approximately 92.5% (198/214) of the participating parents were mothers, and 71% (152/214) were married or living with a partner. The mean age of parents was 44 (SD 6.2) years. Just over half of the parents (120/214, 56.1%) had not completed a university degree. Approximately 60.3% (129/214) of parents self-identified as having a White or European descent, 16.8% (36/214) reported mixed ethnicity, 5.6% (12/214) reported an East or Southeast Asian descent, 4.2% (9/214) reported a South Asian descent, and 3.3% (7/214) reported an indigenous descent. Household income ranged from <CAD $50,000 (US $37,500; 36/214, 16.8%) to >CAD $150,000 (US $112,500; 35/214, 16.4%). Approximately 30.8% (66/214) of parents reported incomes between CAD $50,000 (US $37,500) and CAD $100,000 (US $75,000), and 25.7% (55/214) reported incomes between CAD $100,000 (US $75,000) and CAD $150,000 (US $112,500).

Identifying Digital Phenotypes

Table 1 summarizes the results from the LCA and Figure 4 provides plots for selecting LCA indices for both the child and parent models. Fit indices and interpretability of the classes supported a 4-class solution among children. By contrast, the parent LCA fit indices pointed to a 5- or 6-class solution; however, further evaluation of the potential solutions led to the retention of the 5-class solution. Although 6 of the 11 indices showed the 6-class solution as the best option, the relatively small sample for 2 of the classes suggested an overextraction. Therefore, the 5-class model was retained as the final solution for the parents. In addition, the 5-class solution made more substantive sense. Moreover, the average posterior probability for both child- and parent-selected models ranged between 91% and 99%, indicating well-differentiated classes for the 5-class solution. Thus, our results suggest excellent differentiation between the classes.

Figure 5 shows children's and parents' digital phenotypes (A and B, respectively). Figure 5A shows 4 children's digital phenotypes (N=214): *unengaged*, *dabblers*, *partially engaged*, and *fully engaged*. *Unengaged* (79/214, 36.9% of users) included children who did not interact with most of the app features with exception of check-ins. *Dabblers* (42/214, 19.6% of users) regrouped children who did not use most behavioral features of the app (eg, completing tasks and reading or reflecting on articles) but predominantly interacted with gamified and social features, including collections and the digital coach. *Partially engaged* (61/214, 28.5% of users) included children who were low users of the behavioral features, particularly regarding task completion and reading and reflecting on articles but had greater use of the check-in feature. *Partially engaged* children had mixed interactions with the gamified and social features, with greater use of the collections and the digital coach, respectively, but rarely read stories or completed quizzes. *Fully engaged* (32/214, 15% of users) comprised high users of most app features and included children who engaged the most with the *active ingredients* of the app (ie, the behavioral features such as setting aims and completing tasks).

Figure 5B shows 5 parental digital phenotypes (N=214): *unengaged*, *socially engaged*, *independently engaged*, *partially engaged*, and *fully engaged*. *Unengaged* (103/214, 48.1% of users) included parents who did not use most of the features, with the exception of check-ins. *Socially engaged* (35/214, 16.4% of users) regrouped parents who engaged with the social features of the app by creating posts on the social wall, answering poll questions, and interacting with the live health coach. However, *socially engaged* parents had low use of the behavioral and gamified features and, in particular, did not complete any tasks within the app. *Independently engaged* (18/214, 8.4% of users) comprised parents who made little use

XSL•FO
RenderX

of the social features (the only social feature they used involved direct messages with the live health coach but did not interact with other parents). Instead, *independently engaged* parents focused their attention on the behavioral features of the app and mostly set aims, read articles, and completed check-ins; however, they also interacted with all the behavioral features to some degree. *Partially engaged* (32/214, 15% of users) included parents who had a mixed use of most app features, indicating that their engagement with some behavioral (eg, aims and check-ins) and social (eg, posts) features was evenly split between low and high use. *Partially engaged* parents tended to be high users of the article feature, low users of the answer poll feature, and nonusers of the digital coach feature. Hence, their overall engagement with the behavioral features tended to be greater than with the gamified and social features. Finally, *fully engaged* parents (26/214, 12.1% of users) included users who interacted extensively with all app features, except the digital coach feature.

Table 1. Comparative fit indices between k-class solutions for children and parents.

Classes in the model[a]	LL[b]	AIC[c]	BIC[d]	SABIC[e]	VLMR-LRT[f] *P* value	BLRT[g] *P* value	Entropy[h]	CAIC[i]	AWE[j]	BF[k]	CmP[l]
Children's models											
1	−2337	4725	4809	4730	N/A[m]	N/A	N/A	4758	4771	0.0	0.0
2	−1790	3682	3853	3692	<.001	<.001	.95	3750	3775	0.0	0.0
3	−1644	3441	3700	3456	.10	<.001	.95	3544	3582	0.3	0.2
4[n]	−1563	3331	3678	3351	.008	<.001	.96	3468	3520	16.6	0.8
5[o]	−1521	3300	3734	3325	.76	<.001	.95	3471	3536	71.5	0.0
Parents' models											
1	−1853	3746	3813	3749	N/A	N/A	N/A	3772	3782	0.0	0.0
2	−1455	2992	3130	3000	<.001	<.001	.95	3046	3067	0.0	0.0
3	−1357	2837	3046	2850	.09	<.001	.93	2920	2951	0.8	0.4
4	−1298	2761	3041	2778	.90	<.001	.97	2872	2913	4.4	0.5
5[n]	−1256	2720	3070	2741	.12	<.001	.98	2859	2911	17.7	0.1
6	−1229	2707	3128	2732	.79	.01	.99	2873	2936	22.7	0.0

[a]Model and number of classes in the solution.

[b]LL: log-likelihood.

[c]AIC: Akaike information criterion.

[d]BIC: Bayesian information criterion.

[e]SABIC: sample size–adjusted Bayesian information criterion.

[f]VLMR-LRT: Vuong-Lo-Mendell-Rubin adjusted likelihood ratio test.

[g]BLRT: bootstrapped likelihood ratio test.

[h]Entropy or differentiation between classes.

[i]CAIC: consistent Akaike information criterion.

[j]AWE: approximate weight of evidence.

[k]BF: Bayes factor.

[l]CmP: correct model probability.

[m]N/A: not applicable.

[n]Selected solution based on fit indices, relative sample sizes, and interpretability.

[o]This model was not identified, but the results are reported only for transparency purposes.

XSL•FO
RenderX

Figure 4. Plot of information criterion values across latent classes among children (A) and parents (B). AIC: Akaike information criterion; BIC: Bayesian information criterion; CAIC: consistent AIC; AWE: approximate weight of evidence; SABIC: sample size–adjusted BIC.

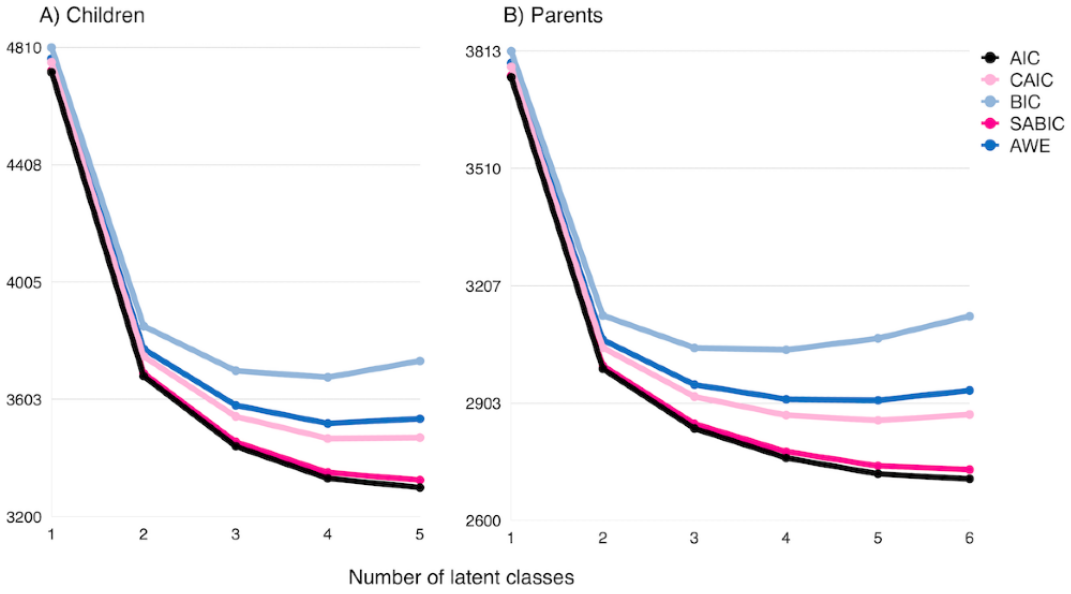
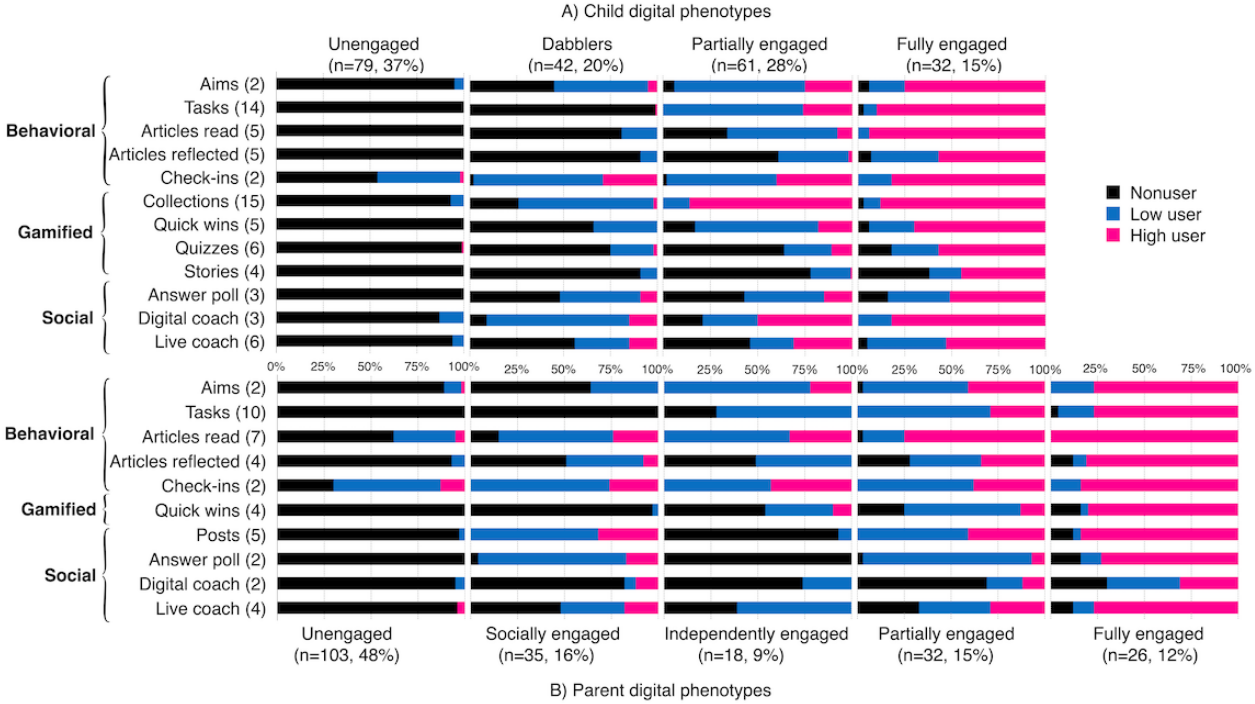


Figure 5. Conditional probability plots showing child (A) and parent (B) digital phenotypes (N=214). Numbers within brackets on the y-axis indicate the median distribution of use for each feature (eg, the median number of tasks completed by parents over 3 months was 10 among low and high users).



Demographic Characteristics Associated With Digital Phenotypes

Table 2 and Table 3 show the distribution of demographic factors across child and parent digital phenotypes, respectively, with relative risk ratios and significance levels available in Multimedia Appendix 2 (Tables S1 and S2). The results are presented separately for children and parents. Children in the *fully engaged* phenotype were 1 to 1.5 years younger than children belonging to the *dabblers* (*P*=.04) and *unengaged* (*P*=.003) phenotypes. Furthermore, children from the *dabblers* phenotype were more likely to be in a household with an income >CAD $80,000 (US $63,771) than children belonging to the *fully* and *partially engaged* phenotypes (*P*=.03 and *P*=.047, respectively). Parents in the *socially engaged* phenotype were 2 to 3 years older than parents in the *fully engaged* (*P*=.02) and *unengaged* (*P*=.01) phenotypes. Moreover, *fully engaged* parents were more likely to be married, common law, or living with a partner than parents belonging to the *independently engaged* (*P*=.02) and *unengaged* (*P*=.01) phenotypes, who were more likely to be single, divorced, or widowed.

Figure 6 shows the distribution of parental digital phenotypes across children's phenotypes, highlighting how their phenotypes were strongly associated. At one end of the spectrum, *fully engaged* children were more likely to have *fully* and *partially*

XSL•FO
RenderX

engaged parents, and at the other end, *unengaged* children were more likely to have *unengaged* parents.

Table 2. Demographic distribution across child digital phenotypes (N=214).

Predictors of child digital phenotypes[a]	Fully engaged	Partially engaged	Dabblers	Unengaged
Phenotype sample size, N	32	61	42	79
Age (years), mean (SD)[b]	12.0 (1.8)	12.9 (2.3)	13.0 (2.4)	13.5 (2.2)
Sex (female), n (%)	19 (59)	30 (49)	19 (45)	42 (53)
Household income (≥CAD $80,000; US $63,771), n (%)[c]	17 (53)	31 (51)	29 (69)	42 (53)
Parental education (more than a Bachelor's degree), n (%)	15 (47)	28 (46)	18 (43)	32 (41)
Parental marital status (married), n (%)	27 (84)	46 (75)	32 (76)	58 (73)
Race or ethnicity (White or European), n (%)	22 (69)	38 (62)	28 (67)	41 (52)

[a]Predictors' reference groups are: male, household income <CAD $80,000 (US $63,771), parental educational attainment lower than a bachelor's degree, single parents, and people who did not self-identify as having a White or European descent.

[b]The age of *fully engaged* children significantly differs from both *dabblers* and *unengaged* children's age.

[c]The household income of both *fully engaged* and *partially engaged* children significantly differs from the household income among *dabblers*.

Table 3. Demographic distribution across parent digital phenotypes (N=214).

Predictors of parent digital phenotypes[a]	Fully engaged	Partially engaged	Independently engaged	Socially engaged	Unengaged
Phenotype sample size, N	26	32	18	35	103
Age (years), mean (SD)[b]	44.5 (7.1)	42.2 (5.6)	44.5 (7.1)	46.7 (6.6)	43.5 (6.0)
Sex (female), n (%)	26 (100)	32 (100)	17 (94.4)	31 (88.6)	92 (89.3)
Household income (≥CAD $80,000; US $63,771), n (%)	16 (61.5)	20 (62.5)	9 (50)	17 (48.6)	57 (55.3)
Parental education (more than a Bachelor's degree), n (%)	9 (34.6)	20 (62.5)	9 (50)	14 (40)	41 (39.8)
Parental marital status (married, common law, or living with a partner), n (%)[c]	25 (96.2)	27 (84.4)	12 (66.7)	26 (74.3)	73 (70.9)
Race or ethnicity (White or European), n (%)	20 (76.9)	19 (59.3)	11 (61.1)	20 (57.1)	59 (57.3)
Recruitment through a clinical setting, n (%)	9 (34.6)	14 (43.8)	8 (44.4)	17 (48.6)	47 (45.6)

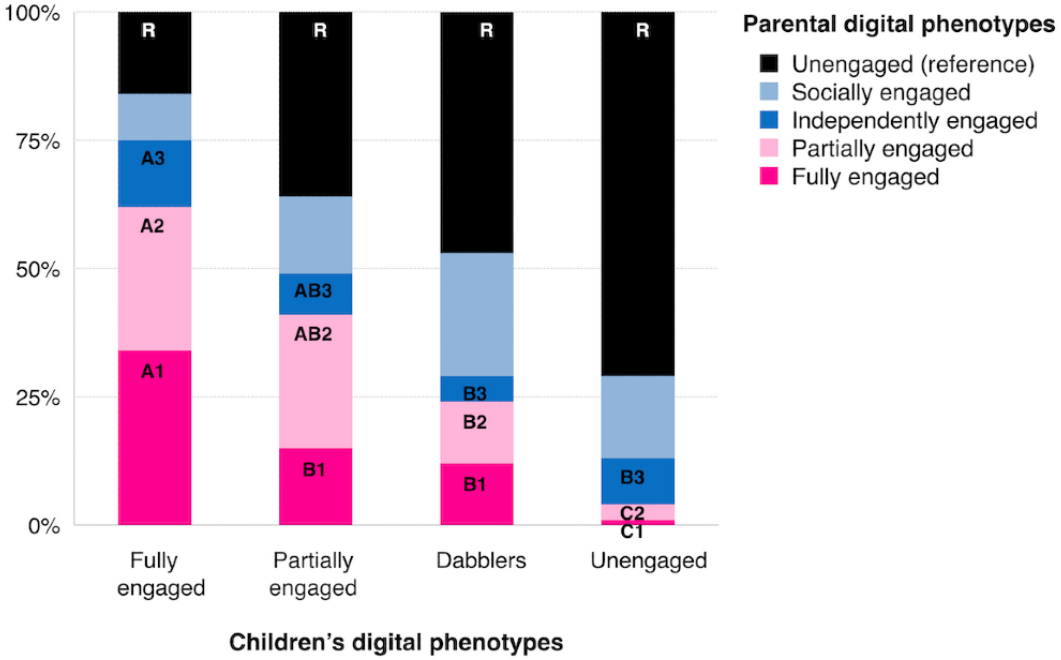
[a]Predictors' reference groups are: male, household income <CAD $80,000 (US $63,771), parental educational attainment lower than a bachelor's degree, single parents, people who did not self-identify as having a White or European descent, and recruitment through Facebook.

[b]The age of both *fully engaged* and *unengaged* parents significantly differs from the age of *socially engaged* parents.

[c]The marital status of *fully engaged* parents significantly differs from both *independently engaged* and *unengaged* parents' marital status.

XSL•FO
RenderX

Figure 6. Associations between children's and parents' digital phenotypes (N=214). Vertical bars represent the proportion of parents with each phenotype with a given child phenotype. Within groups that share the same number and color, groups that do not share the same letter are significantly different from one another and are compared with the reference group (ie, unengaged users).



Changes in Health Outcomes Across Digital Phenotypes

Table 4 summarizes the 3-month changes in *z*BMI, diet, physical activity, and screen time across children's and parents' digital phenotypes, with statistically significant ($P<.05$) or borderline significant ($P<.10$) comparisons shown in Figure 7, where panels A to C show child outcomes across child phenotypes, and panels D and E show child outcomes across parent phenotypes.

Multiple group comparisons showed that changes in the total sugar intake of children differed across phenotypes ($P=.01$; Figure 7A-C). Children belonging to the *fully engaged* ($P=.01$; $f^2=0.04$) or *partially engaged* ($P=.004$; $f^2=0.05$) phenotypes reduced their total sugar intake over 3 months compared with children in the *unengaged* phenotype (reference group), who increased their total sugar intake over time. Regarding children's total daily energy intake and energy intake from sugary beverages, we found borderline differences ($P=.07$ and $P=.09$) that became significant in individual pairwise comparisons. Children from the *fully engaged* ($P=.03$; $f^2=0.01$), *partially engaged* ($P=.04$; $f^2=0.03$), and *dabblers* ($P=.03$; $f^2=0.00$) phenotypes decreased their total energy intake over 3 months compared with the *unengaged* children who increased their daily energy intake over time. Finally, children from the *partially engaged* phenotype decreased their intake of sugary beverages compared with *unengaged* children who did not ($P=.01$; $f^2=0.02$). In this case, *fully engaged* children did not differ significantly from *unengaged* children; however, as shown in Figure 7C, children's intake of sugary beverages in the *fully engaged* group trended downward, whereas *unengaged* children's intake trended upward ($P=.12$).

Differential changes in outcomes among children were also observed across the parental phenotypes (Figure 7D and 7E). Multiple group comparisons showed borderline significant changes in children's *z*BMI and total daily energy intake across parental phenotypes ($P=.06$ and $P=.08$, respectively), which became significant in individual pairwise comparisons. Specifically, children whose parents were *fully engaged* significantly decreased their *z*BMI ($P=.01$; $f^2=0.05$) compared with children with *unengaged* parents (reference group) whose *z*BMI slightly increased. Similarly, children whose parents belonged to the *independently engaged* phenotype decreased their daily caloric intake ($P=.02$; $f^2=0.03$) compared with children with *unengaged* parents whose daily caloric intake increased over 3 months. Figure 7E also shows trends of decreased caloric intake among children with *fully* and *partially engaged* parents compared with children with *unengaged* parents; however, these trends were not statistically significant ($P=.11$ and $P=.07$, respectively).

XSL•FO
RenderX

Table 4. Changes in children's and parents' health outcomes across digital phenotypes (N=214).

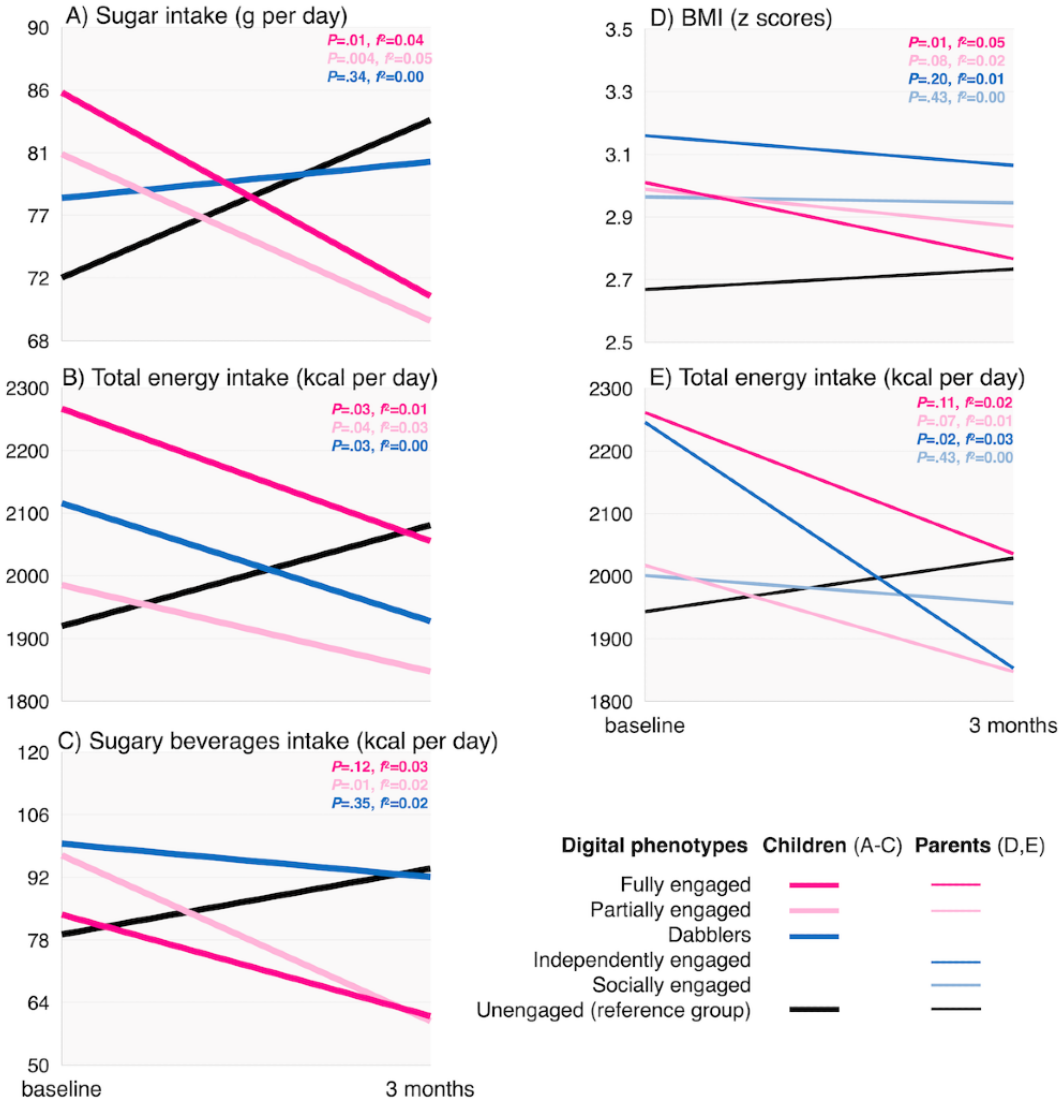
Participants and health outcomes	Child phenotypes		Parent phenotypes	
	Chi-square (*df*)	*P* value	Chi-square (*df*)	*P* value
Children				
BMI *z* scores	0.5 (3)	.93	9.1 (4)[a]	.06[a]
Total energy, daily (kcal per day)	7.2 (3)[a]	.07[a]	8.2 (4)[a]	.08[a]
Healthy Eating Index (range 0-100 points)	0.3 (3)	.96	1.9 (4)	.76
Fruits and vegetables (daily servings)	0.2 (3)	.98	3.6 (4)	.47
Saturated and trans fat (g per day)	5.4 (3)	.15	5.5 (4)	.24
Total fiber (g per day)	1.5 (3)	.68	1.5 (4)	.84
Total sugar (g per day)	11.8 (3)[a]	.01[a]	5.4 (4)	.25
Sugary beverages (kcal per day)	6.7 (3)[a]	.09[a]	5.5 (4)	.24
Total physical activity (minutes per week)	2.5 (3)	.47	0.6 (4)	.96
Fitbit (steps per day)	2.1 (3)	.55	3.2 (4)	.52
Screen time (minutes per day)	4.9 (3)	.18	2.3 (4)	.69
Parents				
Daily frequency of sugary beverages (times per day)	N/A[b]	N/A	1.2 (4)	.88
Daily frequency of fruit juice (times per day)	N/A	N/A	6.5 (4)	.16
Fruit and vegetables (servings per day)	N/A	N/A	2.4 (4)	.67
Walking (minutes per day)	N/A	N/A	4.0 (4)	.41
Moderate and vigorous physical activity (minutes per day)	N/A	N/A	2.0 (4)	.75
Screen time (minutes per week)	N/A	N/A	7.1 (4)	.13

[a]Indicate significant ($P<.05$) or borderline significant ($P<.10$) interactions (time×digital phenotype) for which pairwise comparisons between phenotypes were further explored.

[b]N/A: not applicable.

XSL•FO
RenderX

Figure 7. Changes in children's health outcomes across children's (A-C) and parents' (D and E) digital phenotypes (N=214). Comparison of each phenotype versus unengaged phenotype (reference group). *P* value indicates significance level and f2 indicates Cohen effect size.



Discussion

Principal Findings

This is the first digital phenotyping study of an mHealth intervention targeting health behavior changes among children with overweight or obesity and their parents. We evaluated user typologies based on how children and parents interacted with different features of the Aim2Be app. We found 4 child (*unengaged*, *dabblers*, *partially engaged*, and *fully engaged*) and 5 parent (*unengaged*, *socially*, *independent*, *partially engaged*, and *fully engaged*) phenotypes, which illustrate the ways in which participants used the behavioral, gamified, and social features of the Aim2Be app. As expected, based on Aim2Be's conceptual framework [25], our results demonstrated that specific patterns of use supported behavior change, whereas others did not, meaning that greater engagement with the active ingredients of the app improved children's dietary and weight outcomes.

Comparison With Prior Work

Given the scarcity of research on the digital phenotypes of mHealth users in the context of childhood obesity, it is difficult to compare our findings with those of previous studies. However, our results are similar to a recently published study profiling children's (but not parents') engagement with an older version of Aim2Be [33], where the 4 child profiles that emerged were similar, although our study examined 6 additional app features. Interestingly, the results previously observed in the prevention context [33] were replicated in our study using a clinical sample of children. Importantly, users with distinct patterns of engagement obtained different health benefits depending on whether they interacted with the active ingredients of the app. When lifestyle behavior modification interventions required in-person attendance, dose-response studies identified a minimum of 26 hours of contact for the intervention to improve children's outcomes [4]. However, our digital phenotype analyses illustrate that new approaches are needed to conduct dose-response analyses in the context of mHealth interventions, especially when users have the freedom to select which app features they engage with. As users interact with the Aim2Be app quite differently, this variability must be accounted for when assessing whether the intervention can influence the mediators and outcomes targeted by the app.

XSL•FO
RenderX

In this study, we found that *fully engaged* children with Aim2Be (eg, set goals, completed tasks, and read articles) experienced more desirable behavior changes than *unengaged* users. Specifically, children who engaged more fully with the app decreased their intake of total daily calories, total sugars, and sugary beverages. Our findings align with existing research [22,33] suggesting that mHealth interventions have the potential to improve children's dietary behaviors. Furthermore, in exploratory analyses examining the aims that were most often set and completed among Aim2Be users (data not shown), we found that "Drop sugary drinks" was the most common aim chosen by children, which validates our findings related to lower total sugar and energy from sugary drinks among *fully* and *partially engaged* children. These results highlight the importance of increasing engagement with the app's *active ingredients*, namely, setting specific goals and completing tasks related to those goals to promote health behavior change among children.

In this study, *fully engaged* children were more likely to be younger and have *fully* or *partially engaged* parents. These associations could indicate that the app was more appealing to younger children, as shown by other research [34], or that parents dedicated more attention to their children when they were younger than when they were older. In addition, younger children might be more easily influenced by their parents, which may explain their use of the Aim2Be app. These findings are aligned with previous studies reporting that parental self-monitoring (a behavioral strategy) and adherence to eHealth interventions were significant predictors of adolescents' self-monitoring and adherence [33,64].

We also found that children whose parents were *fully* or *partially engaged* with the app's behavioral features decreased their zBMI and total daily energy intake more than children whose parents only engaged with the social features or who did not engage with the app at all. Our findings are consistent with a qualitative study [65] showing that participation *as a family* is one of the main factors identified by both children and parents to facilitate behavior change. Indeed, current guidelines [4,15,66] for the treatment of childhood obesity include a family-based approach. Taken together, our findings reinforce the critical role that parents play in lifestyle interventions to support their children's adherence and improvement of health outcomes, even in the mHealth context.

This study also revealed that family structure was associated with parental phenotypes. Fewer single-parent households belonged to the *fully* and *partially engaged* phenotypes than parents who lived with a partner or were married, which may reflect that more independent, time-scarce (and therefore task-oriented) parents [67]. Interestingly, single-parent households were also likely to belong to the *independently engaged* phenotype (ie, parents who only engaged with the behavioral app features such as aims and tasks), and children whose parents belonged to this phenotype reduced more of their total daily energy than other phenotypes. In fact, previous research found that parents of young children decreased their use of mHealth apps when they had limited time or only used the app to find specific information of interest [67]. This could explain why *independently engaged* parents did not use the gamified or social domains but used the domain exclusively focused on behavioral change and why their children decreased their energy intake over time.

Limitations and Strengths

This study had several limitations and strengths. First, our sample was relatively small and not powered to detect significant changes across multiple digital phenotypes in these secondary analyses. This could have limited our ability to detect clinically meaningful changes in health outcomes, although some changes were observed. In addition, overall adherence to the app was low, which limited our ability to detect more phenotypes and perhaps to observe some between-group changes. Moreover, our study included a clinical sample (children with overweight or obesity); thus, our findings are limited to this population. Nevertheless, we used a detailed dietary assessment (24-hour dietary recalls), both self-reported and objective measures of physical activity, and objective measures of app usability through app analytics. Finally, we used a novel approach to examine intervention efficacy, which showed positive effects that are not observed [38] using more traditional analysis.

Future Directions

Overall, 3 key messages from our findings point to future directions in mHealth research. First, even in the mHealth context, parental engagement matters as it can increase children's adherence to a lifestyle intervention and provide the household environment that supports behavior change. Thus, whether a lifestyle intervention is delivered in person or on the web, parents should be involved as they are active agents of change. Second, dose-response analyses should assess how (and not only how much) the app is being used by the participants, as users' full engagement with the *active ingredients* of the app seems to be a critical factor for the success of mHealth interventions. Finally, as participants' engagement with specific features of an app is key to promoting behavior change, future research should examine how we design program components that ensure users interact with the active ingredients of the mHealth intervention.

Conclusions

This study showed that distinct patterns of use exist among both parents and children who used a family-based lifestyle mHealth app, namely, Aim2Be. Identifying who uses mHealth apps and how can help us understand and develop more tailored interventions to support various users in a health behavior change process. Our findings point to the importance of optimizing users' full engagement with the *active ingredients* of the app as a critical factor for the success of mHealth interventions and highlight the need for further research to understand program design elements that can influence participant engagement.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank the families who participated in the study.

The Childhood Obesity Foundation received funding, in part, from the Public Health Agency of Canada, with matched financial and in-kind funds from Ayogo Health Inc (Aim2Be developer), Merck Canada Inc, Heart and Stroke, Obesity Canada, Diabetes Canada, Dietitians of Canada, Canadian Society for Exercise Physiology, Craving Change, David Suzuki Foundation, and Pacific Blue Cross Foundation. Supplemental funding was obtained from a Team Grant in Bariatric Care (Team to Address Bariatric Care in Canadian Children—Team ABC3) from the Canadian Institutes of Health Research (Institute of Nutrition, Metabolism, and Diabetes); from Alberta Health Services, Alberta Innovates, Obesity Canada, the Ontario Ministry of Health and Long-Term Care, and the Women and Children's Health Research Institute.

Data Availability

LCM should be contacted at the BC Children's Hospital Research Institute and School of Population and Public Health University of British Columbia (lmasse@bcchr.ubc.ca) regarding access to data and materials.

Authors' Contributions

LCM, J Hamilton, J Ho, AB, KMM, and GDCB designed the Aim2Be randomized controlled trial with LCM as the lead of the randomized controlled trial. LCM oversaw the data collection, and EJB managed part of the data collection. OD-JG and LCM conceptualized the paper with input from CNT-L. CNT-L cleaned the data set, and LCM reviewed the cleaning code with help from OD-JG. OD-JG performed the analyses. OD-JG, CNT-L, EJB, and LCM provided input in the interpretation of the data. OD-JG drafted the manuscript with help from CNT-L, LCM, and EJB. All authors critically reviewed a draft of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

LCM received salary support to conduct this research, which was provided by the BC Children's Hospital Research Institute. OD-JG received a postdoctoral salary from the University of British Columbia and received PhD scholarships from the National Council of Science and Technology (Conacyt) of Mexico and from Universidad Iberoamericana of Mexico City. CNT-L received a postdoctoral fellowship from the Canadian Institutes of Health Research. EJB received a postdoctoral fellowship from the BC Children's Hospital Research Institute. GDCB received funding from an Alberta Health Services Chair in Obesity Research.

Multimedia Appendix 1
CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist.
[PDF File (Adobe PDF File), 1375 KB - jmir_v24i6e35285_app1.pdf]

Multimedia Appendix 2
Relative risk ratios of the predictors of the digital phenotypes.
[DOCX File , 22 KB - jmir_v24i6e35285_app2.docx]

References

1. Roberts KC, Shields M, de Groh M, Aziz A, Gilbert J. Overweight and obesity in children and adolescents: results from the 2009 to 2011 Canadian Health Measures Survey. Health Rep 2012 Sep;23(3):37-41 [FREE Full text] [Medline: 23061263]
2. Baños RM, Oliver E, Navarro J, Vara MD, Cebolla A, Lurbe E, et al. Efficacy of a cognitive and behavioral treatment for childhood obesity supported by the ETIOBE web platform. Psychol Health Med 2019 Jul 16;24(6):703-713. [doi: 10.1080/13548506.2019.1566622] [Medline: 30648879]
3. Kang NR, Kwack YS. An update on mental health problems and cognitive behavioral therapy in pediatric obesity. Pediatr Gastroenterol Hepatol Nutr 2020 Jan;23(1):15-25 [FREE Full text] [doi: 10.5223/pghn.2020.23.1.15] [Medline: 31988872]
4. Guideline Development Panel for Treatment of Obesity, American Psychological Association. Summary of the clinical practice guideline for multicomponent behavioral treatment of obesity and overweight in children and adolescents. Am Psychol 2020;75(2):178-188 [FREE Full text] [doi: 10.1037/amp0000530] [Medline: 32052993]
5. Kang Sim DE, Strong DR, Manzano MA, Rhee KE, Boutelle KN. Evaluation of dyadic changes of parent-child weight loss patterns during a family-based behavioral treatment for obesity. Pediatr Obes 2020 Jun;15(6):e12622. [doi: 10.1111/ijpo.12622] [Medline: 32048808]
6. Com a L, David O, David D. Outcomes and mechanisms of change in cognitive-behavioral interventions for weight loss: a meta-analysis of randomized clinical trials. Behav Res Ther 2020 Jun 02;132:103654. [doi: 10.1016/j.brat.2020.103654] [Medline: 32683134]

XSL•FO
RenderX

7. Vignolo M, Rossi F, Bardazza G, Pistorio A, Parodi A, Spigno S, et al. Five-year follow-up of a cognitive-behavioural lifestyle multidisciplinary programme for childhood obesity outpatient treatment. Eur J Clin Nutr 2008 Sep;62(9):1047-1057. [doi: 10.1038/sj.ejcn.1602819] [Medline: 17554247]
8. Luzier J, Berlin K, Weeks J. Behavioral treatment of pediatric obesity: review and future directions. Children's Health Care 2010 Oct;39(4):312-334. [doi: 10.1080/02739615.2010.516202]
9. Miri SF, Javadi M, Lin C, Griffiths MD, Björk M, Pakpour AH. Effectiveness of cognitive-behavioral therapy on nutrition improvement and weight of overweight and obese adolescents: a randomized controlled trial. Diabetes Metab Syndr 2019;13(3):2190-2197. [doi: 10.1016/j.dsx.2019.05.010] [Medline: 31235156]
10. Partridge SR, Raeside R, Singleton A, Hyun K, Redfern J. Effectiveness of text message interventions for weight management in adolescents: systematic review. JMIR Mhealth Uhealth 2020 May 26;8(5):e15849 [FREE Full text] [doi: 10.2196/15849] [Medline: 32348264]
11. Hammersley ML, Jones RA, Okely AD. Parent-focused childhood and adolescent overweight and obesity eHealth interventions: a systematic review and meta-analysis. J Med Internet Res 2016 Jul 21;18(7):e203 [FREE Full text] [doi: 10.2196/jmir.5893] [Medline: 27443862]
12. Turner T, Spruijt-Metz D, Wen CK, Hingle MD. Prevention and treatment of pediatric obesity using mobile and wireless technologies: a systematic review. Pediatr Obes 2015 Dec;10(6):403-409 [FREE Full text] [doi: 10.1111/ijpo.12002] [Medline: 25641770]
13. Tully L, Burls A, Sorensen J, El-Moslemany R, O'Malley G. Mobile health for pediatric weight management: systematic scoping review. JMIR Mhealth Uhealth 2020 Jun 03;8(6):e16214 [FREE Full text] [doi: 10.2196/16214] [Medline: 32490849]
14. Badawy SM, Kuhns LM. Texting and mobile phone app interventions for improving adherence to preventive behavior in adolescents: a systematic review. JMIR Mhealth Uhealth 2017 Apr 19;5(4):e50 [FREE Full text] [doi: 10.2196/mhealth.6837] [Medline: 28428157]
15. Tate EB, Spruijt-Metz D, O'Reilly G, Jordan-Marsh M, Gotsis M, Pentz MA, et al. mHealth approaches to child obesity prevention: successes, unique challenges, and next directions. Transl Behav Med 2013 Dec;3(4):406-415 [FREE Full text] [doi: 10.1007/s13142-013-0222-3] [Medline: 24294329]
16. Fedele DA, Cushing CC, Fritz A, Amaro CM, Ortega A. Mobile health interventions for improving health outcomes in youth: a meta-analysis. JAMA Pediatr 2017 May 01;171(5):461-469 [FREE Full text] [doi: 10.1001/jamapediatrics.2017.0042] [Medline: 28319239]
17. Duan Y, Shang B, Liang W, Du G, Yang M, Rhodes RE. Effects of eHealth-based multiple health behavior change interventions on physical activity, healthy diet, and weight in people with noncommunicable diseases: systematic review and meta-analysis. J Med Internet Res 2021 Feb 22;23(2):e23786 [FREE Full text] [doi: 10.2196/23786] [Medline: 33616534]
18. 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]
19. Partridge SR, McGeechan K, Hebden L, Balestracci K, Wong AT, Denney-Wilson E, et al. Effectiveness of a mHealth Lifestyle Program With Telephone Support (TXT2BFiT) to prevent unhealthy weight gain in young adults: randomized controlled trial. JMIR Mhealth Uhealth 2015 Jun 15;3(2):e66 [FREE Full text] [doi: 10.2196/mhealth.4530] [Medline: 26076688]
20. Ek A, Delisle Nyström C, Chirita-Emandi A, Tur JA, Nordin K, Bouzas C, et al. A randomized controlled trial for overweight and obesity in preschoolers: the More and Less Europe study - an intervention within the STOP project. BMC Public Health 2019 Jul 15;19(1):945 [FREE Full text] [doi: 10.1186/s12889-019-7161-y] [Medline: 31307412]
21. Delisle Nyström C, Sandin S, Henriksson P, Henriksson H, Maddison R, Löf M. A 12-month follow-up of a mobile-based (mHealth) obesity prevention intervention in pre-school children: the MINISTOP randomized controlled trial. BMC Public Health 2018 May 24;18(1):658 [FREE Full text] [doi: 10.1186/s12889-018-5569-4] [Medline: 29793467]
22. Nollen NL, Mayo MS, Carlson SE, Rapoff MA, Goggin KJ, Ellerbeck EF. Mobile technology for obesity prevention: a randomized pilot study in racial- and ethnic-minority girls. Am J Prev Med 2014 Apr;46(4):404-408 [FREE Full text] [doi: 10.1016/j.amepre.2013.12.011] [Medline: 24650843]
23. Partridge SR, Allman-Farinelli M, McGeechan K, Balestracci K, Wong AT, Hebden L, et al. Process evaluation of TXT2BFiT: a multi-component mHealth randomised controlled trial to prevent weight gain in young adults. Int J Behav Nutr Phys Act 2016 Jan 19;13:7 [FREE Full text] [doi: 10.1186/s12966-016-0329-2] [Medline: 26785637]
24. Tripicchio GL, Ammerman AS, Neshteruk C, Faith MS, Dean K, Befort C, et al. Technology components as adjuncts to family-based pediatric obesity treatment in low-income minority youth. Child Obes 2017 Dec;13(6):433-442 [FREE Full text] [doi: 10.1089/chi.2017.0021] [Medline: 28727927]
25. Mâsse LC, Vlaar J, Macdonald J, Bradbury J, Warshawski T, Buckler E, et al. Aim2Be mHealth intervention for children with overweight and obesity: study protocol for a randomized controlled trial. Trials 2020 Feb 03;21(1):132 [FREE Full text] [doi: 10.1186/s13063-020-4080-2] [Medline: 32014057]
26. Kim Y, Oh B, Shin H. Effect of mHealth with offline antiobesity treatment in a community-based weight management program: cross-sectional study. JMIR Mhealth Uhealth 2020 Jan 21;8(1):e13273 [FREE Full text] [doi: 10.2196/13273] [Medline: 31961335]

XSL•FO
RenderX

27. Nilsen W, Kumar S, Shar A, Varoquiers C, Wiley T, Riley WT, et al. Advancing the science of mHealth. J Health Commun 2012;17 Suppl 1:5-10. [doi: 10.1080/10810730.2012.677394] [Medline: 22548593]
28. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. Nat Biotechnol 2015 May;33(5):462-463. [doi: 10.1038/nbt.3223] [Medline: 25965751]
29. Radhakrishnan K, Kim MT, Burgermaster M, Brown RA, Xie B, Bray MS, et al. The potential of digital phenotyping to advance the contributions of mobile health to self-management science. Nurs Outlook 2020;68(5):548-559. [doi: 10.1016/j.outlook.2020.03.007] [Medline: 32402392]
30. Yang Q, Hatch D, Crowley MJ, Lewinski AA, Vaughn J, Steinberg D, et al. Digital phenotyping self-monitoring behaviors for individuals with type 2 diabetes mellitus: observational study using latent class growth analysis. JMIR Mhealth Uhealth 2020 Jun 11;8(6):e17730 [FREE Full text] [doi: 10.2196/17730] [Medline: 32525492]
31. Teo J, Davila S, Yang C, Hii A, Pua C, Yap J, et al. Digital phenotyping by consumer wearables identifies sleep-associated markers of cardiovascular disease risk and biological aging. Commun Biol 2019;2:361 [FREE Full text] [doi: 10.1038/s42003-019-0605-1] [Medline: 31602410]
32. Ebner-Priemer U, Mühlbauer E, Neubauer A, Hill H, Beier F, Santangelo P, et al. Digital phenotyping: towards replicable findings with comprehensive assessments and integrative models in bipolar disorders. Int J Bipolar Disord 2020 Nov 17;8(1):35 [FREE Full text] [doi: 10.1186/s40345-020-00210-4] [Medline: 33211262]
33. Lin Y, Mâsse LC. A look at engagement profiles and behavior change: a profile analysis examining engagement with the Aim2Be lifestyle behavior modification app for teens and their families. Prev Med Rep 2021 Dec;24:101565 [FREE Full text] [doi: 10.1016/j.pmedr.2021.101565] [Medline: 34976631]
34. Mâsse LC, Watts AW, Barr SI, Tu AW, Panagiotopoulos C, Geller J, et al. Individual and household predictors of adolescents' adherence to a web-based intervention. Ann Behav Med 2015 Jun;49(3):371-383. [doi: 10.1007/s12160-014-9658-z] [Medline: 25270826]
35. Ernsting C, Dombrowski SU, Oedekoven M, O Sullivan JL, Kanzler M, Kuhlmey A, et al. Using smartphones and health apps to change and manage health behaviors: a population-based survey. J Med Internet Res 2017 Apr 05;19(4):e101 [FREE Full text] [doi: 10.2196/jmir.6838] [Medline: 28381394]
36. Smahel D, Elavsky S, Machackova H. Functions of mHealth applications: a user's perspective. Health Informatics J 2019 Sep 10;25(3):1065-1075 [FREE Full text] [doi: 10.1177/1460458217740725] [Medline: 29121831]
37. Chai LK, Collins CE, May C, Ashman A, Holder C, Brown LJ, et al. Feasibility and efficacy of a web-based family telehealth nutrition intervention to improve child weight status and dietary intake: a pilot randomised controlled trial. J Telemed Telecare 2021 Apr;27(3):146-158. [doi: 10.1177/1357633X19865855] [Medline: 31364474]
38. De-Jongh González O, Tugault-Lafleur C, Hamilton J, Ho J, Buchholz A, Morrison K, et al. Efficacy of the Aim2Be mHealth intervention for children with overweight and obesity: a randomized controlled trial. In: Proceedings of the Society of Behavioral Medicine (SBM) 42nd Annual Meeting. 2021 Presented at: Society of Behavioral Medicine (SBM) 42nd Annual Meeting; Apr 12-16, 2021; Virtual URL: https://doi.org/10.1093/abm/kaab020 [doi: 10.1093/abm/kaab020]
39. 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]
40. de Onis M, Onyango AW, Borghi E, Siyam A, Nishida C, Siekmann J. Development of a WHO growth reference for school-aged children and adolescents. Bull World Health Organ 2007 Sep;85(9):660-667 [FREE Full text] [doi: 10.2471/blt.07.043497] [Medline: 18026621]
41. Aim2Be homepage. Aim2Be. URL: https://www.aim2be.ca/ [accessed 2021-10-26]
42. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
43. Deterding S. The lens of intrinsic skill atoms: a method for gameful design. Human Comput Interact 2015 May 15;30(3-4):294-335. [doi: 10.1080/07370024.2014.993471]
44. Ryan R, Rigby C, Przybylski A. The motivational pull of video games: a self-determination theory approach. Motiv Emot 2006 Nov 29;30(4):344-360 [FREE Full text] [doi: 10.1007/s11031-006-9051-8]
45. Designing digital tools for patient engagement. Ayogo. URL: https://ayogo.com/wp-content/uploads/2018/ [accessed 2021-08-26]
46. Tremblay MS, Carson V, Chaput J, Connor Gorber S, Dinh T, Duggan M, et al. Canadian 24-hour movement guidelines for children and youth: an integration of physical activity, sedentary behaviour, and sleep. Appl Physiol Nutr Metab 2016 Jun;41(6 Suppl 3):S311-S327 [FREE Full text] [doi: 10.1139/apnm-2016-0151] [Medline: 27306437]
47. Canada's Food Guide. Government of Canada. URL: https://food-guide.canada.ca/en/ [accessed 2021-11-23]
48. Measuring children's height and weight accurately at home. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/healthyweight/assessing/bmi/childrens_bmi/measuring_children.html [accessed 2021-08-23]
49. Sarkkola C, Rounge TB, Simola-Ström S, von Kraemer S, Roos E, Weiderpass E. Validity of home-measured height, weight and waist circumference among adolescents. Eur J Public Health 2016 Dec;26(6):975-977. [doi: 10.1093/eurpub/ckw133] [Medline: 27578829]

XSL•FO
RenderX

50. Hanning RM, Royall D, Toews JE, Blashill L, Wegener J, Driezen P. Web-based Food Behaviour Questionnaire: validation with grades six to eight students. Can J Diet Pract Res 2009;70(4):172-178. [doi: 10.3148/70.4.2009.172] [Medline: 19958572]
51. Garriguet D. Diet quality in Canada. Health Rep 2009 Sep;20(3):41-52 [FREE Full text] [Medline: 19813438]
52. Canadian Community Health Survey (CCHS) - 2016. Statistics Canada. 2016. URL: https://www23.statcan.gc.ca/imdb/p3Instr.pl?Function=assembleInstr&Item_Id=260675 [accessed 2022-03-29]
53. Fitabase homepage. Fitabase. URL: https://www.fitabase.com/ [accessed 2021-10-20]
54. Kowalski K, Crocker R, Donen R. The Physical Activity Questionnaire for Older Children (PAQ-C) and Adolescents (PAQ-A) Manual. University of Saskatchewan. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.454.4555&rep=rep1&type=pdf [accessed 2021-10-20]
55. Wang JJ, Baranowski T, Lau WP, Chen TA, Pitkethly AJ. Validation of the Physical Activity Questionnaire for older Children (PAQ-C) among Chinese children. Biomed Environ Sci 2016 Mar;29(3):177-186 [FREE Full text] [doi: 10.3967/bes2016.022] [Medline: 27109128]
56. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
57. Rosenberg DE, Norman GJ, Wagner N, Patrick K, Calfas KJ, Sallis JF. Reliability and validity of the Sedentary Behavior Questionnaire (SBQ) for adults. J Phys Act Health 2010 Nov;7(6):697-705. [doi: 10.1123/jpah.7.6.697] [Medline: 21088299]
58. Muthén LK, Muthén BO. Mplus User's Guide. Eighth Edition. Los Angeles, CA: Muthén & Muthén; 2017.
59. Nylund-Gibson K, Choi A. Ten frequently asked questions about latent class analysis. Translational Issues Psychol Sci 2018 Dec;4(4):440-461 [FREE Full text] [doi: 10.1037/tps0000176]
60. Wurpts IC, Geiser C. Is adding more indicators to a latent class analysis beneficial or detrimental? Results of a Monte-Carlo study. Front Psychol 2014 Aug 21;5:920 [FREE Full text] [doi: 10.3389/fpsyg.2014.00920] [Medline: 25191298]
61. How can I estimate effect size for mixed models? UCLA. URL: https://stats.idre.ucla.edu/stata/faq/how-can-i-estimate-effect-size-for-mixed/ [accessed 2021-06-20]
62. Selya AS, Rose JS, Dierker LC, Hedeker D, Mermelstein RJ. A practical guide to calculating Cohen's f(2), a measure of local effect size, from PROC MIXED. Front Psychol 2012;3:111 [FREE Full text] [doi: 10.3389/fpsyg.2012.00111] [Medline: 22529829]
63. Stata Statistical Software: Release 17. College Station, TX: StataCorp LLC; 2017.
64. Tu A, Watts A, Chanoine J, Panagiotopoulos C, Geller J, Brant R, et al. Does parental and adolescent participation in an e-health lifestyle modification intervention improves weight outcomes? BMC Public Health 2017 Apr 24;17(1):352 [FREE Full text] [doi: 10.1186/s12889-017-4220-0] [Medline: 28438202]
65. Watson PM, Dugdill L, Pickering K, Hargreaves J, Staniford LJ, Owen S, et al. Distinguishing factors that influence attendance and behaviour change in family-based treatment of childhood obesity: a qualitative study. Br J Health Psychol 2021 Feb;26(1):67-89. [doi: 10.1111/bjhp.12456] [Medline: 32710510]
66. Report of the commission on ending childhood obesity. World Health Organization. URL: https://www.who.int/publications/i/item/9789241510066 [accessed 2022-03-29]
67. 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;10:397 [FREE Full text] [doi: 10.3389/fendo.2019.00397] [Medline: 31293515]

Abbreviations

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
LCA: latent class analysis
mHealth: mobile health
zBMI: BMI z score

XSL•FO
RenderX

Edited by T Leung; submitted 30.11.21; peer-reviewed by B Shang, S Badawy; comments to author 15.02.22; revised version received 07.04.22; accepted 23.04.22; published 22.06.22.

Please cite as:
De-Jongh González O, Tugault-Lafleur CN, Buckler EJ, Hamilton J, Ho J, Buchholz A, Morrison KM, Ball GDC, Mâsse LC
The Aim2Be mHealth Intervention for Children With Overweight or Obesity and Their Parents: Person-Centered Analyses to Uncover Digital Phenotypes
J Med Internet Res 2022;24(6):e35285
URL: https://www.jmir.org/2022/6/e35285
doi:10.2196/35285
PMID:35731547

©Olivia De-Jongh González, Claire N Tugault-Lafleur, E Jean Buckler, Jill Hamilton, Josephine Ho, Annick Buchholz, Katherine M Morrison, Geoff DC Ball, Louise C Mâsse. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 22.06.2022. 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

Long-term Effects of a Social Media–Based Intervention (Run4Love) on Depressive Symptoms of People Living With HIV: 3-Year Follow-up of a Randomized Controlled Trial

Yan Guo[1,2], PhD; Yingqi Li[1], MB; Chuanchuan Yu[1], MB; He Xu[1], MB; Y Alicia Hong[3], PhD; Xiaolan Wang[1], MB; Nanxiang Zhang[1], MB; Yu Zeng[1], MB; Aliza Monroe-Wise[4], MSc; Linghua Li[5], MD; Cong Liu[5], MSN; Weiping Cai[5], MD; Aihua Lin[1,6], PhD

[1]Department of Medical Statistics, School of Public Health, Sun Yat-sen University, Guangzhou, China

[2]Department of Population and Quantitative Health Sciences, University of Massachusetts Chan Medical School, Worcester, MA, United States

[3]Department of Health Administration and Policy, College of Health and Human Services, George Mason University, Fairfax, VA, United States

[4]Department of Global Health, University of Washington, Seattle, WA, United States

[5]Department of Infectious Diseases, Guangzhou Eighth People's Hospital, Guangzhou, China

[6]Department of Health Service and Management, Guangzhou Xinhua University, Guangzhou, China

Corresponding Author:
Yan Guo, PhD
Department of Medical Statistics
School of Public Health
Sun Yat-sen University
#74 Zhongshan 2nd Road
Guangzhou, 510080
China
Phone: 86 020 87334202
Email: Yan.Guo1@umassmed.edu

Abstract

Background: Emerging studies have shown the effectiveness of mobile health (mHealth) interventions in reducing depressive symptoms among people living with HIV. Most of these studies included only short-term follow-up, with limited data on long-term effects.

Objective: The purpose of this study is to assess the long-term effects of a randomized controlled trial called Run4Love on depressive symptoms among people living with HIV at 1-year and 3-year follow-ups.

Methods: A total of 300 people living with HIV with depressive symptoms were recruited and randomized to an intervention or a control group in Guangzhou, China, from September 2017 to January 2018. The intervention group received a 3-month Run4Love program, including adapted evidence-based cognitive behavioral stress management courses and exercise promotion via WeChat (Tencent), a popular social media app. The control group received usual care and a brochure on nutrition. The primary outcome was reduction in depressive symptoms, measured using the Center for Epidemiological Studies–Depression (CES-D) scale. Data used in this study were collected at baseline and at the 1-year and 3-year follow-ups. Generalized estimating equations were used to examine the group differences at 1-year and 3-year follow-ups.

Results: Approximately half of the participants completed the assessment at 1-year (149/300, 49.7%) and 3-year (177/300, 59%) follow-ups. At 1-year follow-up, participants in the intervention group reported significant reduction in depressive symptoms compared with the control group (CES-D: from 23.9 to 18.1 in the intervention group vs from 24.3 to 23.3 in the control group; mean −4.79, SD 13.56; 95% CI −7.78 to −1.81; *P*=.002). At 3-year follow-up, between-group difference in CES-D remained statistically significant (from 23.9 to 20.5 in the intervention group vs from 24.3 to 24.4 in the control group; mean −3.63, SD 13.35; 95% CI −6.71 to −0.54; *P*=.02). No adverse events were reported during the 3-year follow-up period.

Conclusions: The mHealth intervention, Run4Love, significantly reduced depressive symptoms among people living with HIV, and the intervention effects were sustained at 1-year and 3-year follow-ups. Further research is needed to explore the mechanisms of the long-term effects of mHealth interventions such as Run4Love and to implement these effective interventions among people living with HIV.

XSL•FO
RenderX

Trial Registration: Chinese Clinical Trial Registry ChiCTR-IPR-17012606; https://trialsearch.who.int/Trial2.aspx?TrialID=ChiCTR-IPR-17012606

International Registered Report Identifier (IRRID): RR2-10.2196/10274

(J Med Internet Res 2022;24(6):e36809) doi:10.2196/36809

KEYWORDS

HIV; depressive symptoms; mobile health; mHealth; social media–based; long-term intervention effect

Introduction

More than one-third of people living with HIV experience depressive symptoms, which present a major health concern and are associated with other negative health outcomes such as poor adherence to antiretroviral therapy and increased morbidity and mortality [1-5]. In addition, people living with HIV with depressive symptoms are more likely to engage in HIV-related risk behaviors such as inconsistent use of condoms and substance abuse, which increases the risk of HIV infection and transmission [1,6,7]. As depression is often chronic and requires long-term follow-up, effective interventions to address this common comorbidity in people living with HIV with long-term effects are urgently needed [7].

However, mental health resources are scarce, especially in low-income and middle-income countries, and <17% of people living in these countries have received the needed mental health treatment [8,9]. For example, there is 0.5 psychiatrist per million people in Africa, which has the highest burden of HIV, and 22 psychiatrists per million people in China, compared with 83 psychiatrists per million people in Europe, 105 psychiatrists per million people in the United States, and 147 psychiatrists per million people in Canada [10,11]. Therefore, interventions to provide effective treatments for depression and facilitate the implementation of such treatments for people living with HIV living in low-income and middle-income countries are urgently needed.

A growing number of mobile health (mHealth) interventions have been developed to improve mental health outcomes, especially depressive symptoms, in people living with HIV. However, the existing literature has several limitations. First, few studies have explored the long-term effects of mHealth interventions [12-14]. A recent systematic review and meta-analysis included 14 randomized controlled trials (RCTs) of mHealth interventions for reducing depressive symptoms among people living with HIV [12]. Among these, only one study explored the long-term intervention effects for up to 1 year, and the results did not show any improvement in depressive symptoms [15]. Second, most mHealth studies have small sample sizes <100; only a few studies have sample sizes >200 [15-18]. Third, existing mHealth interventions among people living with HIV have largely been conducted in high-income countries such as the United States and Western Europe, with only a few studies conducted in low-income or middle-income countries [15,16,19-21]. Finally, existing mHealth interventions are mostly telephone-based or simply send SMS text messages, whereas few studies have used web applications or mobile apps [12,15,16,19,22-24].

Our mHealth intervention, Run4Love, is a 3-month internet-based multimedia program delivered via a popular social media app called WeChat. Run4Love was effective in reducing depressive symptoms in people living with HIV at 3-, 6-, and 9-month follow-ups in a 2-arm parallel RCT with a sample size of 300 [16]. After the trial, we continued to follow all the participants for 3 years after the baseline. In this study, we aimed to evaluate the long-term effects of the Run4Love intervention at 1-year and 3-year follow-ups. We hypothesized that the Run4Love intervention had long-term intervention effects in reducing depressive symptoms at 1-year and 3-year follow-ups and that it also had long-term effects on secondary outcomes such as the patients' quality of life (QOL), positive coping, and perceived stress.

Methods

Overview

Participants were recruited from the outpatient department of Guangzhou Eighth People's Hospital, which is the largest infectious disease hospital in Guangzhou, the capital city of Guangdong Province in South China, from September 2017 to January 2018. Participants were randomized to the Run4Love intervention group or the control group in 1:1 ratio.

Ethics Approval

The Run4Love intervention was registered in the Chinese Clinical Trial Registry (ChiCTR-IPR-17012606). The study design is described in the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist in Multimedia Appendix 1. The study protocol was approved by the institutional review board of the School of Public Health at the Sun Yat-sen University (approval number 2015-28).

Participants

Participants were screened for depressive symptoms, and those with depressive symptoms were invited to participate in the study. A total of 300 participants were recruited for the study. Inclusion criteria were the following: (1) aged ≥18 years, (2) being HIV seropositive, (3) having depressive symptoms (Center for Epidemiological Studies–Depression [CES-D] scale score ≥16), (4) willing to provide hair samples, and (5) using WeChat. Exclusion criteria were the following: (1) unable to complete the screening or baseline questionnaire, (2) having trouble in reading or listening to the intervention materials, (3) currently on treatment for depressive symptoms, and (4) unable to participate in physical activities. All the recruited participants signed the informed consent form. Study participants were given up to 10 Yuan (approximately US $1.6) as weekly incentive for

XSL•FO
RenderX

their completion of the Run4Love program; they were also provided 20 to 50 Yuan (US $3.1-US $7.8) for completion of the follow-up questionnaires.

Procedure

Run4Love RCT

After completing the baseline assessment, eligible patients were randomly assigned to the Run4Love intervention group or the wait-list control group. Randomization was performed using SAS software (version 9.4) to generate a randomization list with a block size of 4. Owing to the design of the study, neither the participants nor the researchers were blinded. The Run4Love intervention protocol has been described elsewhere [25]. Briefly, the 3-month Run4Love intervention program was tailored to people living with HIV with depressive symptoms. The intervention included adapted cognitive behavioral stress management (CBSM) courses and physical activity promotion. The adapted CBSM courses consisted of 12 weekly sessions on stress management and coping skills. Physical activity promotion provided guidance and suggestions for regular exercise and healthy diet choices.

Personalized feedback based on completion of the CBSM courses was sent to each participant on a weekly basis via a novel, enhanced WeChat platform developed by the investigators. Corresponding financial incentives were also sent based on their completion status on a weekly basis. In addition, participants in the intervention group received 5 phone calls from the research staff at 1 week and 1, 2, 5, and 8 months after enrollment. The purposes of these phone calls were to confirm, facilitate, and sustain the participation of the patients.

Participants in the wait-list control group received a paper-based brochure on nutrition for people living with HIV at baseline, and they received the Run4Love program at 9 months from baseline, delivered in the same way as in the intervention group for 3 months, with the same materials and frequencies, but without any phone calls and incentives.

Data Collection Procedure

Participants in both groups were invited to complete in-person assessments at 3, 6, and 9 months and web-based assessments at 1 and 3 years after enrollment via the enhanced WeChat platform. Measurements in the follow-up questionnaires were the same across time and took approximately 15 to 25 minutes to complete. Except for the 9-month assessment, participants received a compensation of 20 Yuan (approximately US $3) for completing the survey. Reminders were sent and a follow-up call was made to those who were not responsive for a week.

Measures

Primary Outcome

The primary outcome was reduction in depressive symptoms, which was measured by the Chinese version of the 20-item CES-D scale at baseline and at 3-, 6-, 9-, 12-, and 36-month follow-ups. Changes in CES-D score at each time point from that at baseline were calculated [26,27]. The CES-D scale is a validated measurement that is used among various populations, including Chinese people living with HIV with depressive symptoms [27-29]. The internal consistency of the scale was satisfactory, with most Cronbach α being >.80 when measured at all time points. Items of the scale were assessed on a 4-point Likert scale for the frequency of depressive behaviors or feelings in the past week, such as “I was bothered by things that usually don’t bother me.” The total score ranges from 0 to 60, with score ≥16 indicating depressive symptoms.

Secondary Outcomes

Secondary outcomes included QOL, perceived stress, positive and negative coping, self-efficacy, HIV-related stigma, and depression severity.

QOL was assessed using the 31-item World Health Organization Quality of Life HIV short version (WHOQOL-HIV BREF) [30,31]. Items were assessed on a 5-point Likert scale for QOL in the past 2 weeks, such as “How much do you enjoy life?” with total score ranging from 24 to 120. Perceived stress was measured using the 10-item Perceived Stress Scale (PSS), where the total score ranges between 0 and 40, with high score indicating more stress in the past month [32]. Coping was measured using the Simplified Ways of Coping Questionnaire (SWCQ), which includes a 12-item positive coping subscale (score range 0-36) and an 8-item negative coping subscale (score range 0-24) [33]. Self-efficacy was assessed using the 10-item Chinese version of the General Self-Efficacy Scale. The total score ranges between 10 and 40, with high score indicating better self-efficacy [34]. HIV-related stigma was measured using the HIV Stigma Scale, which includes 7 items on a 4-point Likert scale, such as “I feel guilty because I have HIV.” The total score ranges from 14 to 56, with high scores indicating high levels of HIV-related stigma [35]. The 9-item Patient Health Questionnaire (PHQ-9) is a widely used criteria-based diagnostic tool for depressive symptoms. It was used to assess depression severity over the past 2 weeks using a 3-point Likert scale ranging from *not at all* to *nearly every day*. The total score ranges from 0 to 27, with score ≥5 being considered as having depressive symptoms [36]. We used metabolic equivalents, which were calculated using the Chinese version of the Global Physical Activity Questionnaire to describe the intensity of physical activities of the participants [37].

Statistical Analysis

The intention-to-treat principle was applied to all the analyses. Descriptive statistics for baseline characteristics and psychological outcomes (eg, CES-D, QOL, and PSS scores) were provided. Continuous variables were described as means and SDs for normally distributed variables or medians and IQRs for nonnormally distributed variables and frequencies and percentages for categorical variables. Between-group differences at baseline were reported using the independent samples 2-tailed *t* test, chi-square (χ^2) test, or nonparametric test, as appropriate, and the 95% CIs were calculated. Similar analyses were performed for group differences between participants who completed the 1-year or 3-year follow-up and those who dropped out at the 1-year or 3-year follow-up. For missing values, multivariate imputation by chained equations was performed using R package mice (version 4.0.5; R Foundation for Statistical Computing), and 80 imputed data sets were obtained. The final data set was the average of the 80 imputed data sets.

XSL•FO
RenderX

All the statistical analyses were conducted using the complete data set after imputation. In addition, sensitivity analysis was conducted using data with missing values to assess the robustness of the results.

For primary and secondary outcome analyses, the repeated measures generalized estimating equation (GEE) linear regression models were used to assess the intervention effects [38]. As a model that is well suited for longitudinal data analysis, GEE improves statistical power because it allows for the simultaneous analysis of intervention effects at multiple time points in a single model, with an exchangeable working correlation matrix accounting for potential correlation owing to within-participant dependencies across time. In this study, the main effects of group and time and the interaction effects between group and time were examined using the GEE, using repeated measurements of the 2 groups at baseline and the 5 follow-up points (ie, 3-, 6-, 9-, 12-, and 36-month follow-ups), adjusting for baseline demographic and HIV-related characteristics (ie, age, sex, BMI, education, sexual orientation, family monthly income, marital status, duration of HIV infection, and employment) and psychological outcomes (eg, CES-D scores). The R package geepack (version 4.0.5) was used to conduct the GEE analysis.

Cohen *d* was calculated to measure the effect size of the intervention [39]. The between-group effect size was calculated using the difference between the mean score change of the intervention group from baseline and that of the control group from baseline, which was then divided by the SD of the pooled score changes. Cohen $d > 0.20$ was considered as a small effect size, Cohen $d > 0.50$ as a medium effect size, and Cohen $d > 0.80$ as a large effect size [40].

All analyses were performed using R (version 4.0.5; R Foundation for Statistical Computing). All statistical tests were 2-sided, and $P < .05$ was considered as statistically significant.

Results

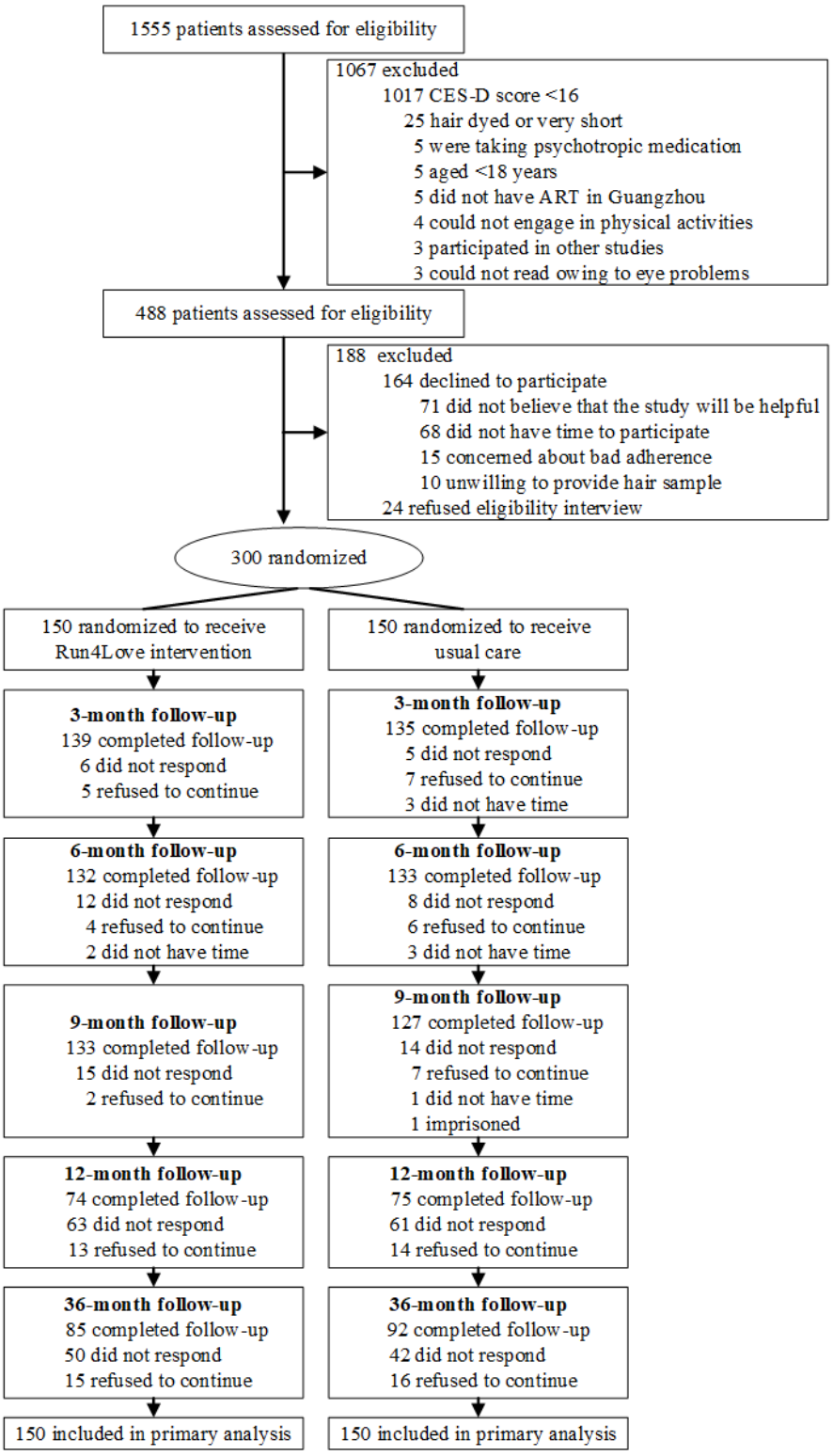
Sample Characteristics

Among the 300 participants, approximately half (n=149, 49.7%) completed the 1-year follow-up evaluation, including 49.7% (74/149) of them in the Run4Love intervention group and 50.3% (75/149) of them in the control group. A total of 59% (177/300) of the participants completed the 3-year follow-up evaluation, including 48% (85/177) of them in the intervention group and 51.9% (92/177) of them in the control group (Figure 1). The baseline characteristics of the participants who completed the baseline, 1-year, and 3-year follow-up surveys are shown in Table 1. All demographic characteristics were balanced between the 2 groups at baseline, except sexual orientation, sex, and family monthly income. The mean completion rate of the CBSM courses was 55% in the Run4Love intervention group and 4% in the wait-list control group.

Similarly, the baseline characteristics were compared between participants who completed the assessments and those who were lost to follow-up. There were significant differences in BMI, duration of HIV infection, and self-efficacy at baseline between participants who completed the 1-year follow-up questionnaire and those who were lost to follow-up. Significant differences were also found in age, BMI, and duration of HIV infection at baseline between participants who completed the 3-year follow-up questionnaire and those who were lost to follow-up (Tables S1 and S2 in Multimedia Appendix 2). Baseline characteristics of participants who did not complete the 1-year or 3-year outcome evaluation between the 2 groups were comparable (Tables S3 and S4 in Multimedia Appendix 2).

XSL•FO
RenderX

Figure 1. Flow chart of the Run4Love trial and the long-term follow-ups. ART: antiretroviral therapy; CES-D: Center for Epidemiological Studies–Depression.



XSL•FO
RenderX

Table 1. Baseline characteristics of participants in the intervention and control groups.

Baseline characteristics	Baseline		1-year follow-up		3-year follow-up	
	Run4Love group (n=150)	Usual care group (n=150)	Run4Love group (n=74)	Usual care group (n=75)	Run4Love group (n=85)	Usual care group (n=92)
Age (years), mean (SD)	28 (5.8)	28.6 (5.9)	28.1 (5.5)	27.9 (6.1)	28.1 (5.3)	27.4 (5.5)
Sex (male), n (%)	142 (94.7)	135 (90)	73 (99)	67 (89)	82 (96)	84 (91)
BMI[a], mean (SD)	20.5 (2.5)	20.1 (2.4)	19.5 (1.7)	19.3 (1.5)	19.5 (1.9)	19.7 (2.3)
Education level greater than high school, n (%)	98 (65.3)	84 (56)	51 (69)	44 (59)	59 (69)	52 (57)
Homosexual, bisexual, or uncertain, n (%)	130 (86.7)	115 (76.7)	67 (91)	60 (80)	75 (88)	73 (79)
Married, n (%)	18 (12)	20 (13.3)	6 (8)	7 (9)	8 (9)	10 (11)
Family monthly income 7000 Yuan (US $1043.7), n (%)	68 (45.3)	56 (37.3)	32 (43)	24 (32)	44 (52)	27 (29)
Duration of HIV infection, median (IQR)	1.7 (0.6-4)	1.7 (0.6-3.6)	1.2 (0.5-1.7)	1.2 (0.9-2.3)	1.2 (0.7-1.7)	1.2 (0.4-2)
CES-D[b] score, mean (SD)	23.9 (6.4)	24.3 (6.9)	23.8 (5.9)	24.7 (7.7)	24.5 (6.2)	24.5 (7)
PHQ-9[c] score, mean (SD)	10.2 (4.5)	10.7 (5.1)	10 (4.1)	10.8 (5.7)	10.5 (4.7)	11 (5.6)
WHOQOL-HIV BREF[d] score, mean (SD)	77.4 (9)	76.6 (9.4)	77.5 (8.3)	75.3 (10.4)	76.9 (8.6)	76.3 (10.2)
GSES[e] score, mean (SD)	24.4 (5.2)	23.3 (5.6)	23.6 (5.2)	22.3 (5.8)	24 (4.9)	23 (5.9)
PSS[f] score, mean (SD)	20 (4.4)	20.7 (4.4)	19.8 (4.4)	21.4 (4.8)	20.1 (4.6)	21.1 (4.8)
HIV Stigma Scale score, mean (SD)	37.1 (7.7)	38 (7.5)	36.5 (7.7)	37.5 (9)	36.7 (8)	38.1 (8.5)
SWCQ positive coping[g] score, mean (SD)	18.4 (5.5)	18.3 (6.2)	18.1 (4.7)	18.1 (6.3)	18.6 (5.2)	18.7 (6.1)
SWCQ negative coping[h] score, mean (SD)	11.8 (3.9)	11.8 (3.9)	11.3 (3.9)	12.1 (4.1)	11.5 (4)	11.9 (4)
Physical activity (metabolic equivalents[i] ≥600), n (%)	65 (43.3)	65 (43.3)	43 (58)	44 (59)	46 (54)	49 (53)

[a]Calculated as weight in kilograms divided by height in meters squared.

[b]CES-D: Center for Epidemiological Studies–Depression.

[c]PHQ-9: 9-item Patient Health Questionnaire.

[d]WHOQOL-HIV BREF: World Health Organization Quality of Life HIV short version.

[e]GSES: General Self-Efficacy Scale.

[f]PSS: Perceived Stress Scale.

[g]SWCQ positive coping: Simplified Ways of Coping Questionnaire positive coping domain.

[h]SWCQ negative coping: SWCQ negative coping domain.

[i]Physical activity was measured by metabolic equivalents. Metabolic equivalents ≥600 indicates recommended physical activity level.

Primary Outcome at 1-Year Follow-up

As previously reported, significant reductions in the primary outcome (depressive symptoms) were observed among participants in the intervention group at the 3-, 6-, and 9-month follow-ups [16]. The sustained effects of reduced depressive symptoms were further observed at 1-year follow-up. At the 1-year follow-up, the Run4Love intervention group showed significant reduction in the CES-D scores compared with that shown by the control group (from 23.9 to 18.1 in the intervention group vs from 24.3 to 23.3 in the control group; mean difference between groups −4.79, SD 13.56; 95% CI −7.78 to −1.81; *P*=.002), with standard effect size (Cohen *d*) of 0.48 in favor of the Run4Love intervention (Table 2).

XSL•FO
RenderX

Table 2. Long-term effects of the Run4Love intervention on primary and secondary outcomes.

Outcomes and time points	Run4Love intervention group (n=150)		Usual care group (n=150)		Between-group difference for mean change from baseline, mean difference (95% CI)	*P* value[a]
	Follow-up, mean (SD)	Within-group changes, mean difference (95% CI)[b]	Follow-up, mean (SD)	Within-group changes, mean difference (95% CI)		
Depressive symptoms (Center for Epidemiological Studies–Depression scale[c])						
Baseline	23.9 (6.4)	N/A[d]	24.3 (6.9)	N/A	N/A	N/A
1-year follow-up	18.1 (11.3)	−5.79 (−7.99 to −3.59)	23.3 (12.4)	−0.99 (−3.27 to 1.29)	−4.79 (−7.78 to −1.81)	.002
3-year follow-up	20.5 (11.2)	−3.47 (−5.64 to −1.31)	24.4 (12.7)	0.15 (−2.07 to 2.38)	−3.63 (−6.71 to −0.54)	.02
Quality of life[e]						
Baseline	77.4 (9)	N/A	76.6 (9.4)	N/A	N/A	N/A
1-year follow-up	82.8 (14.1)	5.42 (2.88 to 7.95)	77.9 (14.7)	1.26 (−1.08 to 3.6)	4.15 (0.81 to 7.50)	.02
3-year follow-up	78.6 (13.7)	1.13 (−1.23 to 3.48)	74.1 (14.3)	−2.50 (−4.96 to −0.05)	3.63 (0.34 to 6.92)	.03
Perceived stress (Perceived Stress Scale[c])						
Baseline	20 (4.4)	N/A	20.7 (4.4)	N/A	N/A	N/A
1-year follow-up	16.6 (6.1)	−3.38 (−4.70 to −2.06)	19.3 (6.6)	−1.41 (−2.74 to −0.08)	−1.97 (−3.67 to −0.26)	.02
3-year follow-up	18 (5.9)	−1.93 (−3.04 to −0.81)	19.9 (6.6)	−0.80 (−1.99 to 0.38)	−1.12 (−2.73 to 0.48)	.17
Simplified Ways of Coping Questionnaire positive coping[e]						
Baseline	18.4 (5.5)	N/A	18.3 (6.2)	N/A	N/A	N/A
1-year follow-up	21 (7.6)	2.59 (0.95 to 4.23)	18.2 (7.1)	−0.16 (−1.59 to 1.27)	2.75 (0.76 to 4.74)	.007
3-year follow-up	19.2 (6.9)	0.78 (−0.84 to 2.4)	17.8 (6.4)	−0.53 (−1.84 to 0.79)	1.31 (−0.72 to 3.34)	.21
Simplified Ways of Coping Questionnaire negative coping[c]						
Baseline	11.8 (3.8)	N/A	11.7 (3.9)	N/A	N/A	N/A
1-year follow-up	11.9 (4.3)	0.11 (−0.81 to 1.03)	11.9 (4.6)	0.19 (−0.76 to 1.15)	−0.08 (−1.38 to 1.21)	.90
3-year follow-up	12.1 (4)	0.29 (−0.67 to 1.25)	11.6 (4)	−0.11 (−1.01 to 0.79)	0.40 (−0.91 to 1.70)	.55
Self-efficacy (General Self-Efficacy Scale[e])						
Baseline	24.4 (5.2)	N/A	23.3 (5.6)	N/A	N/A	N/A
1-year follow-up	26.6 (6.7)	2.22 (0.82 to 3.62)	23.5 (6.4)	0.20 (−1.12 to 1.51)	2.02 (0.19 to 3.85)	.03
3-year follow-up	25.1 (6.4)	0.73 (−0.52 to 1.98)	22.9 (6.5)	−0.42 (−1.72 to 0.87)	1.15 (−0.61 to 2.92)	.19
HIV Stigma Scale[c]						
Baseline	37.1 (7.7)	N/A	38 (7.5)	N/A	N/A	N/A
1-year follow-up	33.5 (9.6)	−3.62 (−5.43 to −1.82)	36.9 (10.4)	−1.10 (−2.95 to 0.76)	−2.53 (−5.17 to 0.09)	.06
3-year follow-up	35.4 (9.9)	−1.70 (−3.43 to 0.03)	37.6 (10.1)	−0.43 (−2.14 to 1.28)	−1.27 (−3.67 to 1.12)	.30
Depression severity (9-item Patient Health Questionnaire[c])						
Baseline	10.2 (4.5)	N/A	10.7 (5.1)	N/A	N/A	N/A
1-year follow-up	7.2 (5.1)	−2.98 (−4.09 to −1.87)	8.5 (5.5)	−2.23 (−3.30 to −1.15)	−0.75 (−2.21 to 0.71)	.31
3-year follow-up	8.4 (5.4)	−1.75 (−2.87 to −0.63)	8.7 (5.9)	−2.06 (−3.16 to −0.97)	0.31 (−1.24 to 1.87)	.69
Physical activity (metabolic equivalents[e])						
Baseline	3225 (6189)	N/A	2675 (6064)	N/A	N/A	N/A
1-year follow-up	4193 (20,017)	969 (−3414 to 5352)	6850 (37,316)	4176 (−2041 to 10,393)	−3206 (−10,734 to 4320)	.40
3-year follow-up	2788 (5470)	−436 (−1814 to 941)	2933 (6860)	258 (−1142 to 1660)	−695 (−2674 to 1283)	.49

XSL•FO
RenderX

[a]Results based on the generalized estimating equations, adjusted for age, sex, BMI, education, sexual orientation, family monthly income, marital status, duration of HIV infection, and employment.

[b]Within-group changes are mean changes.

[c]High score indicates worse outcome.

[d]N/A: not applicable.

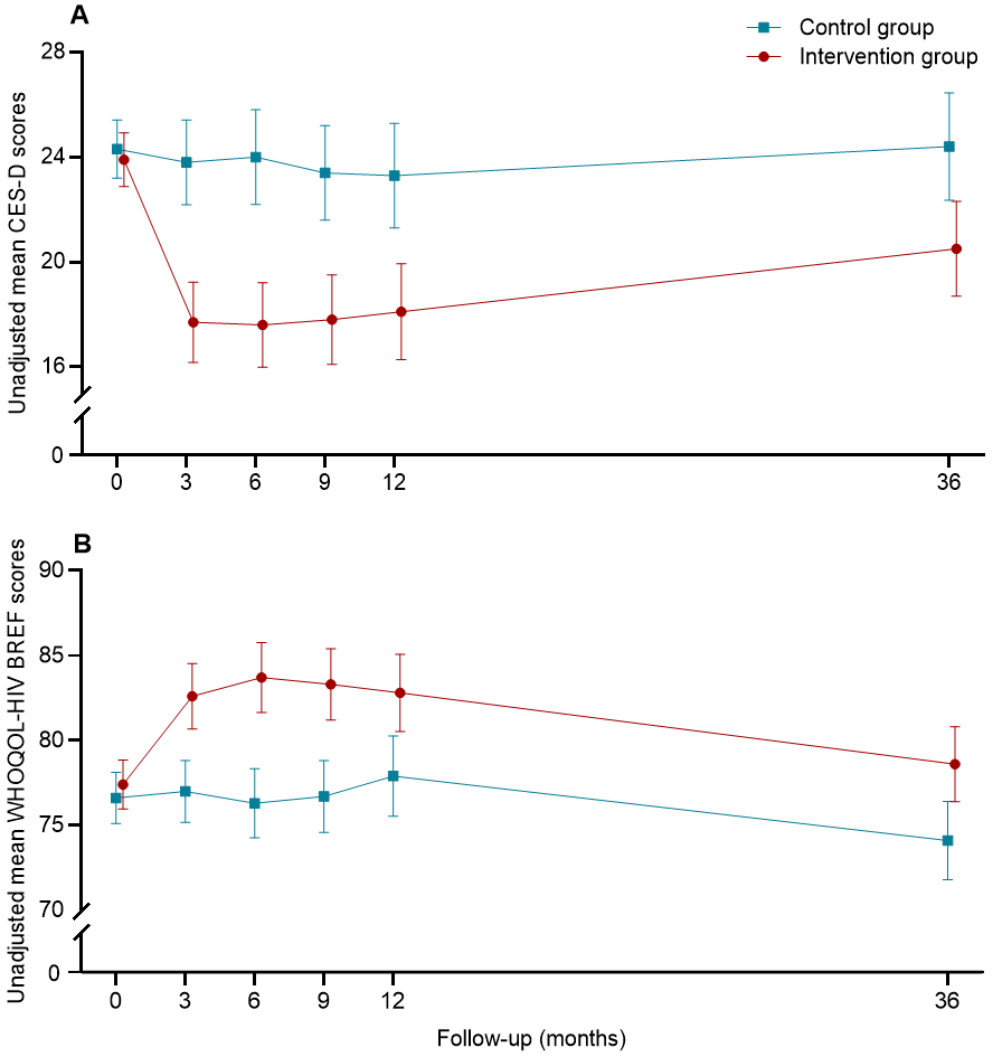
[e]High score indicates better outcome.

Primary Outcome at 3-Year Follow-up

At the 3-year follow-up, between-group difference in the CES-D scores remained statistically significant (mean difference between groups −3.63, SD 13.35; 95% CI −6.71 to −0.54; *P*=.02; Cohen *d*=0.36). Results from the GEE model on the long-term intervention effects on depressive symptoms at the 1-year and 3-year follow-ups are summarized in Table 2. Results indicated that there was no main effect of group or time, but there were significant interaction effects between group and time, with statistically significant between-group differences in the CES-D score for mean changes from baseline, after controlling for baseline characteristics. The results were similar to those obtained from data without multiple imputations for missing values (Table S5 in Multimedia Appendix 2). Significant between-group differences were observed at both 1-year and 3-year follow-up. Results from all the measurement time points (ie, 3, 6, and 9 months and 1 and 3 years), GEE model, and sensitivity analyses showed similar findings (Tables S6-S8 in Multimedia Appendix 2).

Change patterns of depressive symptoms (CES-D) at baseline and follow-ups in the 2 groups are presented in Figure 2. During the 3-year follow-up period, the average CES-D scores of the control group remained fairly stable. By contrast, the average CES-D scores of the intervention group decreased significantly after the Run4Love intervention at 3 months. After the largest decline in depressive symptoms at 3 months, the CES-D scores in the intervention group gradually increased, but never reached as high as they were at baseline, and the between-group differences remained statistically significant throughout the 3-year follow-up.

Figure 2. Measures of depressive symptoms (Center for Epidemiological Studies–Depression [CES-D] scale) and quality of life (World Health Organization Quality of Life HIV short version [WHOQOL-HIV BREF]) at baseline and follow-ups in the Run4Love intervention and control groups.

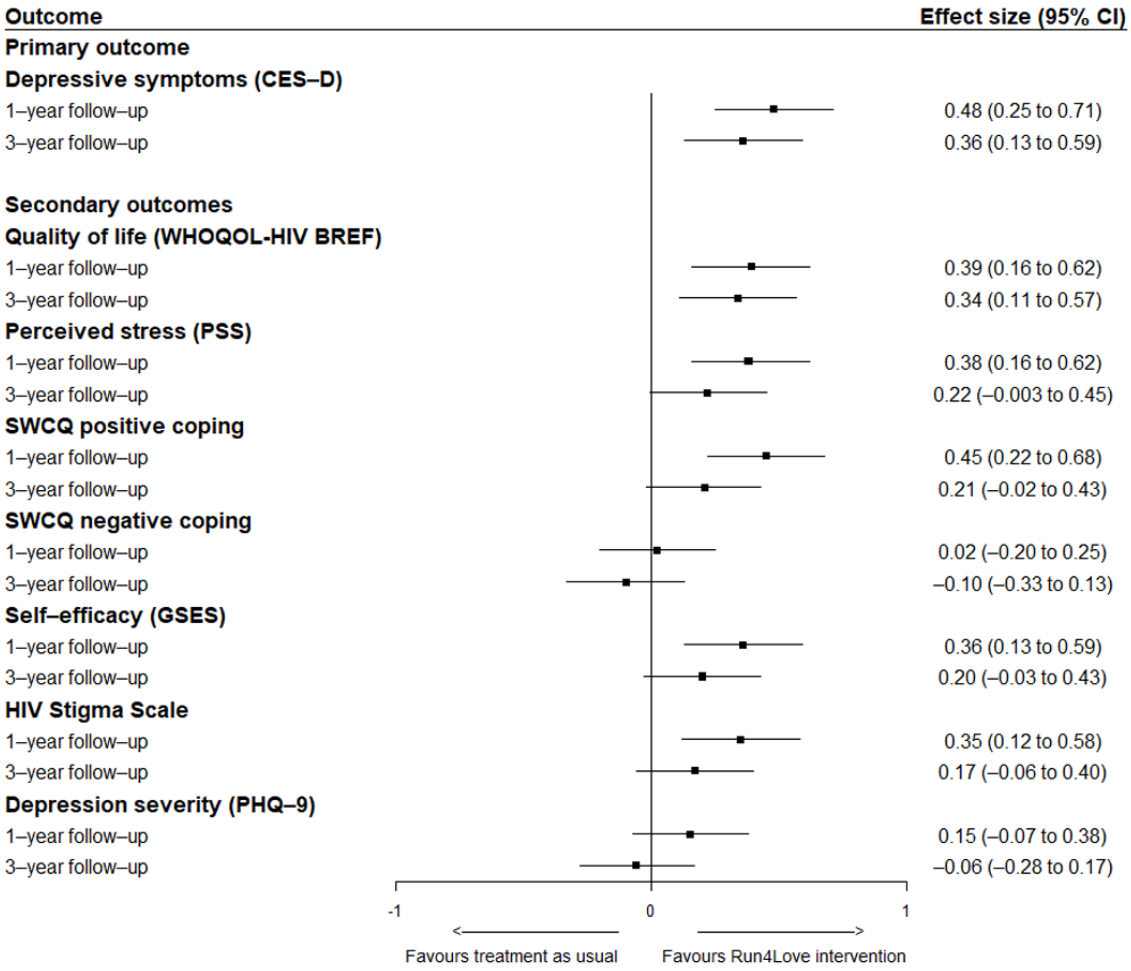


XSL•FO
RenderX

Secondary Outcomes at 1-Year Follow-up

Results of the long-term intervention effects on secondary outcomes are presented in Table 2. At the 1-year follow-up, compared with the control group, participants in the Run4Love intervention group reported significantly improved QOL (WHOQOL-HIV BREF: from 77.4 to 82.8 in the intervention group vs 76.6 to 77.9 in the control group; mean difference between groups 4.15, SD 15.61; 95% CI 0.81-7.50; *P*=.02), with standard effect size of 0.39 in favor of the Run4Love intervention group (Table 2; Figure 3). Similarly, participants in the intervention group also reported significantly reduced perceived stress (PSS: from 20 to 16.6 in the intervention group vs from 20.7 to 19.3 in the control group; mean difference between groups –1.97, SD 8.10; 95% CI −3.67 to −0.26; *P*=.02) and improved SWCQ positive coping (from 18.4 to 21 in the intervention group vs from 18.3 to 17.8 in the control group; mean difference between groups 2.75, SD 10.10; 95% CI 0.76-4.74; *P*=.007) and self-efficacy (General Self-Efficacy Scale: from 24.4 to 26.6 in the intervention group vs from 23.3 to 23.4 in the control group; mean difference between groups 2.02, SD 8.63; 95% CI 0.19-3.85; *P*=.03). There were no significant between-group differences in changes in SWCQ negative coping, HIV-related stigma (HIV Stigma Scale), depression severity (PHQ-9), and physical activity (metabolic equivalents) at 1 year. The results were also similar to those obtained from data without multiple imputations for missing values (Table S5 in Multimedia Appendix 2).

Figure 3. Plot showing the Cohen d effect sizes and 95% CIs for the primary and secondary outcomes of the Run4Love intervention group versus the control group. CES-D: Center for Epidemiological Studies–Depression; GSES: General Self-Efficacy Scale; PHQ-9: 9-item Patient Health Questionnaire; PSS: Perceived Stress Scale; SWCQ: Simplified Ways of Coping Questionnaire; WHOQOL-HIV BREF: World Health Organization Quality of Life HIV short version



Secondary Outcomes at 3-Year Follow-up

At the 3-year follow-up, the intervention effect remained statistically significant for QOL (WHOQOL-HIV BREF: from 77.4 to 78.6 in the intervention group vs 76.6 to 74.1 in the control group; mean difference between groups 3.63, SD 14.53; 95% CI 0.34-6.92; *P*=.03; Cohen *d*=0.34). There were no significant between-group differences in the changes in other measures of the secondary outcomes. The results were similar to those obtained from data without multiple imputations for missing values (Table S5 in Multimedia Appendix 2).

Regarding the change patterns of QOL, in the first 12 months, the average QOL scores in the control group remained moderately stable and declined slightly at the 3-year follow-up. By contrast, significant increase in the average QOL scores was observed in the intervention group at 3 months, immediately upon completion of the Run4Love program, and the score continued to increase, with the largest group difference occurring

XSL•FO
RenderX

at the 6-month follow-up. Although the QOL scores in the intervention group gradually and slowly decreased over time, significant between-group differences were observed at 9-month, 1-year, and 3-year follow-ups (Figure 2). Change patterns of other secondary outcomes are summarized in Figure S1 in Multimedia Appendix 2.

Discussion

Principal Findings

This study is among the first efforts to investigate the long-term effects of an mHealth intervention on depressive symptoms among people living with HIV, with a follow-up period of 3 years. Results showed that, compared with the control, the Run4Love intervention significantly reduced depressive symptoms (CES-D) and improved QOL not only at 3, 6, and 9 months but also at 1-year and 3-year follow-ups [16]. In addition, changes in secondary outcomes including perceived stress, positive coping, and self-efficacy remained significant at the 1-year follow-up, but the between-group differences in these measures were not significant at the 3-year follow-up.

Results Interpretation and Implication

This study addresses several gaps in the literature regarding long-term effects of mHealth interventions on depressive symptoms in people living with HIV. Few studies have investigated the long-term effects (>1 year) of mHealth interventions on depressive symptoms among people living with HIV [12-15,41-45]. A study in the Netherlands found that a 2-month internet-based intervention using cognitive behavioral therapy effectively reduced depressive symptoms among people living with HIV [24]. However, the Netherlands-based study followed the intervention group for 8 months from baseline and the control group for only 5 months. The study found sustained intervention effect on depressive symptoms over time. The only study that we found on mHealth interventions targeting depressive symptoms among people living with HIV, with a follow-up period of ≥1 year was conducted by Li et al [45]. They explored the impact of 1-month positive psychology and social networking intervention on depressive symptoms of HIV-infected men who have sex with men in Chengdu, China, but the results showed no significant improvement in depressive symptoms at the 13-month follow-up.

Compared with mHealth interventions, in-person interventions have shown more sustained long-term effects on depressive symptoms among people living with HIV [46,47]. A total of 2 in-person intervention studies reported significant intervention effects at 1-year follow-up. Steven et al [46] showed that an 11-week cognitive behavioral therapy effectively reduced depressive symptoms among people living with HIV, and the effects were sustained for up to 12 months. Another study of 8 weekly sessions, which lasted 2 to 3 hours each, based on group support psychotherapy also achieved significant improvement in depressive symptoms among people living with HIV [47]. Although effective, these in-person psychotherapy programs are very expensive and difficult to scale up, especially in resource-poor settings. Furthermore, few of these in-person interventions have reported long-term effects for up to 3 years.

From previous qualitative interviews, secondary data analyses, and researchers' experience, we offer some possible explanations for the significant long-term intervention effects found in this study: theory-guided intervention design, combination of web-based and offline interactions, continuous process monitoring, and changed perception and behaviors among the participants. First, the Run4Love program was adapted from evidence-based CBSM courses that are effective in reducing depressive symptoms among people living with HIV [46,48]. Second, the Run4Love intervention combined both offline interactions, between researchers and participants at baseline and 3, 6, and 9 months, and web-based interactions, such as reminders for course completion and 5 phone calls at 1 week and 1, 2, 5, and 8 months from baseline. A trusting relationship between participants and researchers was built at the beginning of the program and maintained over time, which was critical for the success of the intervention [49]. Third, continuous process monitoring ensured a satisfactory level of patient engagement, which is essential for the effectiveness of an mHealth intervention [50]. On average, participants in the intervention group completed 55% of the CBSM coursework, which was comparable with that in other mHealth interventions, whereas those in the wait-list control group completed only 4% of the coursework, as they received no regular monitoring and phone calls [51,52]. Finally, through feedback at the 3-year follow-up, participants shared their personal experience of how their perception of stress changed over time and how they adapted their coping behaviors as a result of the cognitive changes during and after the Run4Love intervention. They also commented that specific stress-reduction skills such as exercise, relaxation, and meditation were very helpful in managing depressive symptoms.

Notably, participants in the intervention group showed significant reduction in depressive symptoms measured by CES-D scores, which was sustained at the 3-year follow-up; however, the between-group differences in PHQ-9 scores were only significant at the 3-month and 6-month follow-ups and were not significant at later follow-ups. This discrepancy may be because of the differences between these 2 measurements. The PHQ-9 is more widely used for depression screening, whereas the CES-D measure has high sensitivity, specificity, and responsiveness in detecting and monitoring changes in depressive symptoms in longitudinal studies compared with the PHQ-9 [53,54]. In addition, the positive effects of physical activity on mental health improvement, including reducing depressive symptoms, are well established in the literature [55,56]; however, we did not observe significant improvement in physical activity in our RCT. It may be possible that the intervention was not effective in improving participants' physical activity or that we did not capture changes in physical activity appropriately and effectively, as physical activity was self-reported. Although in our qualitative interviews at the 3-year follow-up, some participants mentioned that they exercised regularly because of our intervention, the effects of physical activity promotion in our intervention are unclear.

It is imperative and beneficial to promote mHealth interventions with long-term effects on depressive symptoms among people living with HIV, especially during the COVID-19 pandemic,

XSL•FO
RenderX

which not only exacerbates distress in vulnerable populations but also limits their access to health services. mHealth interventions offer a promising alternative to in-person care in the delivery of urgently needed mental health services in underserved communities or areas with shortage of mental health professionals, such as Africa or other low-income and middle-income countries or regions. Furthermore, as depressive symptoms are a common comorbidity with many chronic diseases including HIV infection, evidence-based mHealth interventions such as Run4Love may be adapted and integrated into routine care or depressive symptom management.

Limitations

This study has several limitations. First, the dropout rates at the 1-year and 3-year follow-ups were relatively high, which is a common challenge in long-term follow-ups in interventions, especially in mHealth interventions [57,58]. However, we managed to trace more participants at 3-year follow-up than 1-year follow-up. To examine the robustness of the results at the 1-year and 3-year follow-ups, we used sensitivity analyses and found similar and robust results using data with missing values and data without missing values after multiple imputations. Second, all the participants were recruited from a large hospital in a capital city in South China, and most of them were men, especially young men; therefore, the results from this study may not be generalizable to people living with HIV living in rural areas or women who live with HIV.

Conclusions

In conclusion, this study found that a social media–based mHealth intervention, Run4Love, significantly reduced depressive symptoms and improved QOL among people living with HIV at 1-year and 3-year follow-ups. Further research is needed to explore the mechanisms underlying the long-term effects of mHealth interventions and understand how to implement evidence-based interventions with sustained effects to better serve people living with HIV with depression.

Acknowledgments

This study was supported by grants from the National Natural Science Foundation of China (grant 71573290) and the China Medical Board (grant 17-271). The authors thank all the patients who participated in the study. The authors also thank all the nurses and physicians who recruited participants for the study.

Authors' Contributions

YG, CY, and YL had full access to all the data in the study and took responsibility for the integrity of the data and accuracy of the data analysis. YG and YAH conceptualized and designed the study. YG, CY, YL, and YZ contributed to the acquisition, analysis, and interpretation of the data. YG, CY, YL, and XW drafted the manuscript. AL, YAH, and AMW critically revised the manuscript. CY and YL performed the statistical analysis. YG obtained the funding for this study. LL and WC provided the administrative and technical support. YG and AL supervised the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1265 KB - jmir_v24i6e36809_app1.pdf]

Multimedia Appendix 2
Sensitivity analyses of results on primary and secondary outcomes.
[DOCX File , 756 KB - jmir_v24i6e36809_app2.docx]

References

1. Tao J, Vermund SH, Qian HZ. Association between depression and antiretroviral therapy use among people living with HIV: a meta-analysis. AIDS Behav 2018 May;22(5):1542-1550 [FREE Full text] [doi: 10.1007/s10461-017-1776-8] [Medline: 28439754]
2. Lofgren SM, Bond DJ, Nakasujja N, Boulware DR. Burden of depression in outpatient HIV-infected adults in Sub-Saharan Africa; systematic review and meta-analysis. AIDS Behav 2020 Jun;24(6):1752-1764 [FREE Full text] [doi: 10.1007/s10461-019-02706-2] [Medline: 31720956]
3. Gokhale RH, Weiser J, Sullivan PS, Luo Q, Shu F, Bradley H. Depression prevalence, antidepressant treatment status, and association with sustained HIV viral suppression among adults living with HIV in care in the United States, 2009-2014. AIDS Behav 2019 Dec;23(12):3452-3459. [doi: 10.1007/s10461-019-02613-6] [Medline: 31367965]
4. Galvan FH, Burnam MA, Bing EG. Co-occurring psychiatric symptoms and drug dependence or heavy drinking among HIV-positive people. J Psychoactive Drugs 2003 May;35 Suppl 1:153-160. [doi: 10.1080/02791072.2003.10400510] [Medline: 12825758]

XSL•FO
RenderX

5. Walkup J, Blank MB, Gonzalez JS, Safren S, Schwartz R, Brown L, et al. The impact of mental health and substance abuse factors on HIV prevention and treatment. J Acquir Immune Defic Syndr 2008 Mar 01;47 Suppl 1:S15-S19. [doi: 10.1097/QAI.0b013e3181605b26] [Medline: 18301129]
6. Prince M, Patel V, Saxena S, Maj M, Maselko J, Phillips MR, et al. No health without mental health. Lancet 2007 Sep 08;370(9590):859-877. [doi: 10.1016/S0140-6736(07)61238-0] [Medline: 17804063]
7. Pence BW, Mills JC, Bengtson AM, Gaynes BN, Breger TL, Cook RL, et al. Association of increased chronicity of depression with HIV appointment attendance, treatment failure, and mortality among HIV-infected adults in the United States. JAMA Psychiatry 2018 Apr 01;75(4):379-385 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.4726] [Medline: 29466531]
8. Anthes E. Mental health: there's an app for that. Nature 2016 Apr 07;532(7597):20-23. [doi: 10.1038/532020a] [Medline: 27078548]
9. van Ginneken N, Lewin S, Berridge V. What can be learnt from historical analysis? Community health workers and health policy in South Africa. A response to van Rensburg, H.C.J., Wouters, E., and de Wet, K.'s critique (in this issue). Soc Sci Med 2011 Apr;72(7):1025-1027 [FREE Full text] [doi: 10.1016/j.socscimed.2011.01.025]
10. Lofgren SM, Nakasujja N, Boulware DR. Systematic review of interventions for depression for people living with HIV in Africa. AIDS Behav 2018 Jan;22(1):1-8 [FREE Full text] [doi: 10.1007/s10461-017-1906-3] [Medline: 28900756]
11. Psychiatrists working in mental health sector (per 100,000). The Global Health Observatory, World Health Organization. 2019 Apr 25. URL: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/psychiatrists-working-in-mental-health-sector-(per-100-000) [accessed 2021-10-16]
12. Cheng LJ, Kumar PA, Wong SN, Lau Y. Technology-delivered psychotherapeutic interventions in improving depressive symptoms among people with HIV/AIDS: a systematic review and meta-analysis of randomised controlled trials. AIDS Behav 2020 Jun;24(6):1663-1675. [doi: 10.1007/s10461-019-02691-6] [Medline: 31587115]
13. Xiao Y, Shao Y, Na Z, Zhao W, Wang R, Fang S, et al. A systematic review and meta-analysis of telephone-based therapy targeting depressive symptoms among low-income people living with HIV. AIDS Behav 2021 Feb;25(2):414-426. [doi: 10.1007/s10461-020-02999-8] [Medline: 32809074]
14. Nakimuli-Mpungu E, Musisi S, Smith CM, Von Isenburg M, Akimana B, Shakarishvili A, et al. Mental health interventions for persons living with HIV in low- and middle-income countries: a systematic review. J Int AIDS Soc 2021 Jun;24 Suppl 2:e25722 [FREE Full text] [doi: 10.1002/jia2.25722] [Medline: 34164926]
15. Li J, Mo PK, Kahler CW, Lau JT. A three-arm randomised controlled trial to evaluate the efficacy of a positive psychology and social networking intervention in promoting mental health among HIV-infected men who have sex with men in China. Epidemiol Psychiatr Sci 2021 Mar 19;30:e24 [FREE Full text] [doi: 10.1017/S2045796021000081] [Medline: 33736740]
16. Guo Y, Hong YA, Cai W, Li L, Hao Y, Qiao J, et al. Effect of a WeChat-based intervention (Run4Love) on depressive symptoms among people living with HIV in China: a randomized controlled trial. J Med Internet Res 2020 Feb 09;22(2):e16715 [FREE Full text] [doi: 10.2196/16715] [Medline: 32044751]
17. Heckman TG, Carlson B. A randomized clinical trial of two telephone-delivered, mental health interventions for HIV-infected persons in rural areas of the United States. AIDS Behav 2007 Jan;11(1):5-14. [doi: 10.1007/s10461-006-9111-9] [Medline: 16705479]
18. Heckman TG, Heckman BD, Anderson T, Lovejoy TI, Mohr D, Sutton M, et al. Supportive-expressive and coping group teletherapies for HIV-infected older adults: a randomized clinical trial. AIDS Behav 2013 Nov;17(9):3034-3044 [FREE Full text] [doi: 10.1007/s10461-013-0441-0] [Medline: 23474642]
19. Guo Y, Xu Z, Qiao J, Hong YA, Zhang H, Zeng C, et al. Development and feasibility testing of an mHealth (Text Message and WeChat) intervention to improve the medication adherence and quality of life of people living with HIV in China: pilot randomized controlled trial. JMIR Mhealth Uhealth 2018 Sep 04;6(9):e10274 [FREE Full text] [doi: 10.2196/10274] [Medline: 30181109]
20. Wang H, Zhou J, Huang L, Li X, Fennie KP, Williams AB. Effects of nurse-delivered home visits combined with telephone calls on medication adherence and quality of life in HIV-infected heroin users in Hunan of China. J Clin Nurs 2010 Feb;19(3-4):380-388. [doi: 10.1111/j.1365-2702.2009.03048.x] [Medline: 20500277]
21. Ross R, Sawatphanit W, Suwansujarid T, Stidham AW, Drew BL, Creswell JW. The effect of telephone support on depressive symptoms among HIV-infected pregnant women in Thailand: an embedded mixed methods study. J Assoc Nurses AIDS Care 2013;24(5):e13-e24. [doi: 10.1016/j.jana.2012.08.005] [Medline: 23260038]
22. Schnall R, Cho H, Mangone A, Pichon A, Jia H. Mobile health technology for improving symptom management in low income persons living with HIV. AIDS Behav 2018 Oct;22(10):3373-3383 [FREE Full text] [doi: 10.1007/s10461-017-2014-0] [Medline: 29299790]
23. Drozd F, Skeie LG, Kraft P, Kvale D. A Web-based intervention trial for depressive symptoms and subjective well-being in patients with chronic HIV infection. AIDS Care 2014;26(9):1080-1089. [doi: 10.1080/09540121.2013.869541] [Medline: 24359563]
24. van Luenen S, Garnefski N, Spinhoven P, Kraaij V. Guided Internet-based intervention for people with HIV and depressive symptoms: a randomised controlled trial in the Netherlands. Lancet HIV 2018 Sep;5(9):e488-e497. [doi: 10.1016/S2352-3018(18)30133-4] [Medline: 30135045]

XSL•FO
RenderX

25. Guo Y, Hong YA, Qiao J, Xu Z, Zhang H, Zeng C, et al. Run4Love, a mHealth (WeChat-based) intervention to improve mental health of people living with HIV: a randomized controlled trial protocol. BMC Public Health 2018 Jun 26;18(1):793 [FREE Full text] [doi: 10.1186/s12889-018-5693-1] [Medline: 29940921]
26. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Meas 1977 Jun 1;1(3):385-401. [doi: 10.1177/014662167700100306]
27. Wang M, Armour C, Wu Y, Ren F, Zhu X, Yao S. Factor structure of the CES-D and measurement invariance across gender in Mainland Chinese adolescents. J Clin Psychol 2013 Sep;69(9):966-979. [doi: 10.1002/jclp.21978] [Medline: 23775279]
28. Zhang B, Fokkema M, Cuijpers P, Li J, Smits N, Beekman A. Measurement invariance of the Center for Epidemiological Studies Depression Scale (CES-D) among Chinese and Dutch elderly. BMC Med Res Methodol 2011 May 20;11:74 [FREE Full text] [doi: 10.1186/1471-2288-11-74] [Medline: 21595982]
29. Zimmerman M, Coryell W. Screening for major depressive disorder in the community: a comparison of measures. Psychol Assess 1994 Mar;6(1):71-74. [doi: 10.1037/1040-3590.6.1.71]
30. Hsiung PC, Fang CT, Wu CH, Sheng WH, Chen SC, Wang JD, et al. Validation of the WHOQOL-HIV BREF among HIV-infected patients in Taiwan. AIDS Care 2011 Aug;23(8):1035-1042. [doi: 10.1080/09540121.2010.543881] [Medline: 21500023]
31. WHOQOL-HIV bref, 2012 revision. World Health Organization. 2020. URL: https://apps.who.int/iris/handle/10665/77775 [accessed 2022-06-17]
32. Roberti JW, Harrington LN, Storch EA. Further psychometric support for the 10-item version of the perceived stress scale. J Coll Couns 2006;9(2):135-147. [doi: 10.1002/j.2161-1882.2006.tb00100.x]
33. Cai ZX, Li K, Zhang XC. Workplace stressors and coping strategies among Chinese psychiatric nurses. Perspect Psychiatr Care 2008 Oct;44(4):223-231. [doi: 10.1111/j.1744-6163.2008.00181.x] [Medline: 18826460]
34. Chiu FP, Tsang HW. Validation of the Chinese general self-efficacy scale among individuals with schizophrenia in Hong Kong. Int J Rehabil Res 2004 Jun;27(2):159-161. [doi: 10.1097/01.mrr.0000127640.55118.6b] [Medline: 15167116]
35. Berger BE, Ferrans CE, Lashley FR. Measuring stigma in people with HIV: psychometric assessment of the HIV stigma scale. Res Nurs Health 2001 Dec;24(6):518-529. [doi: 10.1002/nur.10011] [Medline: 11746080]
36. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-IV). Washington, DC, USA: American Psychiatric Association; 1994.
37. Hu B, Lin LF, Zhuang MQ, Yuan ZY, Li SY, Yang YJ, et al. Reliability and relative validity of three physical activity questionnaires in Taizhou population of China: the Taizhou Longitudinal Study. Public Health 2015 Sep;129(9):1211-1217. [doi: 10.1016/j.puhe.2015.03.024] [Medline: 25957853]
38. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 1986 Mar;42(1):121-130. [Medline: 3719049]
39. Thalheimer W, Cook S. How to calculate effect sizes from published research: a simplified methodology. Work-Learning Research. 2002. URL: https://www.bwgriffin.com/gsu/courses/edur9131/content/Effect_Sizes_pdf5.pdf [accessed 2021-10-12]
40. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]
41. McCall HC, Hadjistavropoulos HD, Sundström CR. Exploring the role of persuasive design in unguided Internet-delivered cognitive behavioral therapy for depression and anxiety among adults: systematic review, meta-analysis, and meta-regression. J Med Internet Res 2021 Apr 29;23(4):e26939 [FREE Full text] [doi: 10.2196/26939] [Medline: 33913811]
42. 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. [doi: 10.1001/jamapsychiatry.2020.4364] [Medline: 33471111]
43. Fu Z, Burger H, Arjadi R, Bockting CL. Effectiveness of digital psychological interventions for mental health problems in low-income and middle-income countries: a systematic review and meta-analysis. Lancet Psychiatry 2020 Oct;7(10):851-864 [FREE Full text] [doi: 10.1016/S2215-0366(20)30256-X] [Medline: 32866459]
44. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
45. Deady M, Choi I, Calvo RA, Glozier N, Christensen H, Harvey SB. eHealth interventions for the prevention of depression and anxiety in the general population: a systematic review and meta-analysis. BMC Psychiatry 2017 Aug 29;17(1):310 [FREE Full text] [doi: 10.1186/s12888-017-1473-1] [Medline: 28851342]
46. Safren SA, Bedoya CA, O'Cleirigh C, Biello KB, Pinkston MM, Stein MD, et al. Cognitive behavioural therapy for adherence and depression in patients with HIV: a three-arm randomised controlled trial. Lancet HIV 2016 Nov;3(11):e529-e538 [FREE Full text] [doi: 10.1016/S2352-3018(16)30053-4] [Medline: 27658881]
47. Nakimuli-Mpungu E, Musisi S, Wamala K, Okello J, Ndyanabangi S, Birungi J, et al. Effectiveness and cost-effectiveness of group support psychotherapy delivered by trained lay health workers for depression treatment among people with HIV in Uganda: a cluster-randomised trial. Lancet Glob Health 2020 Mar;8(3):e387-e398 [FREE Full text] [doi: 10.1016/S2214-109X(19)30548-0] [Medline: 32035035]

XSL•FO
RenderX

48. Safren SA, O'Cleirigh CM, Bullis JR, Otto MW, Stein MD, Pollack MH. Cognitive behavioral therapy for adherence and depression (CBT-AD) in HIV-infected injection drug users: a randomized controlled trial. J Consult Clin Psychol 2012 Jun;80(3):404-415 [FREE Full text] [doi: 10.1037/a0028208] [Medline: 22545737]
49. Sakurai-Yageta M, Kawame H, Kuriyama S, Hozawa A, Nakaya N, Nagami F, et al. A training and education program for genome medical research coordinators in the genome cohort study of the Tohoku Medical Megabank Organization. BMC Med Educ 2019 Aug 02;19(1):297 [FREE Full text] [doi: 10.1186/s12909-019-1725-5] [Medline: 31375111]
50. Zeng Y, Guo Y, Li L, Hong YA, Li Y, Zhu M, et al. Relationship between patient engagement and depressive symptoms among people living with HIV in a mobile health intervention: secondary analysis of a randomized controlled trial. JMIR Mhealth Uhealth 2020 Oct 29;8(10):e20847 [FREE Full text] [doi: 10.2196/20847] [Medline: 33118956]
51. Christensen H, Griffiths KM, Farrer L. Adherence in Internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
52. Enrique A, Palacios JE, Ryan H, Richards D. Exploring the relationship between usage and outcomes of an Internet-based intervention for individuals with depressive symptoms: secondary analysis of data from a randomized controlled trial. J Med Internet Res 2019 Aug 01;21(8):e12775 [FREE Full text] [doi: 10.2196/12775] [Medline: 31373272]
53. Chin WY, Choi EP, Chan KT, Wong CK. The psychometric properties of the center for epidemiologic studies depression scale in Chinese primary care patients: factor structure, construct validity, reliability, sensitivity and responsiveness. PLoS One 2015 Aug 7;10(8):e0135131 [FREE Full text] [doi: 10.1371/journal.pone.0135131] [Medline: 26252739]
54. Lambert SD, Clover K, Pallant JF, Britton B, King MT, Mitchell AJ, et al. Making sense of variations in prevalence estimates of depression in cancer: a co-calibration of commonly used depression scales using Rasch analysis. J Natl Compr Canc Netw 2015 Oct;13(10):1203-1211. [doi: 10.6004/jnccn.2015.0149] [Medline: 26483060]
55. Smith PJ, Merwin RM. The role of exercise in management of mental health disorders: an integrative review. Annu Rev Med 2021 Jan 27;72:45-62 [FREE Full text] [doi: 10.1146/annurev-med-060619-022943] [Medline: 33256493]
56. Chekroud AM, Trugerman A. The opportunity for exercise to improve population mental health. JAMA Psychiatry 2019 Nov 01;76(11):1206-1207. [doi: 10.1001/jamapsychiatry.2019.2282] [Medline: 31483446]
57. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
58. van Beugen S, Ferwerda M, Hoeve D, Rovers MM, Spillekom-van Koulil S, van Middendorp H, et al. Internet-based cognitive behavioral therapy for patients with chronic somatic conditions: a meta-analytic review. J Med Internet Res 2014 Mar 27;16(3):e88 [FREE Full text] [doi: 10.2196/jmir.2777] [Medline: 24675372]

Abbreviations

CBSM: cognitive behavioral stress management
CES-D: Center for Epidemiological Studies–Depression
CONSORT–EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
GEE: generalized estimating equation
mHealth: mobile health
PHQ-9: 9-item Patient Health Questionnaire
PSS: Perceived Stress Scale
QOL: quality of life
RCT: randomized controlled trial
SWCQ: Simplified Ways of Coping Questionnaire
WHOQOL-HIV BREF: World Health Organization Quality of Life HIV short version

Edited by A Mavragani; submitted 26.01.22; peer-reviewed by J Mitchell, P Cook; comments to author 08.05.22; revised version received 23.05.22; accepted 24.05.22; published 28.06.22.

Please cite as:
Guo Y, Li Y, Yu C, Xu H, Hong YA, Wang X, Zhang N, Zeng Y, Monroe-Wise A, Li L, Liu C, Cai W, Lin A
Long-term Effects of a Social Media–Based Intervention (Run4Love) on Depressive Symptoms of People Living With HIV: 3-Year Follow-up of a Randomized Controlled Trial
J Med Internet Res 2022;24(6):e36809
URL: https://www.jmir.org/2022/6/e36809
doi:10.2196/36809
PMID:35763324

XSL•FO
RenderX

©Yan Guo, Yingqi Li, Chuanchuan Yu, He Xu, Y Alicia Hong, Xiaolan Wang, Nanxiang Zhang, Yu Zeng, Aliza Monroe-Wise, Linghua Li, Cong Liu, Weiping Cai, Aihua Lin. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.06.2022. 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 Novel Score for mHealth Apps to Predict and Prevent Mortality: Further Validation and Adaptation to the US Population Using the US National Health and Nutrition Examination Survey Data Set

Shatha Elnakib[1*], PhD; Andres I Vecino-Ortiz[1*], MSc, MD, PhD; Dustin G Gibson[1*], PhD; Smisha Agarwal[1*], PhD; Antonio J Trujillo[1*], MPP, PhD; Yifan Zhu[1*], MHS, PhD; Alain B Labrique[1*], MHS, MS, PhD

Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[*]all authors contributed equally

Corresponding Author:
Andres I Vecino-Ortiz, MSc, MD, PhD
Department of International Health
Johns Hopkins Bloomberg School of Public Health
615 N Wolfe Street
E8620
Baltimore, MD, 21205
United States
Phone: 1 4109554711
Email: andres.vecino@gmail.com

Abstract

Background: The *C-Score*, which is an individual health score, is based on a predictive model validated in the UK and US populations. It was designed to serve as an individualized *point-in-time* health assessment tool that could be integrated into clinical counseling or consumer-facing digital health tools to encourage lifestyle modifications that reduce the risk of premature death.

Objective: Our study aimed to conduct an external validation of the C-Score in the US population and expand the original score to improve its predictive capabilities in the US population. The C-Score is intended for mobile health apps on wearable devices.

Methods: We conducted a literature review to identify relevant variables that were missing in the original C-Score. Subsequently, we used data from the 2005 to 2014 US National Health and Nutrition Examination Survey (NHANES; N=21,015) to test the capacity of the model to predict all-cause mortality. We used NHANES III data from 1988 to 1994 (N=1440) to conduct an external validation of the test. Only participants with complete data were included in this study. Discrimination and calibration tests were conducted to assess the operational characteristics of the adapted C-Score from receiver operating curves and a design-based goodness-of-fit test.

Results: Higher C-Scores were associated with reduced odds of all-cause mortality (odds ratio 0.96, $P<.001$). We found a good fit of the C-Score for all-cause mortality with an area under the curve (AUC) of 0.72. Among participants aged between 40 and 69 years, C-Score models had a good fit for all-cause mortality and an AUC >0.72. A sensitivity analysis using NHANES III data (1988-1994) was performed, yielding similar results. The inclusion of sociodemographic and clinical variables in the basic C-Score increased the AUCs from 0.72 (95% CI 0.71-0.73) to 0.87 (95% CI 0.85-0.88).

Conclusions: Our study shows that this digital biomarker, the C-Score, has good capabilities to predict all-cause mortality in the general US population. An expanded health score can predict 87% of the mortality in the US population. This model can be used as an instrument to assess individual mortality risk and as a counseling tool to motivate behavior changes and lifestyle modifications.

(J Med Internet Res 2022;24(6):e36787) doi:10.2196/36787

KEYWORDS

C-Score; validation; mortality; predictive models; mobile phone

XSL•FO
RenderX

Introduction

Background

In the United States, 60% of all adults have at least one chronic condition, and 42% have >1 [1,2], leading to >1.7 million deaths annually [3]. Reliable indicators of current and future health can be integrated into digitally enabled strategies to modify behaviors and reduce the risk of adverse outcomes and death. Therefore, there is a growing demand for evidence-based tools, supported by ubiquitous innovations such as wearable technologies, that could help clinicians and individuals to calculate the risk of disease and predict future health outcomes [4,5]. Such tools and technologies often collect data on risk factors that can be integrated into an index to provide information on current and future disease risks. The advent of wearable technologies and other readily accessible nonclinical sources of anthropometric or biometric data has challenged us to evaluate the value of extending classical metrics to achieve greater precision and predictive accuracy. When accurate, such tools have tremendous potential to inform lifestyle improvements and drive sustained changes in modifiable risk factors that can enhance health status.

In recent years, a number of risk-scoring algorithms and models have demonstrated the capacity to predict adverse health outcomes such as the risk of developing cardiovascular disease [6], diabetes [7], hypertension [8], and very specific cancers [9] and predict complications following surgery [10]. However, existing models or applications are often reserved for use by clinicians or incorporate the mathematical analysis of data points that require invasive testing (eg, blood tests). These models are rarely presented in friendly digital formats or provide advice to clients on specific modifiable behaviors. In addition, most prognostic indices have primarily focused on predicting short-term mortality among older adults and high-risk individuals, whereas fewer indices have focused on prognostic health assessment of the general population [11-16].

The C-Score, derived from metrics that are easily reported by a person and augmented by measures derivable from most smartphones, is designed as a tool for individualized health risk prediction and can be used as a basis for directing targeted lifestyle modifications to reduce the risk of future adverse outcomes. Clift et al [17] developed and validated the C-Score model using a prospective cohort analysis, leveraging the UK Biobank data set [17]. They found that the C-Score had good predictive capabilities for all-cause mortality within 10 years for adults aged between 40 and 69 years. The points-based model had good discrimination with a c-statistic of 0.66, and a Cox model with the C-Score and age had improved discrimination (c-statistic 0.74) and good calibration. Although the UK Biobank data set is an unparalleled resource of extensive health information with >400 peer-reviewed publications to date, its sampling population is volunteer based and hence not entirely representative of the UK population [18]. Keyes et al [19] articulated several concerns related to the nonrepresentativeness of this sample population, whereas Batty et al [20] concluded that risk factor associations in the UK Biobank seem to be generalizable, after comparing with pooled data from the Health Surveys for England and the Scottish Health Surveys.

Objective

In this study, we conducted an external validation of the C-Score in the US population and expanded the original score to improve its predictive capabilities in the US population [17]. The C-Score is a mobile health app that can be used on wearable devices.

For the external validation, we assessed the discrimination and calibration of the original C-Score in the US population using the US National Health and Nutrition Examination Survey (NHANES). For the expansion and adaptation of the model, we reviewed the literature and tested additional predictors of all-cause mortality in the US population to improve the predictive capacity of the model.

Methods

The C-Score

The risk models were developed following an extensive literature review that identified key risk factors for all-cause mortality [17]. The review yielded eight key predictor variables: age, cigarette consumption, alcohol consumption, sleeping duration, self-rated health, waist to height (WtHR) ratio, resting heart rate, and reaction time. Given the interest in modifiable risk factors, age was not included in the calculation of the score. Relative weightings, which were developed by Clift et al [17], using hazard ratios extracted from each identified study, were used to generate a points-based score. The lowest risk was denoted with a 0, with increases in scores indicating higher than optimal risk. The overall score totaled 25 points and was multiplied by 4 to generate a sum of 100 (Table 1). The score operates in a penalizing fashion, with users starting with 100 points and losing points for each health domain in accordance with the hazard ratio extracted from the literature. Thus, the C-Score is an evidence-based consolidated index that uses 7 parameters to predict mortality. The points-based C-Score model performed moderately well in the United Kingdom with an area under the curve (AUC) >0.66 and high calibration [17]. More detailed information on the development of the score can be found elsewhere [17].

XSL•FO
RenderX

Table 1. Points-based score assigned to each explanatory variable for the original C-Score model.[a]

C-Score input	Points assigned, range
Resting heart rate (beats per minute)	0-7.83
Average hours of sleep per night	0-10.26
Waist to height ratio	0-10.8
Self-rated health (ordinal scale: excellent, good, fair, and poor)	0-31.32
Cigarette smoking (status and cigarettes per day)	0-12.96
Alcohol consumption (units per week)	0-19.44
Reaction time	0-6.75

[a]The reaction time variable is not present in the main National Health and Nutrition Examination Survey sample. Therefore, we did not include this in the main analysis. For the sensitivity analysis, we did not include alcohol consumption or sleep duration as these variables were not present in the National Health and Nutrition Examination Survey III.

Data Source and Validation Population

The NHANES is a large cross-sectional population-based survey that combines interviews with physical examinations, thereby serving as a rich source of both self-reported and directly measured biometric data. Each survey round includes a nationally representative sample of approximately 5000 individuals and is conducted regularly. The NHANES questionnaire elicits information pertaining to sociodemographic, dietary, physical, and health-related characteristics. Details of the NHANES study design have been described in previous studies [21,22]. To validate the C-Score, we pooled the NHANES survey data from 2005 to 2014, resulting in data from 28,078 participants.

As mortality data are not readily collected as part of the NHANES, the National Center for Health Statistics has matched 1999 to 2014 data with death certificate records from the National Death Index (NDI), which have been made available for public use. Mortality ascertainment was based on a probabilistic match between the NHANES and NDI death certificate records. These data were, in turn, linked with NDI mortality data using participants' social security number, first name, middle initial name, last name or father's surname, month of birth, day of birth, year of birth, state of birth, state of residence, race, and sex, yielding a sample of 28,033 participants with complete information on mortality. The methodology for the data linkage has been described in detail by the National Center for Health Statistics [23].

We linked the anonymized NHANES survey data with the anonymized NDI mortality data, which included mortality follow-up data from December 31, 2015. The matching yielded a sample of 28,033 participants. This was the sample for which the external validation of the C-Score was conducted. It was also the sample for which the C-Score model was adapted and expanded to improve its performance in the US population.

Following the development of the adapted model, we conducted another round of validation as a sensitivity analysis, using data obtained from NHANES III, a survey conducted from 1988 to 1994, which included the mortality data of 6591 participants. The NHANES III data analysis missed 2 of the 7 variables included in the risk model (sleep duration and alcohol consumption); therefore, the C-Score was calculated in the absence of these risk factors.

Predictor Variables

The explanatory variables in this study were extracted from the questionnaire data and examination data from the 5 NHANES waves. The questionnaire data included age (in years), cigarette consumption (average number of cigarettes per day), alcohol consumption (average number of alcoholic drinks per week), and sleep duration (hours per day). Self-rated health was transformed from a 5-point scale (from poor to excellent) into a 4-point scale in which *excellent* and *very good* health were merged into one category to better match with the UK Biobank variable. The NHANES examination data were collected by trained health technicians, and information was collected on WtHR (waist circumference divided by height) and resting heart rate (beats per minute). Reaction time was missing from the 2005 to 2014 NHANES data but was measured as part of a computerized Neurobehavioral Evaluation System 2.

Expanding the Set of Variables for the Original C-Score Model

We conducted a subsequent literature review of predictors of all-cause mortality in the United States and identified a set of clinical factors and sociodemographic variables for which there is evidence of an association with mortality. As we wanted to ensure the usability of the smartphone app, we sought to create the most parsimonious model with maximal performance based on the combination of the Akaike Information Criterion, AUC, and goodness of fit. In addition to the variables used to construct the original C-Score, we investigated the predictive value of including sociodemographic characteristics such as gender, race or ethnicity, marital status, and educational attainment, as well as simple medical history variables shown to be associated with mortality, such as binary variables *ever diagnosis of high blood pressure* [24,25] and *ever diagnosis with hypercholesterolemia* [26]. Finally, we included interaction terms (C-Score interacting with each of the additional variables) to explore whether a maximally complex model would perform better.

Statistical Analysis

To validate the original C-Score, we tested the model using the pooled NHANES data. However, as NHANES lacks the reaction

XSL•FO
RenderX

time variable, which is one of the variables used to compute the C-Score, we conducted a sensitivity analysis using data from NHANES III, a smaller survey that collected data on reaction time, to measure the marginal effect of the reaction time variable. Following the validation and sensitivity analysis, we incorporated additional variables into the model and investigated their internal and external validity.

Validating the Original C-Score

For all models, we used a complete case approach, whereby the only participants included were those for whom a risk score based on all risk factors could be computed (ie, for whom there were no missing data on any of the included variables). We pooled NHANES data from 2005 to 2014, which included 6 out of 7 variables included in the original C-Score model (missing reaction time). As the NHANES survey did not have the reaction time variable, all individuals were assumed to have the maximum score for that variable in this validation exercise.

In the complete case analysis, there were 21,015 participants (aged 18-85 years) with complete information on mortality, age, and all metrics included in the C-Score. This population with a wide age range was selected as one would expect to see greater variability in the exposure variables, thus permitting better exploration of the models. Furthermore, to produce estimates with a population similar to that in the Clift et al [17] study, participants aged 40 to 69 years were analyzed separately [17]. The complete case analysis for this age-restricted subsample included 9994 participants. For each prediction model, we assessed the model's performance by investigating its discrimination—the extent to which it can adequately discriminate between those who will have the discrete event and those who will not—and calibration—the extent to which the observed and predicted probabilities agree [27,28]. The area under the receiver operating characteristic curves (c-statistics) and a design-based goodness-of-fit test for estimating the F-adjusted mean residual test [29] were used to assess discrimination and calibration, respectively [29]. Unlike the original model, we could not use Cox regressions, given that the NHANES data sets are repeated cross-sections and we did not have the benefits of a longitudinal panel to use Cox. Therefore, our model estimates mortality within a 10-year period (time of follow-up for the NHANES mortality link) instead of the survival time.

In all cases, we ran an additional analysis including both the C-Score and the logarithm of age, as performed by Clift et al [17].

Sensitivity Analysis of the Original C-Score

As the NHANES survey lacks one of the variables used for validation—the reaction time variable—we performed a sensitivity analysis with a different data set. We conducted a sensitivity analysis using data from NHANES III, a survey conducted from 1988 to 1994 containing data for 33,994 people aged ≥2 months, including mortality data, to ascertain the marginal effect of the reaction time variable from the analysis. Owing to the limited number of people with neurobehavioral indicators, we did not impose age limits in this sensitivity analysis.

The NHANES III data set contains the reaction time variable but lacks 2 of the 7 variables included in the risk model (sleep duration and alcohol consumption). The lack of these variables should drive the fit and calibration of the model downward, and therefore, any results in this sensitivity analysis would be conservative. In this sensitivity analysis, we tested the sensitivity of the 5-variable model to the inclusion and exclusion of the reaction time variable. The complete case analysis yielded data from 1440 participants.

All data analyses were performed using Stata 15 (StataCorp), using survey weights to specify the survey and sample design characteristics. In addition, a dummy variable for the survey round was included in the models with pooled data. For all models, P values <.05 were regarded as statistically significant.

Adapting the C-Score to the US Population and Measuring Its Internal and External Validity

We examined the impact of including additional variables on calibration and discrimination [27,28]. We used the area under the receiver operating characteristic curve (AUC), or the c-statistic, to assess the discrimination of the adapted models. We tested both internal and external validities. We used a k-fold cross-validation procedure to assess within-study model validity [30]. We estimated AUC based on 10 random samples (the *test* samples) that were independent of the samples used to train the model (the *training* sample), averaging the AUCs associated with each individual fold and bootstrapping the cross-validated AUCs to obtain 95% CIs. To assess calibration, we used a design-based goodness-of-fit test of logistic regressions, as well as calibration curves developed using locally weighted scatterplot smoothing to compare fitted outcome probabilities with observed outcome probabilities [31]. We also report the Akaike Information Criterion. For the external validation, we assessed the best-performing model using the NHANES III data set.

This study follows the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) guidelines for multivariable prediction models [32].

Ethics Approval

The NHANES survey is approved by the National Center for Health Statistics Institutional Ethics Review Board. Written informed consent was obtained from all adult participants. Ethical approval to conduct this analysis was not required as we used publicly available data. This study was approved by the institutional review board of the Johns Hopkins Bloomberg School of Public Health and was deemed nonhuman subject research (13743).

Data Availability

The data sets analyzed in this study are publicly available on the NHANES website. The C-Scores are proprietary information but can be provided as restricted data to the reviewers.

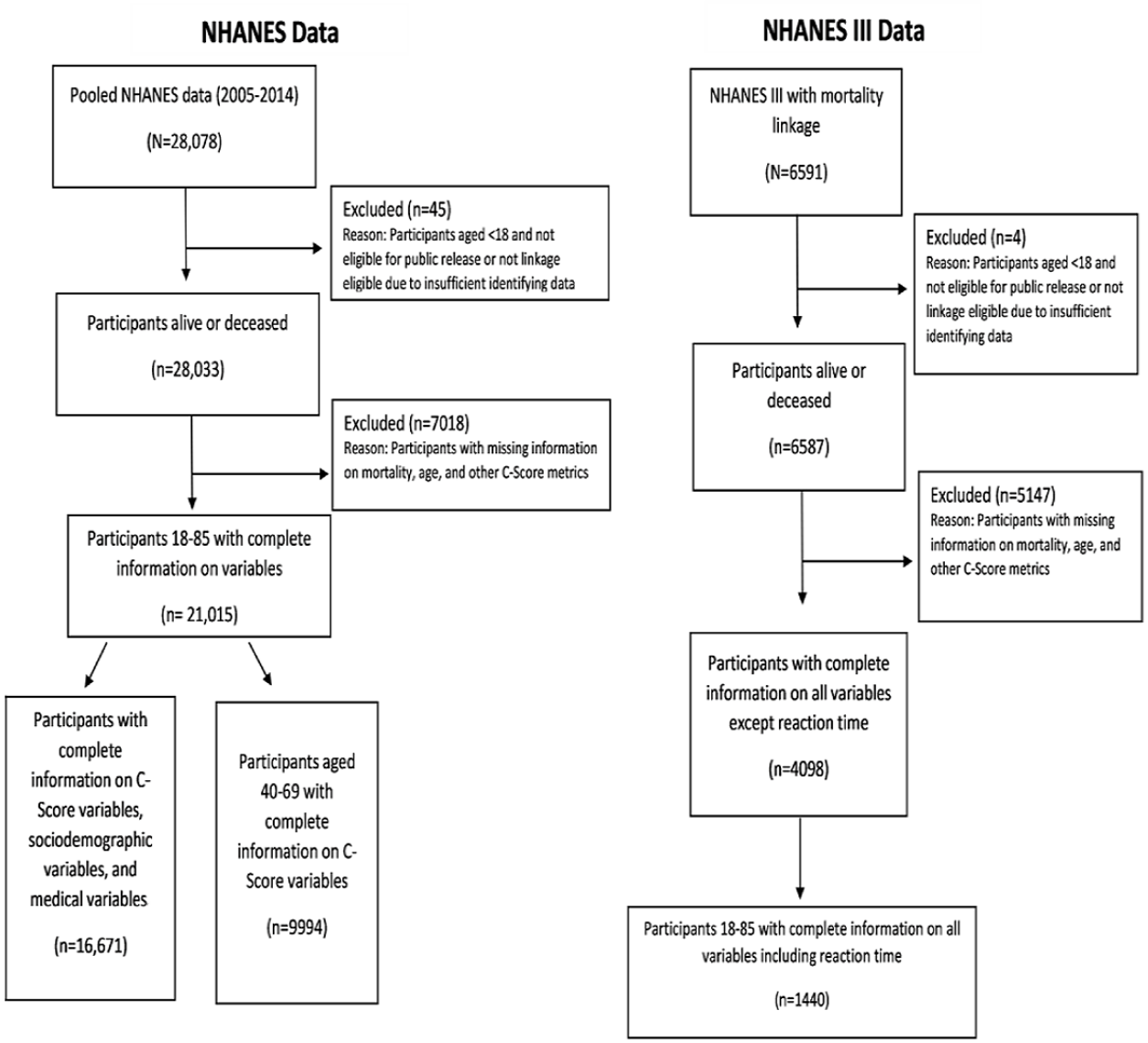
XSL•FO
RenderX

Results

Validating the Original C-Score

From 2005 to 2014, we obtained 28,078 records from the NHANES. Of these, 99.84% (28,033/28,078) were matched with mortality data and 74.84% (21,015/28,078) had complete information on all variables. A flowchart of the sample sizes for the main analysis, sensitivity analysis, and adaptation of the model is shown in Figure 1. The basic characteristics of the study sample are presented in Table 2.

Figure 1. Flowchart for sample sizes for National Health and Nutrition Examination Survey (NHANES) study samples.



XSL•FO
RenderX

Table 2. Descriptive statistics for the different samples used in the study.[a]

Variable	Full study sample (N=21,015)	Age-restricted sample (40-69 years; N=9994)	NHANES[b] III subsample for the sensitivity analysis (N=1440)
Age (years), mean (SD)	47.43 (17.97)	53.78 (8.53)	47.85 (5.80)
Sex, n (%)			
Male	10,094 (48.03)	4764 (47.67)	655 (45.49)
Female	10,921 (51.97)	5230 (52.33)	785 (54.51)
Ethnicity, n (%)			
Mexican American	3334 (15.86)	1621 (16.22)	373 (25.9)
Other Hispanic	1891 (9)	997 (9.98)	36 (2.5)
Non-Hispanic White	9519 (45.3)	4215 (42.18)	615 (42.71)
Non-Hispanic Black	4413 (21)	2314 (23.15)	403 (27.99)
Other race—including multiracial	1858 (8.84)	847 (8.48)	13 (0.9)
Resting heart rate, mean (SD)	72.83 (12.11)	72.42 (11.94)	69.14 (10.80)
Waist to height ratio, mean (SD)	0.59 (0.10)	0.60 (0.09)	0.58 (0.09)
Weekly alcohol intake, mean (SD)	3.63 (8.36)	3.88 (9.07)	N/A[c]
Sleep duration, mean (SD)	6.85 (1.40)	6.73 (1.38)	N/A
Self-rated health, n (%)			
Excellent or very good	8169 (38.87)	3515 (35.17)	558 (38.75)
Good	8425 (40.09)	4007 (40.09)	538 (37.36)
Fair	3790 (18.03)	2078 (20.79)	292 (20.28)
Poor	631 (3)	394 (3.94)	52 (3.61)
Number of cigarettes per day, mean (SD)	3.28 (7.60)	3.94 (8.63)	6.67 (11.71)
Comorbidities, n (%)	3790 (18.41)	2217 (22.27)	183 (12.81)

[a]Survey weights are not included in this descriptive analysis.

[b]NHANES: National Health and Nutrition Examination Survey.

[c]N/A: not applicable.

There were 21,015 participants in the pooled data with complete information on mortality, age, and other C-Score metrics. The mean age of the sample was 47.43 (SD 17.97) years, the mean resting heart rate was 72.83 (SD 12.11) beats per minute, the mean WtHR was 0.59 (SD 0.10), mean weekly alcohol intake was 3.63 (SD 8.63) drinks per week, and mean sleep duration was 6.85 (SD 1.40) hours. For self-rated health, 38.87% (8169/21,015) were *excellent*, 40.09% (8425/21,015) were *good*, 18.03% (3790/21,015) were *fair*, and 3% (631/21,015) were *poor*. There were 48.03% (10,094/21,015) men and 51.97% (10,921/21,015) women. In the study sample, 18.41% (3790/21,015) had existing comorbidities such as diabetes, stroke, coronary heart disease, angina, or heart attack. In terms of the main study outcome, 6.07% (1276/21,015) of patients had died as of December 31, 2015.

In the validation subsample (among participants aged 40-69 years), there were 9994 participants with a mean age of 53.78 (SD 8.53) years, mean resting heart rate of 72.42 (SD 11.94) beats per minute, mean WtHR of 0.60 (SD 0.09), mean weekly alcohol intake of 3.88 (SD 9.07) drinks per week, and mean sleep duration of 6.73 (SD 1.38) hours. For self-rated health, 35.17% (3515/9994) were *excellent*, 40.09% (4007/9994) were *good*, 20.79% (2078/9994) were *fair*, and 3.94% (394/9994) were *poor*. There were 47.67% (4764/9994) of men and 52.33% (5230/9994) of women. In terms of comorbidities, 22.27% (2217 or 22.27%) reported a diagnosis of diabetes, stroke, coronary heart disease, angina, or heart attack. In terms of the study outcome, 95.38% (9532/9994) of participants were alive, and 4.32% (462/9994) had died as of December 31, 2015.

Table 3 shows that in the study sample, higher C-Scores were related to a reduction in the occurrence of all-cause mortality (odds ratio 0.96, $P<.001$, 95% CI 0.95-0.96). The C-Score model showed a good fit for all-cause mortality in this population, with an AUC of approximately 0.72 (95% CI 0.70-0.73). After adding the log of age as a covariate in this model, the calibration test rejected the null hypothesis of good fit; however, the AUC increased to 0.86 (95% CI 0.85-0.87).

XSL•FO
RenderX

Table 3. Performance of the C-Score models for all-cause mortality by subsample.[a]

Outcome	C-Score model					C-Score plus log (age)				
	Score OR[b] (*P* value)	*F*-adjusted test statistic		AUC[c] (95% CI)	AIC[d]	Score OR (*P* value)	*F*-adjusted test statistic		AUC (95% CI)	AIC
		F test (*df*)	*P* value (fit)				*F* test (*df*)	*P* value (fit)		
Full study sample (N=21,015)	0.96 (<.001)	0.52 (9,71)	.86 (good)	0.72 (0.70-0.73)	8897.78	.96 (<.001)	7.25 (9,71)	<.001 (poor)	0.86 (0.85-0.87)	7272.50
Age-restricted sample (40-69 years; N=9994)	0.95 (<.001)	1.16 (9, 71)	0.34 (good)	0.72 (0.70-0.75)	3458.24	.95 (<.001)	0.50 (9,71)	.87 (good)	0.75 (0.73-0.77)	3366.48

[a]All models include dummy variables for the survey rounds. Survey weights were included in all analyses.

[b]OR: odds ratio.

[c]AUC: area under the curve.

[d]AIC: Akaike Information Criterion.

Table 3 shows that in the full study sample, the model demonstrated a good fit when not including the logarithm of age. Among the participants aged between 40 to 69 years, C-Score models, both with and without log age, had a good fit for all-cause mortality. Values of AUC ranged between 0.72 (95% CI 0.70-0.75) to 0.75 (95% CI 0.73-0.77).

Sensitivity Analysis

In the sensitivity analysis, we obtained data from NHANES III (1988-1994) on 6591 participants, of whom 21.85% (1440/6591) had complete data to conduct the validation. Table 4 shows the C-Score model had generally a good fit for all-cause mortality and an AUC of 0.68 (95% CI 0.65-0.72). The addition of reaction time worsened the model fit. The tables show that in the predictive C-Score model without reaction time but with age, all-cause mortality had a good fit, with an AUC of 0.72 (95% CI 0.69-0.75). After adding reaction time, the AUC for all-cause mortality did not differ.

Table 4. Sensitivity analysis on all-cause mortality for the marginal effect of the reaction time variable using NHANES[a] III (N=1440).[b]

Outcome	C-Score model				
	Score OR[c] (*P* value)	*F*-adjusted test statistic		AUC[d] (95% CI)	AIC[e]
		F test (*df*)	*P* value (fit)		
C-Score model performance with reaction time	0.92 (<.001)	2.97 (9,41)	.01 (poor)	0.68 (0.65-0.72)	1556.57
C-Score model performance without reaction time	0.91 (<.001)	1.82 (9,41)	.09 (good)	0.68 (0.65-0.72)	1555.48
C-Score model plus log age performance with reaction time	0.92 (<.001)	0.86 (9,41)	.56 (good)	0.72 (0.69-0.75)	1438.43
C-Score model plus log age performance without reaction time	0.92 (<.001)	0.97 (9,41)	.48 (good)	0.72 (0.69-0.75)	1485.04

[a]NHANES: National Health and Nutrition Examination Survey.

[b]All models included a dummy variable for the survey rounds. Survey weights were included in all analyses. The C-Score was calculated using five out of seven covariates: waist to height ratio, self-rated health, resting heart rate, smoking, and reaction time. The C-Score was calculated using 4 out of 7 covariates.

[c]OR: odds ratio.

[d]AUC: area under the curve.

[e]AIC: Akaike Information Criterion.

Adapting the C-Score to the US Population and Measuring Its Internal and External Validity

Overview

Of the 21,015 participants with complete information on the C-Score metrics, 20,626 (98.15%) had information on sociodemographic characteristics and of those, 16,671 (80.82%) had complete information on medical history variables. Thus, the final analytic sample in which the C-Score was adapted comprised 16,671 participants. Table 5 outlines the characteristics of this sample. The average age of the respondents in this sample was 50.43 (SD 17.32) years, and a little more than half (8831/16,671, 52.97%) were female. The mean resting heart rate was 72.53 (SD 12.04) beats per minute, mean WtHR was 0.59 (SD 0.10), mean weekly alcohol intake was 3.32 (SD 7.37) drinks per week, and mean sleep duration at night was 6.84 (SD 1.40) hours. For self-rated health, 39.48% (6581/16,671) reported *excellent health*, 40.09% (6617/16,671) reported *good health*, 18.03% (2948/16,671) reported *fair health*, and 3% (525/16,671) reported *poor health*. Approximately 21.03% (3497/16,671) of the respondents had existing comorbidities such as diabetes, stroke, coronary heart disease, angina, or heart attack. There were 6.3% (1062/16,671) deaths recorded in the analytic sample.

XSL•FO
RenderX

Table 5. Characteristics of the research sample (N=16,671).

Variable	Analytical sample
Age (years), mean (SD)	50.43 (17.32)
Sex, n (%)	
Male	7840 (47.03)
Female	8831 (52.97)
Ethnicity, n (%)	
Mexican American	2142 (12.85)
Other Hispanic	1447 (8.68)
Non-Hispanic White	7944 (47.65)
Non-Hispanic Black	3543 (21.25)
Other race (including multiracial)	1595 (9.57)
Resting heart rate (beats per minute), mean (SD)	72.53 (12.04)
Waist to height ratio, mean (SD)	0.59 (0.096)
Weekly alcohol intake (drinks per week), mean (SD)	3.32 (7.37)
Sleep duration (hours per night), mean (SD)	6.84 (1.39)
Self-rated health, n (%)	
Excellent or very good	6581 (39.48)
Good	6617 (39.69)
Fair	2948 (17.68)
Poor	525 (3.15)
Number of cigarettes per day, mean (SD)	2.97 (7.25)
Comorbidities, n (%)	3497 (21.03)
Deaths, n (%)	1062 (6.3)

The addition of sociodemographic variables and medical history variables (model 3), in contrast, similarly increased the AUC of the original C-Score model from 0.72 to an AUC of 0.87 (95% CI 0.86-0.88), although without a loss in the goodness of fit.

Upon inclusion of interaction terms between each of the covariates and the C-Score variable, we did not obtain significant increases in AUC or fit, indicating that this more complex model does not offer much improvement compared with a more parsimonious model. In addition, the C-Score odds ratio was not significant, implying no change in the odds of all-cause mortality associated with the change in the C-Score.

Table 6 and Figures 2 and 3 compare the performance of the expanded models with that of the basic C-Score model. The addition of basic sociodemographic variables to the C-Score model in model 2 increased discrimination considerably, as evidenced by the c-statistic of 0.87 (95% CI 0.85-0.88) compared with 0.72 (95% CI 0.71-0.73) yielded by the original C-Score model. However, although the addition of sociodemographic variables lowered the Akaike Information Criterion, the model was not well calibrated as the calibration test rejected the null hypothesis of good fit (P=.04).

XSL•FO
RenderX

Table 6. Performance of original C-Score versus expanded models for all-cause mortality.[a]

Model	Independent variables	Participants, N	Score OR[b] (*P* value)	Goodness of fit (*P* value)	AUC[c] (95% CI)	AIC[d]
1	C-Score[e]	21,015	0.96 (<.001)	Good fit (.86)	0.72 (0.70-0.73)	8897.78
2	C-Score[e]+sociodemographic variables[f]	20,626	0.97 (<.001)	Poor fit (.04)	0.87 (0.85-0.88)	6977.07
3	C-Score[e]+sociodemographic variables[f]+medical history[g]	16,671	0.96 (<.001)	Good fit (.06)	0.87 (0.86-0.88)	5705.134
4	C-Score[e]+sociodemographic variables[f]+medical history[g]+interactions[h]	16,671	1.0 (.25)	Good fit (.19)	0.87 (0.86-0.89)	5693.319

[a]All models include dummy variables for the survey rounds. Survey weights were included in all analyses.

[b]OR: odds ratio.

[c]AUC: area under the curve.

[d]AIC: Akaike Information Criterion.

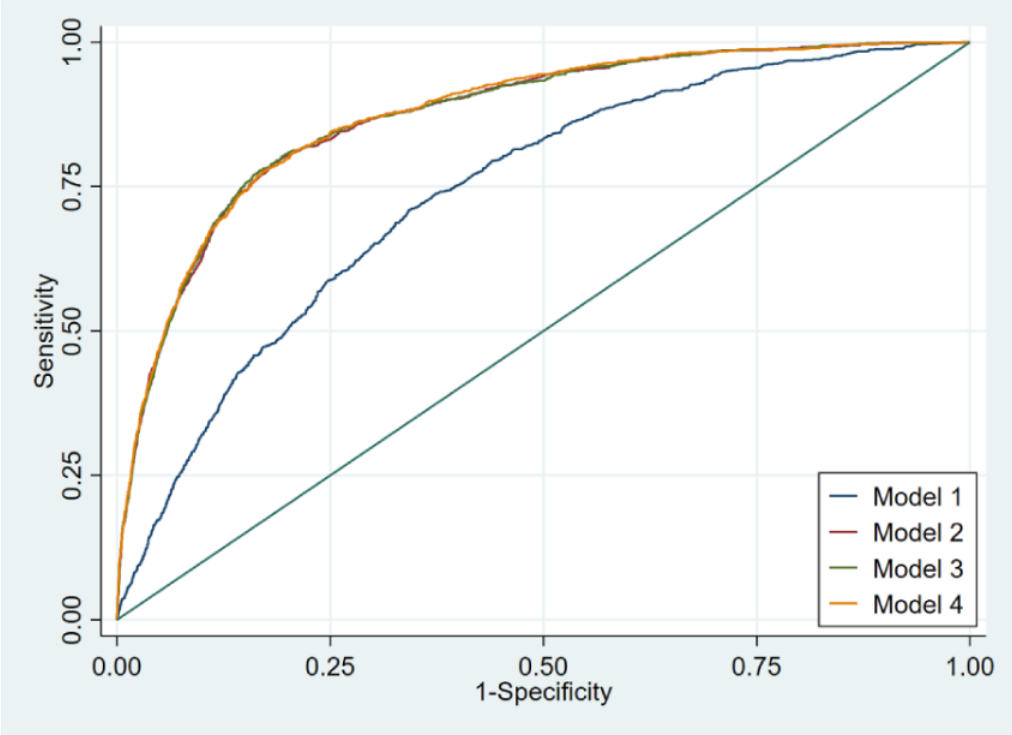
[e]C-Score included six variables: cigarette consumption, alcohol consumption, sleep duration, self-rated health, waist to height ratio, and resting heart rate.

[f]Sociodemographic variables included age, gender, race or ethnicity, marital status, and educational attainment.

[g]Medical history variables were *ever diagnosis of high blood pressure* and *ever diagnosis with hypercholesterolemia*.

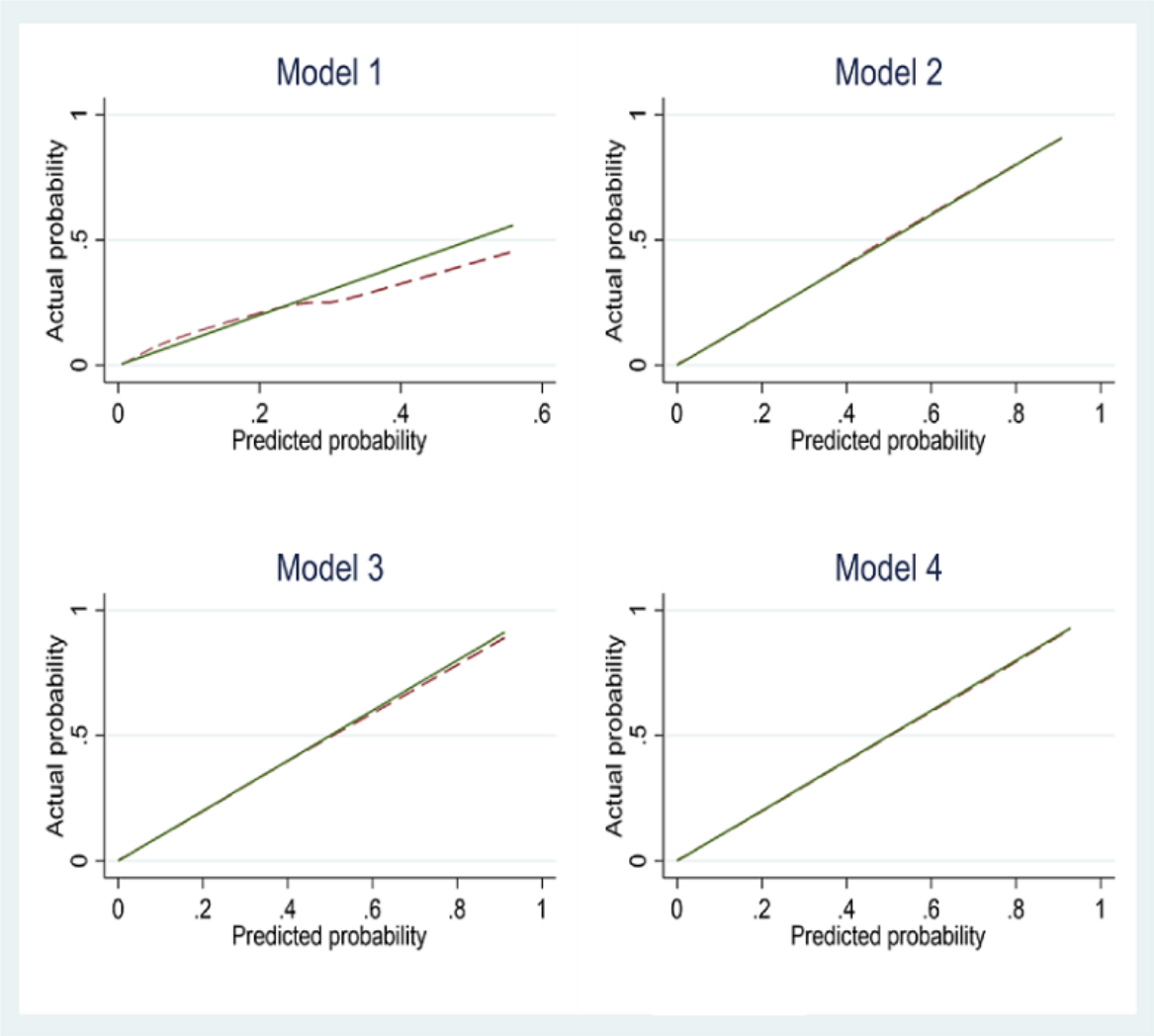
[h]Each sociodemographic variable and medical history variable interacted with the C-Score.

Figure 2. Receiver operating characteristic curve for original C-Score versus expanded models for all-cause mortality. Model 1: C-Score; model 2: C-Score+sociodemographic variables; model 3: C-Score+sociodemographic variables+medical variables; model 4: C-Score+sociodemographic variables+medical history+interactions.



XSL•FO
RenderX

Figure 3. Calibration plots of predicted versus observed probabilities for original C-Score versus expanded models for all-cause mortality. Model 1: C-Score; model 2: C-Score+sociodemographic variables; model 3: C-Score+sociodemographic variables+medical variables; model 4: C-Score+sociodemographic variables+medical history+interactions.

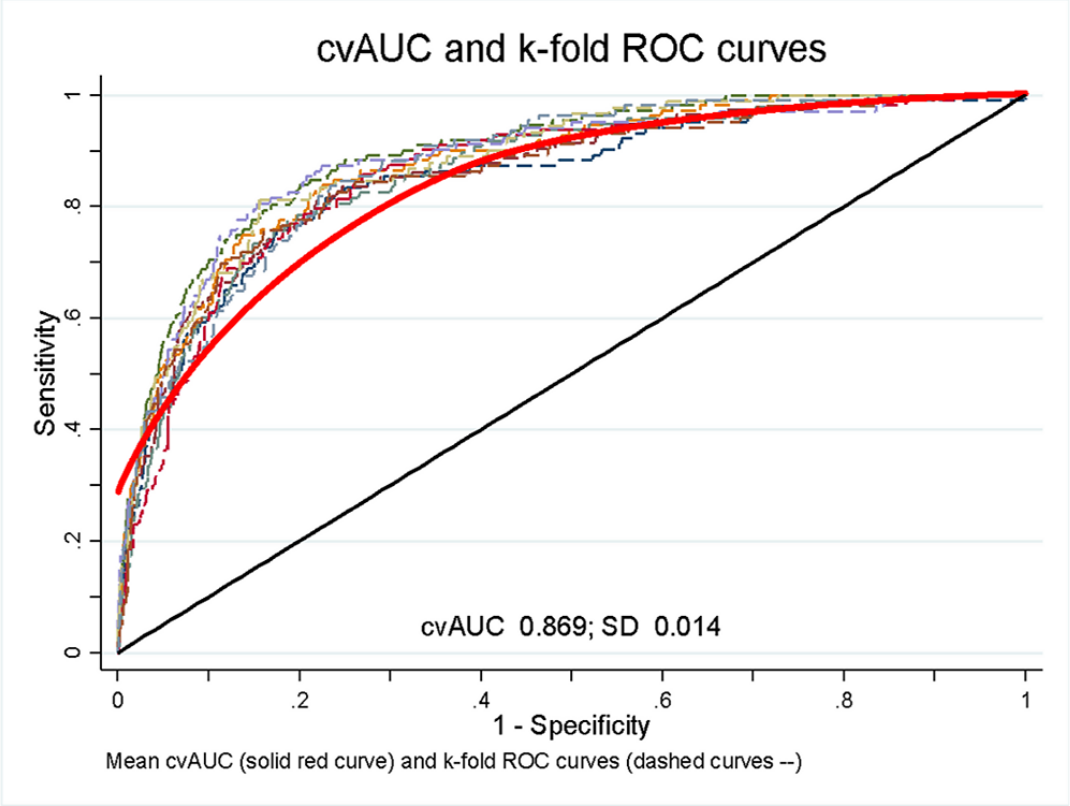


Internal Validation

The validity of our final model (model 3) was assessed using k-fold cross-validation. We used 10 random samples to determine the discrimination capability of the model in predicting the future incidence of all-cause mortality. The AUCs for these random samples ranged from 0.85 to 0.87, showing high consistency in the discrimination of the model (Figure 4). The mean cross-validation AUC was 0.869, indicating a strong capability of the model to discriminate the incidence of all-cause mortality.

XSL•FO
RenderX

Figure 4. Internal validation using k-fold procedure (folds=10). cvAUC: cross-validation area under the curve; ROC: receiver operating characteristic.

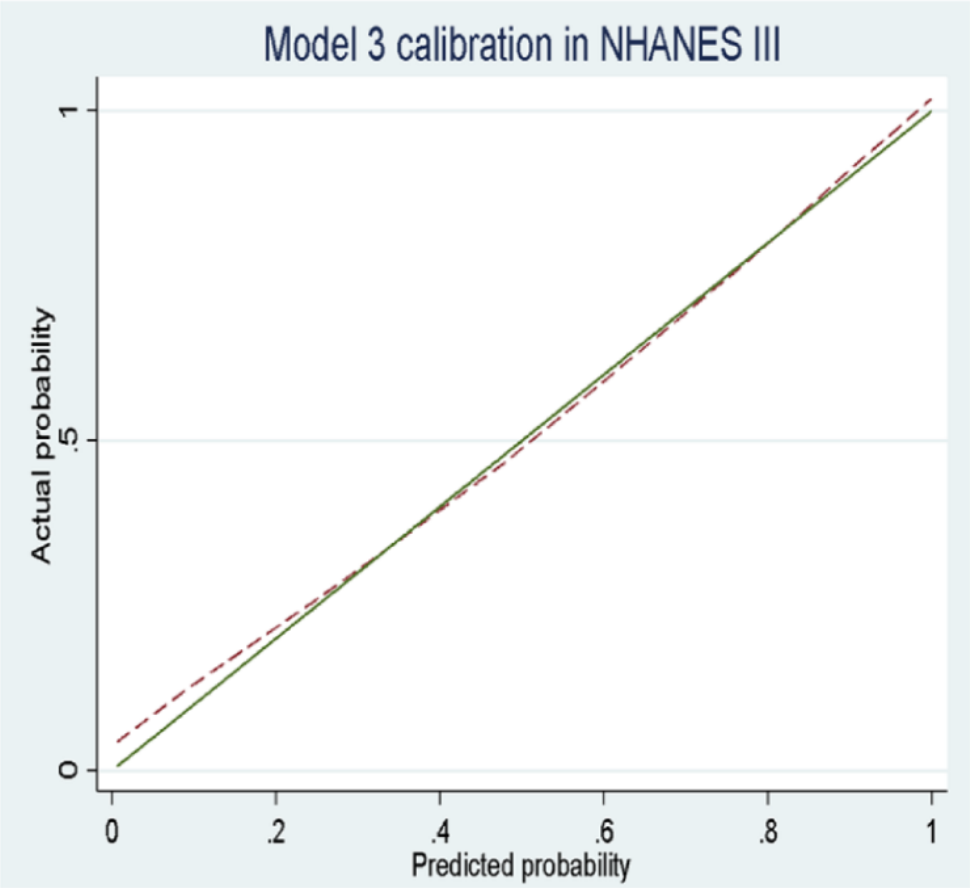


External Validation

The best-performing model (model 3) of the main analysis was used for external validation. Figure 5 shows a calibration plot displaying the predicted versus observed probabilities of all-cause mortality. A comparison between the model performance in the research sample and the external validation sample reveals that the C-score using NHANES 2005-2014 has a good fit with *P*=.06, AUC of 0.87 (95% CI 0.86-0.88) and an Akaike Information criteria of 5705.13. The C-score on the NHANES III survey has a good fit with *P*=.45, AUC of 0.89 (95% CI 0.88-0.90) and an Akaike Information criteria of 3420.19. These results imply that the model performed very well in the external validation sample. It was both well-calibrated and had a high AUC, which is even higher than that identified in the first sample.

XSL•FO
RenderX

Figure 5. Calibration plot of predicted versus observed probabilities of all-cause mortality for model 3 for all-cause mortality. Model 3: C-Score+sociodemographic variables+medical variables. NHANES: National Health and Nutrition Examination Survey.



Discussion

Principal Findings

In this study, we conducted external validation of the C-Score in the US population and expanded the original score to improve its predictive capabilities in the US population.

We found that the C-Score had generally good prediction and calibration capabilities and that it is a promising model that could provide fast and accurate information on all-cause mortality through a digital health app. Our results reveal similar AUCs compared with those found in the United Kingdom by Clift et al [17].

Given the lack of the reaction time variable in the main NHANES sample, we conducted a sensitivity analysis with another survey (NHANES III), which contains the reaction time variable, to assess its marginal effect in predicting all-cause mortality. The results suggest that the absence of the reaction time variable did not meaningfully change the calibration or the discrimination attributes of the assessed model. We believe that the marginal effect is likely to be low as part of the variance explained by the reaction time variable might be captured by other variables in the C-Score.

In addition, we showed that the incorporation of a set of basic sociodemographic and medical history variables greatly boosted the model's predictive performance in the US general population. The AUC for our final model greatly increased from 0.72 (95% CI 0.71-0.73) for the basic C-Score model to 0.87 (95% CI 0.86-0.88) in the expanded model. We further assessed the internal and external validity of the expanded model and found that the model performed equally well in the 10-fold cross-validation sample and the external NHANES III data set.

The incorporation of this model into a user-friendly digital health app can motivate users to predict their current and future health status and take actions to modify their health, thus potentially shaping their future trajectories. Consumer demand for technological innovations that measure health status and predict health outcomes is evidenced by the recent proliferation in the use of commercial wearable technologies, ranging from simple activity or exercise monitors to more sophisticated home-based connected medical devices [4,5]. These devices may function independently or leverage sophisticated back end analytics to analyze user trends and provide feedback [33]. In addition to catering to consumer demand for quick, robust, and user-friendly health assessment, these digital health strategies also engage health care providers by sending client-generated data directly into electronic health records, enabling their integration into care plans [34,35]. The past decade has seen a clear increase in obesity and other chronic diseases worldwide, especially in the US population, where cardiovascular disease, cancer, chronic respiratory illness, and diabetes are leading

XSL•FO
RenderX

causes of death and morbidity [36]. An increasing proportion of adults and children worldwide are overweight or obese, exacerbating the risk of future noncommunicable diseases (NCDs) [37]. The availability of scores that can help individuals reliably estimate current (and potentially future) risk of adverse outcomes could be helpful in interventions to improve individual and, thus, population health in the United States and worldwide. Thus, our validation of the C-Score serves to validate a promising predictive model that can be easily accessed by a lay audience to predict individualized clinical risk and take action to make beneficial lifestyle changes and consequently reduce the risk of future adverse outcomes.

Recent evidence confirms the utility of wearable technology in predicting clinical outcomes with high accuracy [4,38]. Previous studies have capitalized on wearable technologies to provide reliable and accurate measurements of established predictors of mortality and adverse health outcomes [39-42]. For example, Smirnova et al [42] found that wearable technologies provide reproducible and unbiased measures of physical activity, which, in turn, outperform traditional predictors of 5-year mortality among older adults in the US population [42]. The adapted C-Score model had the added strength of using variables that are routinely captured in baseline data collected from users of wearables or inpatient records maintained by health care systems. In addition, such data are more uniformly measured and available across different settings outside the United States and the United Kingdom. Given the overall goal of increasing the generalizability of this score, this is a step in the right direction toward making this a more universally feasible model. Previous models that leveraged complete blood counts and metabolic profiles achieved similar performance (AUC 0.83-0.90) at a presumably much higher cost and logistical complexity [43]. Other studies that integrated a wide range of cognitive, demographic, lifestyle, and clinical factors also achieved similar, if not lower, performance. For example, Ajnakina et al [44] achieved an AUC of 0.74 for all-cause mortality prediction in the general population using 13 prognostic factors. Models that apply increasingly more complex methods such as machine learning are able to slightly improve discrimination, yielding AUCs between 0.78 and 0.79 [45].

Our findings should be viewed in light of some limitations. First, we used a cross-sectional survey that did not follow individuals over time. NHANES is the only survey that is nationally representative of the US general population, which contains most of the variables present in the original C-Score model. The NHANES survey contains 6 out of the 7 variables included in the original UK population-based model, potentially leading to a C-Score that artificially underperforms when predicting all-cause mortality. However, our sensitivity analysis showed that the reaction time variable did not marginally provide additional value to the C-Score in this sample. Even if the subsample in which we tested the reaction time variable did not have the external validity to inform the results of the NHANES subsample, the lack of the reaction time variable would likely lead to an underperforming score, implying that the ability of the score to predict all-cause mortality would be higher, if the reaction time variable had been available in the main NHANES data set. Moreover, although the association between death and other covariates has been investigated using Cox proportional hazards models in other publications, including the original C-Score model [45,46]—we focused on a binary all-cause mortality variable instead of time to death as (1) time to event data was not available, (2) logistic models are easier to communicate to a lay audience, and (3) they avoid the assumptions made by Cox models that may not be met [42]. They have also been shown to perform as well as more complex models [42,47]. Ideally, we would have preferred to use a data set that provides longitudinal estimates; however, we used NHANES, a cross-sectional survey, as it is the only US survey that is nationally representative of the general population and contains the variables present in the original C-Score model (with the exception of reaction time). It also provides a large data set with population-based data.

Conclusions

Limitations notwithstanding, the findings of this validation indicate that the performance of the C-Score is fairly good for predicting all-cause mortality in the US population. The adapted risk score had even better prediction capabilities, as evidenced by the finding that it predicted 87% of the mortality in the US population.

In conclusion, our study findings validate and expand a novel risk-scoring algorithm that can predict the risk of all-cause mortality among adults in the general population with high accuracy and which could be incorporated into a digital health application. The use of high-performing risk scores could be instrumental in clinical counseling, choice of care pathways, and even patient-driven behavior change interventions targeting modifying lifestyles and promoting behavioral change. Despite known effective strategies to reduce NCD-related deaths worldwide, chronic and preventable NCDs continue to drive adult mortality. High-performing risk scores that trigger behavior change could be instrumental in stemming this tide of death and decreased global productivity.

Acknowledgments

This study was supported by Huma, formerly Medopad (award number 133913). The content is the responsibility of the authors and does not necessarily reflect the views of Huma. The funders had no role in the study design, data collection and analysis, decision to publish, or manuscript preparation.

XSL•FO
RenderX

Authors' Contributions

SE conducted the data analysis for both the validation and adaptation phases, contributed to the literature review, contributed to the methodological design, and drafted the manuscript. AIVO contributed to the methodological design, coordinated the project, and drafted the manuscript. DGG contributed to the methodological design and to the drafting of the manuscript. SA contributed to the methodological design and to the drafting of the manuscript. AJT contributed to the methodological design and to the drafting of the manuscript. YZ conducted the data analysis for the validation phase, contributed to the literature review, contributed to the methodological design, and drafted the manuscript. ABL is the principal investigator of the project and contributed to the methodological design and drafting of the manuscript.

Conflicts of Interest

None declared.

References

1. Multiple Chronic Conditions in the United States. Santa Monica, CA: RAND Corporation; 2017.
2. DuGoff E, Canudas-Romo V, Buttorff C, Leff B, Anderson G. Multiple chronic conditions and life expectancy: a life table analysis. Med Care 2014 Aug;52(8):688-694. [doi: 10.1097/MLR.0000000000000166] [Medline: 25023914]
3. Raghupathi W, Raghupathi V. An empirical study of chronic diseases in the United States: a visual analytics approach. Int J Environ Res Public Health 2018 Mar 01;15(3):431 [FREE Full text] [doi: 10.3390/ijerph15030431] [Medline: 29494555]
4. Burnham J, Lu C, Yaeger L, Bailey T, Kollef M. Using wearable technology to predict health outcomes: a literature review. J Am Med Inform Assoc 2018 Sep 01;25(9):1221-1227 [FREE Full text] [doi: 10.1093/jamia/ocy082] [Medline: 29982520]
5. Loncar-Turukalo T, Zdravevski E, Machado da Silva J, Chouvarda I, Trajkovik V. Literature on wearable technology for connected health: scoping review of research trends, advances, and barriers. J Med Internet Res 2019 Sep 05;21(9):e14017 [FREE Full text] [doi: 10.2196/14017] [Medline: 31489843]
6. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. BMJ 2017 May 23;357:j2099 [FREE Full text] [doi: 10.1136/bmj.j2099] [Medline: 28536104]
7. Hong Kong Diabetes Registry, Yang X, So WY, Tong PC, Ma RC, Kong AP, et al. Development and validation of an all-cause mortality risk score in type 2 diabetes. Arch Intern Med 2008 Mar 10;168(5):451-457. [doi: 10.1001/archinte.168.5.451] [Medline: 18332288]
8. Martínez-Díaz AM, Palazón-Bru A, Folgado-de la Rosa DM, Ramírez-Prado D, Navarro-Juan M, Pérez-Ramírez N, et al. A one-year risk score to predict all-cause mortality in hypertensive inpatients. Eur J Intern Med 2019 Jan;59:77-83. [doi: 10.1016/j.ejim.2018.07.010] [Medline: 30007839]
9. Modlin IM, Gustafsson BI, Pavel M, Svejda B, Lawrence B, Kidd M. A nomogram to assess small-intestinal neuroendocrine tumor ('carcinoid') survival. Neuroendocrinology 2010;92(3):143-157 [FREE Full text] [doi: 10.1159/000319784] [Medline: 20733279]
10. Hippisley-Cox J, Coupland C. Development and validation of risk prediction algorithms to estimate future risk of common cancers in men and women: prospective cohort study. BMJ Open 2015 Mar 17;5(3):e007825 [FREE Full text] [doi: 10.1136/bmjopen-2015-007825] [Medline: 25783428]
11. Eagle KA, Lim MJ, Dabbous OH, Pieper KS, Goldberg RJ, Van de Werf F, GRACE Investigators. A validated prediction model for all forms of acute coronary syndrome: estimating the risk of 6-month postdischarge death in an international registry. JAMA 2004 Jun 09;291(22):2727-2733. [doi: 10.1001/jama.291.22.2727] [Medline: 15187054]
12. Ganna A, Ingelsson E. 5 year mortality predictors in 498 103 UK Biobank participants: a prospective population-based study. Lancet 2015 Aug;386(9993):533-540. [doi: 10.1016/s0140-6736(15)60175-1]
13. Yourman L, Lee S, Schonberg M, Widera E, Smith A. Prognostic indices for older adults: a systematic review. JAMA 2012 Jan 11;307(2):182-192 [FREE Full text] [doi: 10.1001/jama.2011.1966] [Medline: 22235089]
14. Lloyd-Jones DM. Cardiovascular risk prediction. Circulation 2010 Apr 20;121(15):1768-1777. [doi: 10.1161/circulationaha.109.849166]
15. Simon GJ, Peterson KA, Castro MR, Steinbach MS, Kumar V, Caraballo PJ. Predicting diabetes clinical outcomes using longitudinal risk factor trajectories. BMC Med Inform Decis Mak 2020 Jan 08;20(1):6 [FREE Full text] [doi: 10.1186/s12911-019-1009-3] [Medline: 31914992]
16. Man B, Schwartz A, Pugach O, Xia Y, Gerber B. A clinical diabetes risk prediction model for prediabetic women with prior gestational diabetes. PLoS One 2021 Jun 25;16(6):e0252501 [FREE Full text] [doi: 10.1371/journal.pone.0252501] [Medline: 34170930]
17. Clift AK, Le Lannou E, Tighe CP, Shah SS, Beatty M, Hyvärinen A, et al. Development and validation of risk scores for all-cause mortality for a smartphone-based "General health score" app: prospective cohort study using the UK Biobank. JMIR Mhealth Uhealth 2021 Feb 16;9(2):e25655 [FREE Full text] [doi: 10.2196/25655] [Medline: 33591285]

XSL•FO
RenderX

18. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. Am J Epidemiol 2017 Nov 01;186(9):1026-1034 [FREE Full text] [doi: 10.1093/aje/kwx246] [Medline: 28641372]
19. Keyes KM, Westreich D. UK Biobank, big data, and the consequences of non-representativeness. Lancet 2019 Mar 30;393(10178):1297 [FREE Full text] [doi: 10.1016/S0140-6736(18)33067-8] [Medline: 30938313]
20. Batty G, Gale C, Kivimäki M, Deary I, Bell S. Comparison of risk factor associations in UK Biobank against representative, general population based studies with conventional response rates: prospective cohort study and individual participant meta-analysis. BMJ 2020 Feb 12;368:m131 [FREE Full text] [doi: 10.1136/bmj.m131] [Medline: 32051121]
21. Zhang N, Yang X, Zhu X, Zhao B, Huang T, Ji Q. Type 2 diabetes mellitus unawareness, prevalence, trends and risk factors: national health and nutrition examination survey (NHANES) 1999-2010. J Int Med Res 2017 Apr;45(2):594-609 [FREE Full text] [doi: 10.1177/0300060517693178] [Medline: 28415936]
22. Gregg EW, Sorlie P, Paulose-Ram R, Gu Q, Eberhardt MS, Wolz M, 1999-2000 national healthnutrition examination survey. Prevalence of lower-extremity disease in the US adult population >=40 years of age with and without diabetes: 1999-2000 national health and nutrition examination survey. Diabetes Care 2004 Jul;27(7):1591-1597. [doi: 10.2337/diacare.27.7.1591] [Medline: 15220233]
23. Linkage methods and analytical support for NCHS linked mortality data. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/nchs/data-linkage/mortality-methods.htm [accessed 2022-05-11]
24. Lewington S, Clarke R, Qizilbash N, Peto R, Collins R, Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. Lancet 2002 Dec 14;360(9349):1903-1913. [doi: 10.1016/s0140-6736(02)11911-8] [Medline: 12493255]
25. Miura K, Daviglus ML, Dyer AR, Liu K, Garside DB, Stamler J, et al. Relationship of blood pressure to 25-year mortality due to coronary heart disease, cardiovascular diseases, and all causes in young adult men: the Chicago Heart Association Detection Project in Industry. Arch Intern Med 2001 Jun 25;161(12):1501-1508. [doi: 10.1001/archinte.161.12.1501] [Medline: 11427097]
26. Prospective Studies Collaboration, Lewington S, Whitlock G, Clarke R, Sherliker P, Emberson J, et al. Blood cholesterol and vascular mortality by age, sex, and blood pressure: a meta-analysis of individual data from 61 prospective studies with 55,000 vascular deaths. Lancet 2007 Dec 01;370(9602):1829-1839. [doi: 10.1016/S0140-6736(07)61778-4] [Medline: 18061058]
27. Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. Stat Med 2014 Feb 10;33(3):517-535 [FREE Full text] [doi: 10.1002/sim.5941] [Medline: 24002997]
28. Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, et al. Discrimination and calibration of clinical prediction models: users' guides to the medical literature. JAMA 2017 Oct 10;318(14):1377-1384. [doi: 10.1001/jama.2017.12126] [Medline: 29049590]
29. Archer KJ, Lemeshow S. Goodness-of-fit test for a logistic regression model fitted using survey sample data. Stata J 2006 Feb 01;6(1):97-105. [doi: 10.1177/1536867x0600600106]
30. Refaeilzadeh P, Tang L, Liu H. Cross-validation. In: Encyclopedia of Database Systems. Boston, MA: Springer; 2009.
31. Everitt B. Scatterplot smoothers. In: Encyclopedia of Statistics in Behavioral Science. Hoboken, New Jersey, United States: Wiley; 2005.
32. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015 Jan 06;162(1):55-63 [FREE Full text] [doi: 10.7326/M14-0697] [Medline: 25560714]
33. Godfrey A, Hetherington V, Shum H, Bonato P, Lovell NH, Stuart S. From A to Z: wearable technology explained. Maturitas 2018 Jul;113:40-47 [FREE Full text] [doi: 10.1016/j.maturitas.2018.04.012] [Medline: 29903647]
34. Quesada-González D, Merkoçi A. Mobile phone-based biosensing: an emerging "diagnostic and communication" technology. Biosens Bioelectron 2017 Jun 15;92:549-562. [doi: 10.1016/j.bios.2016.10.062] [Medline: 27836593]
35. Dinh-Le C, Chuang R, Chokshi S, Mann D. Wearable health technology and electronic health record integration: scoping review and future directions. JMIR Mhealth Uhealth 2019 Sep 11;7(9):e12861 [FREE Full text] [doi: 10.2196/12861] [Medline: 31512582]
36. Noncommunicable Diseases (NCD) country profiles. World Health Organization. URL: http://www.who.int/nmh/publications/ncd-profiles-2018/en/ [accessed 2022-05-11]
37. Overweight. World Health Organization. URL: https://gateway.euro.who.int/en/indicators/h2020_6-overweight/visualizations/#id=1707 [accessed 2022-05-11]
38. Bae S, Dey A, Low C. Using passively collected sedentary behavior to predict hospital readmission. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2016 Presented at: UbiComp '16: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sep 12 - 16, 2016; Heidelberg Germany. [doi: 10.1145/2971648.2971750]
39. Beltrame T, Amelard R, Wong A, Hughson RL. Extracting aerobic system dynamics during unsupervised activities of daily living using wearable sensor machine learning models. J Appl Physiol (1985) 2018 Feb 01;124(2):473-481 [FREE Full text] [doi: 10.1152/japplphysiol.00299.2017] [Medline: 28596271]

XSL•FO
RenderX

40. Rodríguez-Rodríguez I, Chatzigiannakis I, Rodríguez JV, Maranghi M, Gentili M, Zamora-Izquierdo M. Utility of big data in predicting short-term blood glucose levels in type 1 diabetes mellitus through machine learning techniques. Sensors (Basel) 2019 Oct 16;19(20):4482 [FREE Full text] [doi: 10.3390/s19204482] [Medline: 31623111]
41. Sopic D, Aminifar A, Aminifar A, Atienza D. Real-time event-driven classification technique for early detection and prevention of myocardial infarction on wearable systems. IEEE Trans Biomed Circuits Syst 2018 Jul 16 (forthcoming). [doi: 10.1109/TBCAS.2018.2848477] [Medline: 30010598]
42. Smirnova E, Leroux A, Cao Q, Tabacu L, Zipunnikov V, Crainiceanu C, et al. The predictive performance of objective measures of physical activity derived from accelerometry data for 5-year all-cause mortality in older adults: national health and nutritional examination survey 2003-2006. J Gerontol A Biol Sci Med Sci 2020 Sep 16;75(9):1779-1785 [FREE Full text] [doi: 10.1093/gerona/glz193] [Medline: 31504213]
43. Horne BD, May HT, Muhlestein JB, Ronnow BS, Lappé DL, Renlund DG, et al. Exceptional mortality prediction by risk scores from common laboratory tests. Am J Med 2009 Jun;122(6):550-558. [doi: 10.1016/j.amjmed.2008.10.043] [Medline: 19486718]
44. Ajnakina O, Agbedjro D, McCammon R, Faul J, Murray RM, Stahl D, et al. Development and validation of prediction model to estimate 10-year risk of all-cause mortality using modern statistical learning methods: a large population-based cohort study and external validation. BMC Med Res Methodol 2021 Jan 06;21(1):8 [FREE Full text] [doi: 10.1186/s12874-020-01204-7] [Medline: 33407175]
45. Weng SF, Vaz L, Qureshi N, Kai J. Prediction of premature all-cause mortality: a prospective general population cohort study comparing machine-learning and standard epidemiological approaches. PLoS One 2019 Mar 27;14(3):e0214365 [FREE Full text] [doi: 10.1371/journal.pone.0214365] [Medline: 30917171]
46. Seccareccia F, Lanti M, Menotti A, Scanga M. Role of body mass index in the prediction of all cause mortality in over 62,000 men and women. The Italian RIFLE pooling project. Risk factor and life expectancy. J Epidemiol Community Health 1998 Jan;52(1):20-26 [FREE Full text] [doi: 10.1136/jech.52.1.20] [Medline: 9604037]
47. Dinh A, Miertschin S, Young A, Mohanty SD. A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. BMC Med Inform Decis Mak 2019 Nov 06;19(1):211 [FREE Full text] [doi: 10.1186/s12911-019-0918-5] [Medline: 31694707]

Abbreviations

AUC: area under the curve
NCD: noncommunicable disease
NDI: National Death Index
NHANES: National Health and Nutrition Examination Survey
TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis
WtHR: weight to height ratio

Edited by A Mavragani; submitted 25.01.22; peer-reviewed by I Mircheva, B Wong, M Rajendran; comments to author 06.04.22; revised version received 14.04.22; accepted 28.04.22; published 14.06.22.

Please cite as:
Elnakib S, Vecino-Ortiz AI, Gibson DG, Agarwal S, Trujillo AJ, Zhu Y, Labrique AB
A Novel Score for mHealth Apps to Predict and Prevent Mortality: Further Validation and Adaptation to the US Population Using the US National Health and Nutrition Examination Survey Data Set
J Med Internet Res 2022;24(6):e36787
URL: https://www.jmir.org/2022/6/e36787
doi:10.2196/36787
PMID:35483022

©Shatha Elnakib, Andres I Vecino-Ortiz, Dustin G Gibson, Smisha Agarwal, Antonio J Trujillo, Yifan Zhu, Alain B Labrique. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.06.2022. 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

Usability Issues in Evidence-Based Psychosocial Interventions and Implementation Strategies: Cross-project Analysis

Sean A Munson[1,2], PhD; Emily C Friedman[2,3], CPE, MID; Katie Osterhage[2,4], MMS; Ryan Allred[2], BA; Michael D Pullmann[2,3], PhD; Patricia A Areán[2,3], PhD; Aaron R Lyon[2,3], PhD; UW ALACRITY Center Researchers[5]

[1]Department of Human Centered Design & Engineering, University of Washington, Seattle, WA, United States

[2]Advanced Laboratories for Accelerating the Reach and Impact of Treatments for Youth and Adults with Mental Illness (ALACRITY) Center, University of Washington, Seattle, WA, United States

[3]Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, United States

[4]Department of Family Medicine, University of Washington, Seattle, WA, United States

[5]See Acknowledgments

Corresponding Author:
Sean A Munson, PhD
Department of Human Centered Design & Engineering
University of Washington
Campus Box 352315
Seattle, WA, 98195
United States
Phone: 1 206 221 7699
Email: smunson@uw.edu

Abstract

Background: People often prefer evidence-based psychosocial interventions (EBPIs) for mental health care; however, these interventions frequently remain unavailable to people in nonspecialty or integrated settings, such as primary care and schools. Previous research has suggested that usability, a concept from human-centered design, could support an understanding of the barriers to and facilitators of the successful adoption of EBPIs and support the redesign of EBPIs and implementation strategies.

Objective: This study aimed to identify and categorize usability issues in EBPIs and their implementation strategies.

Methods: We adapted a usability issue analysis and reporting format from a human-centered design. A total of 13 projects supported by the National Institute of Mental Health—funded Accelerating the Reach and Impact of Treatments for Youth and Adults with Mental Illness Center at the University of Washington used this format to describe usability issues for EBPIs and implementation strategies with which they were working. Center researchers used iterative affinity diagramming and coding processes to identify usability issue categories. On the basis of these categories and the underlying issues, we propose heuristics for the design or redesign of EBPIs and implementation strategies.

Results: The 13 projects reported a total of 90 usability issues, which we categorized into 12 categories, including complex and/or cognitively overwhelming, required time exceeding available time, incompatibility with interventionist preference or practice, incompatibility with existing workflow, insufficient customization to clients/recipients, intervention buy-in (value), interventionist buy-in (trust), overreliance on technology, requires unavailable infrastructure, inadequate scaffolding for client/recipient, inadequate training and scaffolding for interventionists, and lack of support for necessary communication. These issues range from minor inconveniences that affect a few interventionists or recipients to severe issues that prevent all interventionists or recipients in a setting from completing part or all of the intervention. We propose 12 corresponding heuristics to guide EBPIs and implementation strategy designers in preventing and addressing these usability issues.

Conclusions: Usability issues were prevalent in the studied EBPIs and implementation strategies. We recommend using the lens of usability evaluation to understand and address barriers to the effective use and reach of EBPIs and implementation strategies.

International Registered Report Identifier (IRRID): RR2-10.2196/14990

(J Med Internet Res 2022;24(6):e37585) doi:10.2196/37585

XSL•FO
RenderX

KEYWORDS

evidence-based psychosocial interventions; usability; implementation strategies; mental health; human-centered design; implementation science

Introduction

Background

Many people seeking care for mental health problems prefer psychosocial interventions [1-5]. Evidence-based psychosocial interventions (EBPIs) have been shown to be effective in numerous studies; however, despite these preferences and evidence supporting their use, EBPIs remain unavailable to people in most service settings, especially nonspecialty or integrated settings, such as primary care [6] and schools [7,8], where most mental health care is delivered [9].

Non–mental health settings differ from mental health service contexts for which EBPIs are typically developed. The differences in service settings often contribute to a poor contextual fit, low rates of adoption, and sustained use [10,11]. Furthermore, most EBPIs are complex interventions that require ongoing support to ensure their quality. Clinicians face difficulties in learning and adopting these new practices, which further limits EBPI availability [12]. Despite decades of research intended to address these and other barriers, the many implementation strategies that target these barriers are often cumbersome and costly processes to deliver [13,14], resulting in few cost-effective strategies [15] and a lingering science-to-service gap.

Within the University of Washington (UW) Advanced Laboratory for Accelerating the Reach and Impact of Treatments for Youth and Adults with Mental Illness (ALACRITY) Center [16], we believe that new approaches are required to address these barriers. We set out to apply a core concept from human-centered design (HCD) to understand and improve EBPIs: *usability*. Usability is the degree to which a program can be used easily, efficiently, and with satisfaction/low user burden by a particular end user [17]. The concept of usability and techniques for assessing the usability of a system and then ideating, designing, implementing, and evaluating usability fixes has been central to the widespread success of modern digital tools, and we argue that usability metrics and assessment methods are broadly relevant.

Drawing on the results from 13 projects in the Center, we identified 13 categories of usability issues in EBPIs and their implementation strategies. We propose heuristics to prevent or mitigate such issues in the future design of EBPIs and their implementation strategies. We also present our development and use of the usability issue format as a resource for researchers.

Overview of EBPIs

EBPIs are complex health interventions involving interpersonal or informational activities and techniques [12]. In mental and behavioral health, EBPIs target biological, behavioral, cognitive, emotional, interpersonal, social, or environmental factors to reduce the symptoms of these disorders and improve functioning or well-being. Examples of EBPIs to address mental health concerns include parent training protocols to address children's disruptive behavior problems [18] and cognitive behavioral therapies for adults with anxiety or depression [19].

Over the past few decades, hundreds of EBPIs have been developed for use with youth and adults; however, their adoption, high-fidelity delivery, and sustained use in routine service delivery remain low [20-22]. This is because of a variety of multilevel barriers and facilitators (ie, determinants), ranging from the characteristics of the interventions themselves to broader social, political, and policy influences [23]. Recognizing these struggles, researchers and practitioners have increasingly worked to develop *implementation strategies* such as methods or techniques used to enhance the adoption, implementation, and sustainment of a clinical program or practice by addressing determinants [14,24]. Implementation strategies are complex interventions that are largely socially mediated and interpersonal or informational [14,25]. More than 70 unique implementation strategies have been developed for use across a variety of health care delivery settings [26,27]. Examples of strategies include informational meetings or training, audits, and feedback processes.

The consistent and effective use of both EBPIs and implementation strategies has been suboptimal [18,28], and additional approaches are needed to improve practitioner and service recipient engagement with these innovations. Both EBPIs and strategies have a common disconnect between the settings in which they were developed, which are most often academic medical centers or other research settings, and the real-world contexts in which they may ultimately be deployed [29-31]. Few contemporary approaches exist that highlight and address the design quality of both EBPIs and implementation strategies [16].

HCD and Usability

HCD (or the closely related field of user-centered design) is a field that has produced methods of developing compelling, intuitive, easily adopted, and engaging products, services, and tools [32]. HCD has appropriated, adapted, and developed methods for systematically understanding users and other interested or affected parties and contexts, ideating and designing innovations to address needs, opportunities, and problems, and then evaluating the resulting innovations. As a field, HCD excels in improving innovation-user and innovation-context fit.

Usability—the extent to which a system or service can be used by specific people to achieve specified goals with effectiveness, efficiency, and satisfaction within a specified context of use (ISO 9241-11:2018) [17]—is a core concept at all stages of the HCD process. Understanding what makes existing solutions unusable for some people, or what would make an experience more usable, can drive problem identification and innovation. As innovations move forward in the design process, they must be evaluated for usability at each stage, or the resulting

XSL•FO
RenderX

innovation may have limited adoption or harmful or even fatal errors.

Several approaches have been developed to ensure the usability of designed systems. This includes usability evaluation (or usability testing), in which users interact with prototypes or implemented systems in scenarios or in open-ended use. This use could occur in the laboratory or field. Evaluations often begin before target users are directly engaged. Designers are trained in common usability issues and heuristics, or guidelines, that help prevent these issues. Designers and user experience professionals may use these guidelines in the formal heuristic evaluation of prototypes by walking through parts of the interface to identify and address potential issues. Outside formal heuristic evaluation, most designers have internalized common heuristics [33], which helps pre-empt many potential usability issues as they work. In cognitive walk-throughs [34], which is another form of early-stage evaluation, domain experts walked through the proposed interface, often guided by specific scenarios, to identify potential usability issues.

Although these techniques were developed primarily in the design and evaluation of digital systems (or hybrid digital-physical systems), including in mental health [35], they are also promising for understanding what facilitates and inhibits the successful implementation, adoption, and use of EBPIs. Recent studies have applied HCD to EBPIs to enhance usability, decrease burden, and increase contextual appropriateness [1,13,25]. This research has also shown how HCD evaluation approaches, such as cognitive walk-throughs, can be adapted to identify usability barriers in implementation strategies [25].

In this study, we sought to apply usability evaluation methods to identify usability issues across a range of EBPIs and their implementations. We propose that this research can help to understand the ways in which EBPI design and implementation create barriers to their use by patients, clinicians, and other primary and secondary users and that such an understanding can led to the development of heuristics that improve future EBPI design and implementation.

Specifically, we asked the following: (1) what are the common usability issues in EBPIs and implementation strategies for mental and behavioral health, and (2) what are the characteristics of the highest-severity usability issues?

In this paper, we present the results and reflections of our investigation on the development and use of a usability issue–reporting format for mental and behavioral health interventions.

Methods

Overview

Drawing from 13 projects examining the usability of EBPIs, such as behavioral activation (BA) and problem-solving therapy (PST), and their implementation strategies, project teams and Center researchers identified 90 usability issues. We clustered these into 13 categories of issues. From these clusters, we developed heuristics that we propose can be used by EBPI designers and implementers to guide the design, redesign, and implementation of EBPIs across a range of accessible community settings.

Discover, Design/Build, and Test Process and Data Sources

Overview

The UW ALACRITY Center organized the work in these projects according to the Discover, Design/Build, and Test (DDBT) model, which draws from both implementation science (IS) and HCD [16]. The DDBT model is a phased approach to intervention and strategy redesign that uses methods from HCD and is informed by the Consolidated Framework for Implementation Research model, including the Consolidated Framework for Implementation Research intervention, individual, inner setting, and process constructs [23]. In the Discover phase, redesign teams, comprising EBPI experts and HCD researchers, worked to identify the usability challenges they experienced when interacting with the EBPI. Strategies for usability identification included but were not limited to focus groups, in situ use of the EBPI, cognitive walk-throughs, interviews, and contextual observations. Solutions for the usability challenges were identified in the Design/Build phase, whereby redesign teams worked together to create initial prototype solutions that users iteratively used and redesigned.

Data Sources

Data were provided by investigators on 13 projects: 2 pilot projects and 11 seed projects funded through an open call at the UW ALACRITY Center. These studies focused on (1) improving the implementation or usability of an existing EBPI and (2) ≥1 key issue of interest to the Center (clinician capacity, usability, and sustained quality). Each project specified a conceptual model with one or more hypothesized mechanisms that improved the clinical or implementation outcomes. All proposals were peer reviewed and assigned a mentor from the ALACRITY Center's core team of investigators. The projects and processes by which project teams described and reported usability issues are explained in the following sections.

DDBT Procedures

Although we standardized the format for describing usability issues, we did not prescribe specific methods for each project team to use in the stages of DDBT. Consequently, teams used a range of methods, including interviews, focus groups, cognitive walk-throughs [34], behavioral rehearsals [36], asynchronous remote communities [37], and various other approaches to understanding the usability of EBPIs and implementation strategies [38]. Different methods are appropriate for different groups (eg, for people with barriers to accessing technology, the asynchronous remote community method, which requires ongoing engagement through technology, would hinder participation, although it was ideal for long-term engagement with a group that regularly uses social media to keep in touch [37]). The Center provided teams with support from Center investigators, mentors, and professional staff for the selection and application of HCD and implementation of scientific methods in their work. We also recommended the use of the Intervention Usability Scale [39], System Usability Scale [40], User Burden Scale [41],

XSL•FO
RenderX

Implementation Strategy Usability Scale [25], Acceptability of Intervention Measure, Feasibility of Intervention Measure, and Intervention Appropriateness Measure [42] to assess the usability and implement ability of an intervention.

We also did not prescribe the design *fidelity* of the interventions and implementation strategies to be investigated. Design fidelity (not to be confused with treatment fidelity) refers to the details, completeness, and functionality of a prototype. In design, lower-fidelity prototypes can support exploring several different design options at a lower cost. They are valuable for identifying major incompatibilities between design and user needs. In the projects in our study, low-fidelity prototypes included storyboards, mock-ups of digital and/or paper tools intended to support a treatment approach, or even just textual scenarios describing potential changes. Other teams used mock-ups of digital or other artifacts designed to support the implementation of an EBPI or full interactive prototypes of digital services or training designed to support the implementation of an EBPI. However, other teams reviewed the recordings or transcripts of therapy sessions and interviewed clients using existing therapy in practice.

Overview of the Projects

The 13 projects focused on EBPIs and implementation strategies (Table 1). Among the studied EBPIs, approximately half examined BA [43-45] (3/13, 23% of projects), PST [46] (2/13, 15% of projects), or another EBPI based on one of these (1/13, 8% of projects). These interventions were accessible in many ways: PST and BA are used in nonspecialty settings such as primary care, and our research partners (eg, Seattle Children's Hospital) also often use BA.

XSL•FO
RenderX

Table 1. Our data set included usability issues reported by 13 University of Washington Accelerating the Reach and Impact of Treatments for Youth and Adults with Mental Illness Center projects (N=90 issues).

Project	Setting	EBPI[a]	Implementation strategy	Methods	Issues reported, n
Task sharing with BA[b] (R34, Areán and Gonzalez)	Rural primary care clinics	BA	Task sharing: shifting more tasks from therapist to care manager to more efficiently implement BA	• Qualitative interviews with therapists and care managers during the "Discover" phase	6
PST[c] support tool (R34, Bennett, Raue, and Munson)	Primary care clinics	PST	PST aid: a web-based tool designed to support the use of PST	• Observations and qualitative interviews with clinicians during the "Discover" phase	6
Designing and evaluating an asynchronous remote communities approach to behavioral activation with clinicians and adolescents at risk for depression (R03, Jenness and Kientz)	A hospital or large urban health system	BA	Asynchronous remote communities: offer peer, automated, and clinician support between sessions	• Discover: 2 asynchronous remote community studies [47] • Design and build: iterative design, build, and usability evaluation of interactive prototype • Test: a pilot study, collecting data on feasibility, usability, user burden, acceptability, and symptom outcomes [48]	3
Using human-centered design for technology-enabled behavioral treatment of depression in urban and rural cancer centers (R03, Hsieh and Bauer)	Urban and rural cancer centers delivering collaborative care	BA	N/A[d]	• Discover: interviews with 29 stakeholders across 3 groups • Design: parallel journeys framework as a conceptual design framework	11
Discovering the capacity of primary care frontline staff to deliver a low-intensity technology-enhanced intervention to treat Geriatric depression (R03, Renn and Zaslavsky)	Primary care	Mobile motivational physical activity–targeted intervention (based on BA)	Task sharing: implementation using frontline primary care staff such as nurses and medical assistants	• Discover: focus groups and interviews with 24 stakeholders • Design/build: halted because of the COVID-19 pandemic	2
mHealth[e] in West Africa: developing an evidence-based psychosocial intervention toolkit (R03, Ben-Zeev and Snyder)	Ghanaian prayer camps	Multiple digitally delivered components of evidence-based interventions for psychosis	N/A	• Discover: observations and qualitative interviews with 18 healers [49] • Design: co-design sessions with 12 healers • Build: prototype • Test: usability testing with 12 healers	3
Iterative redesign of a behavioral skills training program for use in educational settings (R03, Bearss and Locke)	Elementary school special education classrooms	The RUBI[f] protocol	N/A	• Discover: demonstration study of RUBI with mixed methods feedback [50] • Design: collaborative redesign feedback sessions; demonstration study of revised RUBI in Educational Settings with mixed methods feedback	4

XSL•FO
RenderX

Project	Setting	EBPI[a]	Implementation strategy	Methods	Issues reported, n
Increasing the usability and cultural relevance of an EBPI for suicidality in schools (R03, Brewer and Jones)	High schools	CAMS[g]	N/A	• Discover: contextual observations and qualitative interviews with school-based clinicians and focus groups with high school students [51] • Design/build: usability testing of unadapted CAMS SSF[h] with school-based clinicians, followed by a co-design session with school-based clinicians and usability testing of the adapted CAMS SSF with school-based clinicians	4
Improving the usability of decision support for PTSD[i] in primary care (R03, Chen and Williams)	Primary care	Prolonged exposure; cognitive processing therapy	Veterans Affairs SDM[j] protocol	• Discover: HCD approach interviews with 22 clinicians and 25 patients with PTSD • Design/build: usability testing of iterative adaptations of an electronic health record template for conducting SDM with primary care–based mental health clinicians	12
Modification of a parenting intervention for primary care–based delivery to women with perinatal depression and anxiety: PFR[k] (R03, Bhat and Oxford)	Primary care and prenatal clinics	PFR	N/A	• Discover: focus group with users of PFR to identify PFR features to be modified • Design/build: iterative design of the PFR-Brief protocol in collaboration with an end user participatory design group alternating with consumer feedback in microtrials	3
Supporting iterative design of homework in problem solving therapy (R03, Agapie and Areán)	Individual therapy sessions with older adults in an urban setting	PST	N/A	• Discover: interviews with patients and analysis of recordings of sessions [52].	20
Improving usability of a comprehensive self-management intervention to address anxiety and depression among persons with irritable bowel syndrome (R03, Kamp and Levy)	Primary care and gastroenterology clinics across the Pacific Northwest	Comprehensive self-management intervention for irritable bowel syndrome	N/A	• Discover: interviews with 12 patients and 14 health care providers • Design/build: usability testing of an interactive digital prototype	5
Iterative (re)design of a virtual postpartum depression intervention with Latina mothers (R03, Gonzalez and Ramirez)	Digital space (Ginger.io)	Mothers and Babies Program	N/A	• Discover: surveys and qualitative interviews with Latina mothers in the postpartum period • Design: prototype of a web-based mental health platform	4

[a]EBPI: evidence-based psychosocial intervention.

[b]BA: behavioral activation.

[c]PST: problem-solving therapy.

[d]N/A: not applicable.

[e]mHealth: mobile health.

[f]RUBI: Research Units in Behavioral Intervention.

[g]CAMS: Collaborative Assessment and Management of Suicidality.

[h]SSF: Suicide Status Form.

[i]PTSD: posttraumatic stress disorder.

XSL•FO
RenderX

[j]SDM: shared decision-making.

[k]PFR: Promoting First Relationships.

A total of 6 projects also recommended redesigns based on technology. This reflects a few factors internal and external to the Center. Internally, researchers from the Paul G Allen School of Computer Science and Engineering and the Department of Human Centered Design and Engineering—2 units extensively focused on the study and design of sociotechnical systems—were central to our team. Externally, technology is often an appealing way of increasing the reach of interventions (although, as discussed in the *Results* section, this can sometimes be too optimistic a view [53]). This became even more important as many projects were operational during the COVID-19 pandemic, making telehealth and remote care more common.

Finally, the overarching Center mission—the identification of usability challenges in underresourced, non–mental health settings—informed the selection of the projects included in this study. Teams were led by investigators from different disciplines, many of whom were from HCD and technology design fields; as a result, solutions identified may have been informed by that discipline. Investigators proposed projects with the intent of uncovering usability challenges where they saw opportunities for improvement and redesign of interventions or implementation strategies rather than focusing on the identification of intervention or strategy features that were usable. The Center's focus on nonspecialty settings also influenced which issues were identified: many of the EBPIs and implementation strategies studied may work in specialty mental health settings but become unusable in the nonspecialty settings where most people access care.

Adaptation of Usability Issue Concept and Reporting Format

We adapted a common usability issue–reporting format drawn from HCD and human-computer interaction. This template was adapted by the center's methods and design team and solicited the following information:

- Descriptive title of the issue or problem
- Severity of the problem, rated from catastrophic to subtle on a scale adapted from a study by Dumas and Reddish [54]; this resulted in a 5-point scale, where L0=catastrophic, risks causing harm; L1=prevents completion of task; L2=causes significant delay and/or frustration; L3=minor effect on usability; and L4=subtle, possible future improvement
- Scope of the problem: who is affected (all users or some) and how much (eg, every session or just when starting and a particular module in an intervention or widespread)
- Complexity of the issue: how straightforward it would be to address the issue, which includes both how well-understood the problem is and how much it interacts with different components of the EBPI or implementation strategy
- Evidence for the issue: includes both qualitative and quantitative evidence (as available) to support the reader in understanding the problem, who it affects, and its consequences

If the teams knew of related research, we encouraged them to reference it. We also encouraged them to describe any next steps or known solutions. Our initial format drew on SM's experience in teaching usability and EF's expertise as a user experience professional.

Development of the Usability Issue Format

We piloted the usability issue format with 2 of the smaller-scale projects, asking investigators to report issues from their data using a survey and accompanying guidance on usability issue reporting. We then examined these preliminary issues to clarify guidance and added examples in a survey that we would use with all teams. Changes made at this stage included describing problems with the intervention or implementation strategy rather than with users and revising to be as specific as possible in describing the affected components of the intervention and/or implementation strategy and the consequences. This also led us to suggest describing problems in the following format: *When [PRECURSOR], the [COMPONENT] is / has / is experienced as / results in / etc. [PROBLEM] which [CONSEQUENCE].* Multimedia Appendix 1 [33,54,55] presents the final usability issue survey and guidance.

Usability Issue Identification and Reporting

We introduced the usability issue survey and guidance at a workshop for all Center investigators. During the workshop, we presented example issues and revisions that would make them more precise and attribute faults to the EBPI or implementation strategy rather than to interventionists or recipients. We encouraged investigators to apply the survey to any preliminary results they had and to ask questions.

Following the workshop, we asked the investigators to use the survey to report the issues they identified in their projects. We encouraged them to complete the survey primarily after the Discover phase—the phase during which they worked to understand the current EBPI or implementation strategy's use in its destination context—and to complete it again with any additional issues identified in the iterative design and build phases. Teams emailed the completed surveys to the Center investigators.

As individual project teams analyzed data from their projects to complete the survey, some were asked whether to report an issue according to our guidance as follows:

> *As a center, we are interested in usability issues with the interventions or implementation strategies [e.g., Problem Solving Therapy is fatiguing to do all day], not more run of the mill usability issues with the way it is delivered [e.g., the button should be bigger; the handout's colors clash] unless these issues significantly interfere with a user's ability to accomplish the core tasks of the intervention.*

The criteria for issues of interest were difficult to discern in some projects. For example, when a worksheet used during an intervention did not provide enough space for recipients to respond to a question, Center and project researchers determined

XSL•FO
RenderX

that this should not be reported through consensus. However, when the same worksheet consistently did not support the key elements of an intervention, the team determined that this was an important finding to report.

For many investigators, this project was their first exposure to the concept of usability and reporting on usability issues; therefore, we found it necessary for Center researchers (AL, SM, and EF) to work iteratively and collaboratively with teams to refine usability issues before finalizing them. Common challenges included framing usability issues in a way that blamed the interventionist or recipient (eg, their lack of training or time rather than the EBPI or strategy requiring more training or timing than someone had), writing compound usability issues that could be separated, and describing the potential solution rather than the issue that necessitated a solution.

Analysis of Usability Issues From Projects: Affinity Diagramming Process

Using an affinity diagramming process [56], we coded the usability issues reported by each project to identify clusters of usability issues. We recorded 72 issues from 11 projects as sticky notes in a digital, collaborative space. We then created 3 copies, which 3 researchers (EF, RA, and KO) each used to independently group issues across projects and label the resulting groups. Throughout this process, researchers worked to identify themes that inhibited the usability of EBPIs and implementation strategies rather than bucketing issues according to specific parts of therapy, implementation, or particular artifacts.

The coding process comprised 3 rounds of review, starting with the initial coding and discussion among the coders (ECF, RA, and KO) to compare preliminary categories and build a consensus [57]. This resulted in 10 preliminary categories and corresponding definitions. The broader research team met and discussed the identified codes and categories. In a few instances, this meant returning to project teams to request additional clarification about the usability issue. ECF, RA, and KO incorporated this feedback in the second round of coding while still on their individual boards. In this round, coders also considered subcategories and any potentially missing high-level categories, resulting in a revised list of 12 categories.

The coders then individually recategorized usability issues into these 12 categories. Disagreements in the resulting categorization were resolved through discussion by coders, resulting in consensus coding, with a few remaining disagreements that were deliberated with the broader research team. To address some issues, the coders asked for further clarifications from the project teams. Through this process, we came to see 7 reported issues as *compound issues*; that is, they were multiple issues, and thus, we broke them apart. As we completed this coding, 2 project teams reported an additional 11 issues that fit within the existing categories. The final list included 90 issues.

After reviewing the final codes, the researchers suggested that each category of issues might reflect one or more design heuristics. On the basis of this discussion, the coders examined whether and how these categories were mapped to 3 sets of heuristics: the classic 10 usability heuristics by Nielsen [33], the 15 principles for good service design [58], and initial heuristics for implementable EBPIs [13]. The research team developed draft heuristics based on the sets and categories that did not seem to be covered by existing heuristics. The research team continued to refine the heuristics as we developed this manuscript.

Ethics Approval

Individual projects were reviewed or determined exempt by the University of Washington (study numbers 2824, 3795, 4236, 4274, 6044, 6748, 6839, 7463, 7735, 7754, 7853, and 9463) and Seattle Children's (study number 1890). Each project's protocol documents included details about sharing data with the Center for secondary analysis.

Results

Overview

We identified 12 categories of usability issues in the EBPIs and implementation strategies, as summarized in Textbox 1.

XSL•FO
RenderX

Textbox 1. Our analysis identified 12 categories of usability issues.

Complex and/or cognitively overwhelming: The intervention or implementation strategy is too overwhelming to the user or the interventionist.

Required time exceeds the available time: The intervention or implementation strategy demands more time than is available.

Incompatibility with interventionist preference or practice: The intervention or implementation strategy is not compatible with how the interventionist prefers—or has been trained—to work and deliver interventions.

Incompatibility with existing workflow: The intervention or implementation strategy is not compatible with the interventionists' existing workflows.

Insufficient customization to clients or recipients: The intervention or implementation strategy cannot be tailored to client/recipient needs or does not provide enough guidance for interventionists and clients/recipients to customize it.

Intervention buy-in (value): Intervention or implementation strategy does not sufficiently build client/recipient buy-in for its value.

Interventionist buy-in (trust): The intervention or implementation strategy does not build the client's/recipient's trust in the interventionist.

Overreliance on technology: Intervention or implementation strategy relies on technology that creates barriers for some clinicians or recipients or that is not available to all clients or recipients.

Requires unavailable infrastructure: Intervention or implementation strategy requires physical, systemic, or organizational infrastructures that are not available.

Inadequate scaffolding for client/recipient: This involves a lack of preparation and support for the client/recipient. The intervention or implementation strategy lacks support for the client/recipient to understand and succeed in the required activities of the intervention.

Inadequate training and scaffolding for interventionists: The intervention or implementation strategy's training and scaffolding do not provide enough initial and/or ongoing support to deliver the invention as designed or to know how to respond to emergent challenges.

Lack of support for necessary communication: The intervention or implementation strategy requires but does not sufficiently facilitate communication between interventionist and client/recipient.

Complex or Cognitively Overwhelming

Too complex or cognitively overwhelming EBPIs or implementation strategies were the most common types of usability issues uncovered in all the projects surveyed, accounting for 12 of 90 issues submitted to the Center. In these issues, the complexity of the intervention or implementation strategy exceeded the interventionist and/or client's ability to manage it, creating an excessive cognitive or emotional burden.

An example of this came up in PST, where therapists and clients worked together to complete a worksheet that began by asking them to clearly identify the problem the client hoped to address that week and then their goal. If the problem identification step resulted in problems that were too complex, clients and therapists could be unable to complete the next steps. Complex problems could take more time to discuss than is afforded in a session, leading to running out of time. The selection of a complex problem could also overwhelm clients, who then become stuck in the next steps, with interventionists unsure of how to best support them.

In another project that focused on a digital decision support tool to support clients with shared decision-making (SDM) and to support the initiation of and adherence to EBPIs for posttraumatic stress disorder within primary care, clinicians and clients found the presented information to be overwhelming. They reported that the tool contained too much information, too many words, charts that were too busy, and insufficient visual or audio support for interpreting this information.

Required Time Exceeds Available Time

Of the 90 issues, 11 focused on EBPIs or implementation strategies that exceeded the time available for their delivery. Although *time* is the most commonly identified implementation barrier in the literature [26], *time* here refers specifically to excessive time demands of the intervention or implementation strategy.

This category of issues frequently occurs in projects adapting an intervention to a new setting, such as primary care, where the providers have less time with the recipient than what was originally designed. For example, although a comprehensive self-management intervention addressing anxiety, depression, and gastrointestinal symptoms among individuals with irritable bowel syndrome had previously been successful when delivered by clinicians in a research setting, delivery typically took longer than the time allotted for it in a primary care visit (<10 minutes). As a result, primary care clinicians skipped or rushed the content or abandoned the intervention altogether.

Interventions also asked too much of the interventionists or recipients between sessions. For example, one project adapted the Research Units in Behavioral Intervention program, an evidence-based parent-mediated intervention that improves disruptive behavior in children with autism, for use by educators within schools [50]. However, individualized visual support, which is a core component of the Research Units in Behavioral Intervention program, requires more time to develop than educators could allocate. Generating the content required creating the pictures, developing the visual in a computer program, printing and laminating it, cutting, and velcroing to finalize the visual. Teachers do not have the time to create these materials alongside their other responsibilities, reducing their likelihood of using this core component of the intervention.

Insufficient Customization to Clients

The reported issues also highlighted the need for interventions to be adaptable and accessible to different client profiles (eg, age, race, learning styles, and education levels) and provide interventionists with guidance for tailoring to individual preferences and needs. The lack of customization was an issue

XSL•FO
RenderX

in a project examining Collaborative Assessment and Management of Suicide, a suicide-specific intervention that enables clinicians to quickly assess and treat suicidality. This intervention is extremely trainable and resource efficient, making it appealing for use in high schools. Although initially designed for use with adults, the Collaborative Assessment and Management of Suicide sometimes used words/phrases that were too difficult or vague for high school students (eg, anguish, agitation, and significant loss). This resulted in low comprehension, less engagement, and uncertainty regarding how to answer the questions.

A project designing and evaluating asynchronous, remote support for BA with adolescents at risk for depression also identified a need for customization. Clinicians reported personalizing the manualized BA prompts that appeared on worksheets for each teenager in their current use of BA and that building them into digital tools prevented this personalization. This inflexibility created barriers to teenagers in applying the concepts to their own lives and situations.

Incompatibility With Interventionist Preference or Practice

Usability issues also occur when an intervention or implementation strategy is not compatible with how the interventionist prefers—or has been trained—to do things. For example, some interventionists perceive the processes for PST and engagement as rigid, repetitive, or frustrating (eg, setting the agenda weekly feels repetitive). Consequently, these interventionists felt that this repetitive work was not stimulating and sometimes adapted the therapy process to skip it, reducing intervention fidelity.

Therapists also sometimes end up leading the selection of the problems, goals, and solutions when a client faces difficulty with a step in these manualized interventions, as they believe that this is a good use of their expertise to keep the process moving. However, this is a misapplication of the therapy process; the client should lead identification of problems to work on. When the interventionist takes the lead, they may select problems that are not important to the client or cause the client to lose ownership of the process and reduce the client's self-efficacy for that or subsequent steps.

Incompatibility With Existing Workflows

Sometimes, an intervention or implementation strategy is not compatible with the interventionists' current workflows. This was an issue in a project focused on encouraging collaboration and task sharing between therapists and care managers in the treatment of depression and trauma in rural primary care clinics. Care managers had a variety of tasks, such as providing health education and resources to patients; thus, they could not support therapists in providing treatment to the extent that therapists expected. This highlighted the need to clarify roles, the division of labor, and workflows, including the types of cases that care managers could take on.

In another example, a project examined the use of BA to treat depression in individuals with cancer. Social workers were responsible for providing BA therapy alongside a wide range of services, including navigational support (eg, transportation, housing, and financial support). These navigational needs are often emergent and can crowd out the BA agenda for a session, preventing its successful delivery.

Intervention Buy-in (Value)

In some situations, the value of the intervention is not clear or acceptable to the patient. In a project evaluating the usability of an existing Veterans Affairs (VA) web-based SDM aid, study participants reported that sections of the web-based decision tool contained confusing and misleading content, which undermined their interest in engaging with the support it could offer. At other times, the mechanism of an intervention can prevents buy-in. For example, some PST recipients, expecting something more similar to talk therapy, felt that the therapy was childish—or not even a therapy—because of its manualized approach and extensive scaffolding. This reduced their willingness to engage in the homework required by the therapy or even to return to subsequent sessions.

Interventionist Buy-in (Trust)

The results also highlighted the need for EBPIs to support building rapport between interventionists and recipients. Although trust is not explicitly a step in most therapy and implementation processes, it is implied. For instance, some recipients did not trust their therapists' expertise, did not feel comfortable sharing with therapists, or doubted whether a therapist could help.

In the project evaluating the use of homework in PST and engagement, recipients were sometimes reluctant to disclose sensitive information to therapists. When this prevented them from disclosing problems or barriers they faced in addressing problems, it prevented recipients from effectively engaging in the therapy process. Lack of rapport and trust also sometimes prevented the intervention from being completed, as designed in a project evaluating the EBPI Promoting First Relationships, a parenting intervention for women with perinatal depression and anxiety. A core component of the Promoting First Relationships intervention is recording the interaction of mothers with their infants and using the recording to evaluate and provide feedback about their interactions. Not all mothers felt comfortable enough with their providers to record and share this interaction, which caused them to miss a core element of the intervention.

Overreliance on Technology

Other interventions and implementation strategies relied on technologies to which the intended interventionists or recipients did not have reliable access or were not comfortable using. This was particularly a barrier in rural areas and among older patients in accessing tele–mental health or web-based tools and resources, resulting in reduced access to care and support.

Requires Unavailable Infrastructure

Projects also found that interventions or implementation strategies required physical, systemic, or organizational infrastructure that was not available to the client, recipient, or organizational context. For example, in a project examining the use of BA to treat depression in patients being treated for cancer, treatment schedules for BA and cancer sometimes did not align.

XSL•FO
RenderX

When a patient completed their cancer treatment or changed cancer treatments, which could sometimes mean changing locations, this could disrupt their relationship with their BA interventionist, as access to a cancer center's mental health team was sometimes conditioned on the patient actively being in cancer treatment with that center. BA is not designed for transitions in providers; thus, this could end access to mental health therapy before the end of the mental health treatment plan.

Supervision was a critical element of this integrated care in the project designing for collaboration and task sharing between therapists and care managers in the treatment of depression and trauma in rural primary care clinics. However, therapists who would provide the supervision did not have training as supervisors and did not have supervision as part of their job responsibilities or time allocated in their work. The lack of dedicated time and training inhibited effective coordination between therapists and care managers.

Inadequate Scaffolding for Client/Recipient

Some interventions and implementation strategies lacked support for the recipient to understand and succeed in the core activities of the intervention. Both projects examining PST identified issues in this category. Some of the core concepts of PST, such as distinguishing among problems, goals, and solutions, were unclear to recipients. Although skilled interventionists could support navigating these distinctions during sessions, clients who remained unclear about these concepts at the end of this short-term intervention did not feel confident in applying PST skills without ongoing support. Other PST recipients were unsure whether certain topics (eg, relationship, divorce, and intimacy issues) were appropriate for PST. Without an explicit invitation to bring that topic into therapy, this delayed or prevented them from addressing problems most relevant to their needs and also caused them to become frustrated as they worked on less meaningful problems.

These projects also highlighted the need for scaffolding actions. PST, BA, and many other EBPIs require documenting and maintaining action plans during sessions so that clients can engage with them between sessions; thus, the interventionist can refer to it in subsequent sessions. These plans are not always documented or may be documented in a format only accessible to the client. This can prevent clients from pursuing plans between sessions, thus reducing the essential component of the intervention.

Inadequate Training and Scaffolding for Interventionists

Some interventions and implementation strategies did not provide sufficient training and scaffolding for the interventionist to deliver the therapy as designed or to respond to emergent challenges.

For example, in a project to support integrated care among therapists and care managers in the delivery of BA, not all care managers had sufficient training to deliver BA confidently and with fidelity. As therapists had limited availability to provide supervision (as described under infrastructure), these care managers did not have resources to which they could reliably turn when a situation exceeded the limits of their training. This resulted in handing off the patient to a therapist rather than improving their skills.

Similarly, the VA's SDM tool for posttraumatic stress disorder provided much information to support SDM but limited scaffolding for the process. Consequently, clinicians without training or significant experience in SDM tended to fall back on what they knew and initiated SDM less frequently in their interactions with clients.

Lack of Support for Necessary Communication

Interventions require effective communication between the interventionist and the recipient; however, they do not always sufficiently facilitate communication. For example, the VA's SDM tool provided much information to interventionists and recipients but did not support collaborative review and the reaching of a shared understanding of that information. This prevented the effective use of information.

The period between the sessions also presents communication challenges. Projects focused on BA noted that patients could encounter new situations or barriers to their action plans for which they were not prepared. Without a clear communication pathway to an interventionist or other asynchronous support, clients often deferred their actions until the next session. By that time, clients had trouble recounting details of the challenging situation, further limiting their ability to fully engage in treatment. Such situations interact with the infrastructure for providing mental health care; access to interventionists between sessions is certainly not always possible, but redesigns that support documentation of the situation for later recall and collaboration could better support communication.

Summary: Severity, Scope, and Complexity

In addition to reporting issues, project teams characterized them according to severity, scope, and complexity. Table 2 summarizes these ratings.

All reported issues were of severity that prevented the completion of a task (L1; 29 issues), which caused significant delay and/or frustration for the client or recipient (L2; 50 issues), or minor annoyances (L4; 11 issues). None of the teams reported catastrophic usability issues or issues that threatened recipient safety. Teams also did not report potential future improvements; we believe this is because they focused more on identifying current barriers to the delivery of EBPIs and implementation strategies.

XSL•FO
RenderX

Table 2. Severity, scope, and complexity by issue category (N=90).

Category	Number, n	Severity[a], n			Scope, n			Complexity, n		
		L1[b]	L2[c]	L3[d]	Global	Medium	Local	High	Medium	Low
Complex and/or cognitively overwhelming	12	6	4	2	10	0	2	3	5	4
Required time exceeds available time	10	4	5	1	7	1	2	3	4	3
Incompatibility with interventionist preference or practice	7	1	6	0	2	0	5	2	4	1
Incompatibility with existing workflow	4	1	2	1	4	0	0	1	0	3
Insufficient customization to clients/recipients	9	0	8	1	6	3	0	1	5	3
Intervention buy-in (value)	8	3	5	0	4	0	4	4	1	3
Interventionist buy-in (trust)	7	1	6	0	2	0	5	0	5	2
Overreliance on technology	5	3	1	1	3	0	2	2	2	1
Requires unavailable infrastructure	10	5	4	1	7	2	1	3	4	3
Inadequate scaffolding for client/recipient	6	1	2	3	3	0	3	0	2	4
Inadequate training and scaffolding for interventionists	6	4	2	0	4	2	0	2	3	1
Lack of support for necessary communication	6	0	5	1	5	0	1	2	4	0

[a]No project teams reported L0 (catastrophic, risks causing harm) or L4 (subtle, future enhancement) usability issues.

[b]Prevents completion of task.

[c]Causes significant delay and/or frustration.

[d]Minor effect on usability.

Prevalent, high-severity categories included the issues of *complex and/or cognitively overwhelming, required time exceeds available time*, and *requires unavailable infrastructure.* These were also among the categories with the highest complexity ratings. The first 2 categories were often interrelated: when part of an EBPI was complex or overwhelming for the interventionist or recipient, this would often lead to them running out of time in a session or other interactions. Alternatively, when aspects of a session not accounted for by the EBPI left less time during a session, complex components quickly overwhelmed interventionists and recipients, who tried to make the most of the limited time. In either case, the effects were severe: steps of the intervention were skipped or never started, likely decreasing its efficacy.

Requiring unavailable infrastructure was prevalent in projects adapting EBPIs or implementation strategies for different settings. In many of these issues, the original EBPI or implementation strategy prescribed (or assumed) resources and infrastructure that were not available in the new setting; thus, substantial redesign would be required for successful implementation. This also relates to the fourth category with high severity—*overreliance on technology*. Some interventions or implementations required technology that was not available in new settings (eg, expecting reliable, low-latency internet connectivity in rural settings) or for some users (eg, for people to have a newer device or internet connectivity at home); this distinction is also reflected in the bimodal distribution between global scope (all recipients for a setting) or local scope (just the recipients who did not have access to the necessary technology). When the intervention depended on access to such technology with no alternative ways of completing it, it prevented some recipients or entire settings from benefiting.

Finally, the different severity ratings for *inadequate training and scaffolding for interventionists* and *inadequate training and scaffolding for the client/recipient* highlight the different effects of usability issues directly affecting the recipient versus the interventionist. Interventionists with sufficient training and scaffolding could support a client through an intervention with reasonable treatment fidelity, even if the intervention's scaffolding for the client was insufficient. However, when the interventionist did not have sufficient training or scaffolding, they could quickly become frustrated and abandon the intervention or continue delivering it but with low fidelity, thus likely decreasing its efficacy.

The issues also varied in scope. Scope can refer to either how widespread an issue is with respect to a service or product design *or* what proportion of users are affected. Among the reported issues, the project teams weighed the user group most heavily in assessing the scope. Of the 90 issues, 57 issues were global in scope. These issues affected everyone within and across user groups (eg, the infrastructure or necessary scaffolding was not available to anyone in the project setting or something about the intervention or implementation's design was overwhelming or too time intensive for all interventionists or all recipients in a setting). Of the 90 issues, 25 issues were local in scope. These issues affected only a specific user group (eg, were not culturally responsive for *some* recipients in a setting or required a technology that only some interventionists or recipients lacked) or one aspect of the intervention or implementation strategy with which only some interventionists or recipients interacted. The remaining 8 issues were medium in scope. These issues

XSL•FO
RenderX

tended to affect all interventionists or recipients to some degree but were only severe for some.

Categories with the highest complexity included *complex and or cognitively overwhelming* and *required time exceeding available time*, as modifications to address these issues could affect the overall delivery and flow of an intervention. Some also required consideration of how much time and cognition an intervention could demand alongside other activities that might need to happen in nonspecialty care. However, even these categories included medium- and low-complexity issues that project teams believed they could address with simpler modifications to the intervention or implementation strategy. In terms of severity, issues that affected the interventionist tended to be more complex than those that affected the client; taken together, this emphasizes how a well-prepared interventionist can smooth out rough edges in an intervention for a client while usability issues for the interventionist flow through also become issues for the recipient.

The importance of designing to support the therapist has been noted in previous work on digital mental health technologies, which emphasizes the need to build on their existing workflows, avoid burdensome time demands, and support communication, although within reasonable boundaries [35]. We found that these needs are not unique to digital interventions, and usability issues related to these design principles are present in many EBPIs and implementation strategies.

Discussion

Principal Findings

Our results indicate widespread usability challenges for EBPIs and their implementation strategies. To some extent, this is a result of the way the Center shaped the research: investigators looked to use cases and settings in which people were known or anticipated to experience difficulties, with the goal of mitigating those difficulties and improving outcomes. However, they are consistent with other research noting usability barriers to the successful implementation and use of EBPIs [39,59,60], as well as other health interventions in complex settings [61]. Addressing these issues is an urgent concern for improving mental health care in nonspecialized settings.

Although a comprehensive discussion of how to redesign to address the identified usability issues is beyond the scope of this paper, we note that addressing these usability issues in intervention design and implementation may nonetheless not be easy. Several issues point to the need for interventions to better fit individual provider preferences and workflows, the context of a particular clinic, or the constraints of an individual patient, all of which might suggest more tailorable or customizable interventions and implementations. This parallels earlier guidance for digital mental health interventions to be more adaptable [35]. However, added customizability often drives complexity—to borrow from the discussion of customization in intelligence interfaces by Woods [62]: "[customization] is likely to provide the illusion of assistance while creating a new layer of burdens and complexities." This challenge is particularly salient as other usability issues point to the need to reduce the complexity of interventions so that they are less overwhelming during and between sessions. Innovative intervention designs and implementation strategies may yet be able to achieve both of these goals; however, it will be more difficult than addressing only one.

In addition, although previous research has examined the role of buy-in for both adopting a new or changed intervention or implementation strategy in an organization [63,64] and for whether someone seeks a particular health intervention, our results emphasize the need for ongoing attention to buy-in to support engagement. In the projects in our study, interventionists and recipients typically expressed initially favorable reactions to an intervention or implementation strategy—interventionists were seeking better practices, and recipients, as noted in the *Introduction* section, favored psychosocial interventions over alternatives but then found their confidence in the intervention, or the interventionist's ability to deliver it, waning as they engaged with specifics. Addressing these challenges is important to sustain engagement.

Our work with 13 project teams and the results of their work add to the growing literature on the value of adapting methods and constructs from HCD to examine and design EBPIs and implementation strategies. Our approach is consistent with the view that IS can contribute broad, multilevel frameworks that guide researchers and design and implementation teams, whereas HCD contributes specific methods to engage and learn from users [65-67]. That previous work also noted that although the two fields are complementary, efforts to align and scaffold the work of applying them would be necessary to achieve their benefits. In our work with the ALACRITY Center projects, we found that adapting usability issue analysis and reporting surveys facilitated teams in applying this lens to their work; however, it was not sufficient on its own. Teams that did not already have HCD expertise needed support from Center researchers to select HCD methods; apply effective but efficient analysis techniques; and, in most cases, iterate with Center researchers on reported usability issues. Consequently, we urge organizations planning to apply our DDBT approach to (1) develop resources that scaffold the process for design and implementation teams and (2) ensure that HCD and IS experts are available to support the design and implementation teams in navigating key decisions if they do not have that expertise internally. We are working on this within the UW ALACRITY Center, and we believe that this work will require contributions from across the field.

Limitations and Future Work

Overview

As described in the *Methods* section, Center experts extensively supported project teams as they developed their plans for assessing usability issues and as they worked to describe usability issues. On the basis of our interactions with teams, we observed tendencies that could prevent teams from working more independently from correctly attributing problems and, thus, from addressing the right problems. The most common example of this was framing usability issues in ways that attributed the problem to the user (usually an interventionist or a recipient) rather than to the intervention's design or implementation.

XSL•FO
RenderX

In this study, we addressed these challenges through dialog and coaching. We do not assume that the same level of support would be available to all teams working on designing and implementing interventions, and future research should continue to iterate on our process and supports for identifying and acting on usability issues so that teams with less support can successfully identify and address usability issues with fewer resources.

In addition, all data were collected from projects of the UW ALACRITY Center, led by university-based researchers, and funded through a call that required interdisciplinary teams. As noted in the *Methods* section, this may have biased our results toward EBPIs and implementation strategies with opportunities for improvement, toward EBPIs in which UW researchers have expertise, and toward technology-mediated interventions.

Stages of Design at which Usability Should Be Assessed and Issues Addressed

The projects in our sample engaged with interventions and implementation strategies at different stages, including examining interventionist and recipient experiences with existing interventions and implementation processes and testing prototypes of changed interventions or strategies that varied in their fidelity (level of detail, completeness, and functionality).

Owing to the limitations of prototypes, people are often unable to engage with low-fidelity prototypes in situations with high external validity; thus, there is a risk that usability issues that only emerge in real-world situations will go unobserved [68]. In addition, limited engagement with a low-fidelity prototype can make it difficult to distinguish between an initial adoption barrier and an ongoing barrier to its successful use. Therefore, it is important to assess usability issues at increasing levels of fidelity and in situations with increasing external validity. This might involve moving from *laboratory-based* testing (ie, users interacting with an intervention in circumscribed scenarios) to in vivo testing with fewer controls. However, this move is likely to require additional resources as it requires real-world intervention or implementation strategy deployment.

On the basis of our experiences, we recommend assessing EBPIs and implementation strategies for usability at all stages of their design and implementation, including initial development, refinement, pilot testing, and larger deployment. HCD has developed a mature understanding of the types of questions that should or should not be asked based on a prototype's capabilities and fidelity [69-73]. We believe that much of this guidance can be applied to EBPI design and implementation, such as the general principles articulated previously, although developing more specific guidance should be an area for future research.

From Issues to Designs: Heuristics and Assessing Improvements

In the usability of digital artifacts and services, research and accumulated experiences with usability issues have led to the development of *heuristics* or guiding principles for evaluating whether products are usable [74]. In addition to the classic 10 heuristics for usability engineering by Nielsen [33], there are also specialized heuristics for different types of products, such as chatbots [75], public displays [76], and voice user interfaces [77], and services [58]. Most relevant to our work, Lyon and Koerner [13] proposed preliminary heuristics for the implementability of EBPIs.

These heuristics are commonly taught for use in a process known as heuristic evaluation [78], in which experts in user research or design use them to review a prototype or product and identify potential problems. Practitioners also develop familiarity with common sets of heuristics or those specific to the domain in which they work and use them to guide design and not only evaluate it.

Previous work in implementation science began to articulate potential heuristics for the design and implementation of EBPIs [13] and used them to evaluate EBPIs [1]. On the basis of the results of our study, which contribute to an expanded understanding of usability issues, we propose a revised and expanded set of heuristics for the design and implementation of EBPIs (Table 3).

We propose that these heuristics guide the work of those designing new or refined EBPIs, as well as groups working on implementing and adapting existing EBPIs. However, a limitation of heuristic evaluations in other domains is that having a design expert examine a product is often insufficient for observing usability issues that emerge from complex behaviors or situations; therefore, in-depth usability evaluation in both controlled settings and the field remains necessary [68,79,80]. This is particularly true for EBPIs and implementation strategies, both of which are complex health interventions [81]. Future research should test whether the intervention and implementation strategy design guided by teams using these heuristics leads to better outcomes.

In addition to assessing whether the use of heuristics leads to better outcomes, future research should identify designs that are effective in addressing the common and severe usability issues identified in this study. Many of the teams that participated in this research work on doing so for their particular settings, interventions, adaptations, and users. Examining successful solutions for commonalities may lead to transferable design and implementation approaches.

XSL•FO
RenderX

Table 3. The proposed heuristics that could prevent or mitigate each category of usability issues.

Usability issue category	Proposed heuristic
Complex and/or cognitively overwhelming	Low cognitive load: The intervention should be simple, with clear, concise instructions, to minimize the amount of thinking required to complete a task. Minimize tasks and steps.
Time required exceeds time available	Efficiently uses time: The intervention should be designed to be completed within the time constraints of the delivery format, with attention to (1) other activities that may need to be completed in a contact point and (2) how much clients/recipients are asked to complete between contact points.
Incompatibility with interventionist preference or practice	Responsive to existing practices: Interventions should be familiar and responsive to a variety of interventionists' work styles. *Corollary:* interventions and implementation strategies should communicate prerequisites, with respect to provider practices, for their success.
Incompatibility with existing workflow	Responsive to existing system constraints: When possible, intervention structures should be flexible to different existing workflows. *Corollary:* Interventions and implementation strategies should communicate prerequisites, with respect to provider and setting workflows, for their success.
Insufficient customization to clients	Flexible and adaptable: Interventions and their implementation strategies should be adaptable and accessible to different client/patient profiles (eg, disability, age, culture, education, or income) and provide guidance for how to match and/or adapt to appropriate clients.
Intervention buy-in (value)	Demonstrates value: The intervention goal and process should be clear and acceptable for the needs and expectations of the client/patient, and to communicate its value.
Interventionist buy-in (trust)	Satisfaction and trust: The intervention should include space for the interventionist to establish a relationship and build rapport so the client/patient can assess trust and fit.
Overreliance on technology	Avoid technology choices that exclude: Interventions mediated by, implemented in, or otherwise relying on a technology should support users with a range of ability, comfort, and access and assess whether technology prerequisites are met and, if not, either add technology support or recommend another intervention or implementation
Requires unavailable infrastructure	Minimal infrastructure: Organizational infrastructure varies and cannot be guaranteed. Interventions should have ways to assess available infrastructure and adapt to accommodate differences or recommend alternative interventions/implementations if prerequisites for success cannot be met.
Inadequate scaffolding for the client	Learnable for recipients: The intervention/tool should include elements that support the client/patient in learning the concepts and workflow necessary for the client/patient to successfully carry out their role and activities.
Inadequate training and scaffolding for provider	Learnable for interventionists: The intervention/tool should include enough training, instructions, and in the moment support so the interventionist can successfully carry out their role and responsibilities.
Lack of support for necessary communication	Enhances communication and feedback: The intervention should include mechanisms to connect the client/patient and interventionist, allow for feedback to be shared about the process, and support adjustment of the treatment plan based on what is or is not working well.

Expanding to Other Interventions

As noted in the *Methods section*, our sample of projects is not representative of the entire space of EBPIs and implementation strategies. It is biased by the local expertise in our Center, and consequently, EBPIs such as BA and PST are overrepresented. In addition, our Center focused on nonspecialty settings, especially primary care. Finally, most HCD experts affiliated with the Center also work in human-computer interaction; thus, of the 13 projects, 8 were oriented toward those in which investigators hypothesized that some use of technology could better support the intervention or its implementation or was already being used in this way.

We believe that the issues we describe in this paper, as well as the approaches to identifying them in other EBPIs and implementation strategies, will be found in other types of therapies, whether they use technology to support their delivery. However, future research should assess this issue. A broader understanding of usability issues in mental health care will help to better understand the prevalence, severity, and implications of different types of usability issues.

EBPIs, implementation strategies, and digital technologies are all types of health service research products that can benefit from usability evaluation [30]. Although most of our Center's projects focused on the redesign of EBPIs by incorporating digital solutions, projects that focused directly on strategies indicate that the usability evaluation methods apply to them as well. However, some differences between EBPIs and implementation strategies may lend themselves to somewhat different testing techniques. Given that strategies relative to client-facing psychosocial interventions tend to involve a more diverse array of system levels, interested parties, and interactions, the direct evaluation of components via techniques such as behavioral rehearsals may be less feasible. Instead, implementation strategies might be most readily evaluated using techniques such as cognitive walk-through [25], which can focus on broader processes such as ways that organizational leaders influence the implementation climate.

We also anticipate that an important next step in the usability evaluation of complex health interventions such as EBPIs and strategies will be to extend this study to other domains of health. Although mental and behavioral health interventions are among the most complex in contemporary health care, often

XSL•FO
RenderX

representing reciprocal, socially mediated processes delivered over many months (eg, psychotherapy protocols lasting 12-16 sessions), other fields also use interventions with a high degree of complexity (eg, 6-month lifestyle interventions for women at high risk of breast cancer [82]). Although additional research is needed to determine the extent to which these interventions might demonstrate comparable usability issues, as well as be improved with similar heuristics, it is likely that many of our findings apply to improving the intervention implementability more broadly. Some usability issue categories identified in our research (eg, overreliance on technology and unavailable infrastructure, incompatibility with interventionist preference or practice, incompatibility with a setting's workflows, insufficient support for communication, and supporting) parallel concerns noted in other health systems research (eg, need to attend to workflows and communication; organizational policies, procedures, and culture; and computing infrastructure [61]).

Conclusions

Previous research has indicated that usability may explain the low adoption of EBPIs in nonspecialty settings [39]. A total of 13 projects examining EBPIs and associated implementation strategies identified 90 usability issues, which our team clustered into 12 categories. Of the 90 issues, 29 could prevent the completion of part of an EBPI, and 50 could cause significant delay or frustration in care. We contribute to an approach for analyzing and reporting usability issues in future projects, categories of usability issues that EBPI and implementation strategy designers should seek to avoid, and heuristics to support more usable EBPI and implementation strategy designs.

Acknowledgments

This study was supported by the National Institute of Mental Health (grant P50MH115837). The funding body had no role in the study design, manuscript writing, or decision to submit the paper for publication.

The contributors associated with the UW ALACRITY Center Researchers are as follows: Elena Agapie, Ian Bennett, Amy M Bauer, Karen Bearss, Dror Ben-Zeev, Amritha Bhat, Arpita Bhattacharya, Stephanie Brewer, Jessica A Chen, James Fogarty , John C. Fortney, Carmen Gonzalez, Gary Hsieh, Kendra Kamp, Julie Kientz, Jessica Jenness, Janine Jones, Rona L. Levy, Jill J. Locke, Brittany Mosser, Monica Oxford, Maggie Ramirez, Patrick J Raue, Brenna Renn, Jaime Snyder, Jina Suh, Emily C Williams, and Oleg Zaslavsky.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Usability issue guidance and survey.
[PDF File (Adobe PDF File), 165 KB - jmir_v24i6e37585_app1.pdf]

References

1. Raue PJ, Weinberger MI, Sirey JA, Meyers BS, Bruce ML. Preferences for depression treatment among elderly home health care patients. Psychiatr Serv 2011 May;62(5):532-537 [FREE Full text] [doi: 10.1176/ps.62.5.pss6205_0532] [Medline: 21532080]
2. Arean PA, Raue PJ, Sirey JA, Snowden M. Implementing evidence-based psychotherapies in settings serving older adults: challenges and solutions. Psychiatr Serv 2012 Jun;63(6):605-607 [FREE Full text] [doi: 10.1176/appi.ps.201100078] [Medline: 22638006]
3. Houle J, Villaggi B, Beaulieu M, Lespérance F, Rondeau G, Lambert J. Treatment preferences in patients with first episode depression. J Affect Disord 2013 May;147(1-3):94-100. [doi: 10.1016/j.jad.2012.10.016] [Medline: 23167975]
4. McHugh RK, Whitton SW, Peckham AD, Welge JA, Otto MW. Patient preference for psychological vs pharmacologic treatment of psychiatric disorders. J Clin Psychiatry 2013 Jun 15;74(06):595-602. [doi: 10.4088/jcp.12r07757]
5. Quiñones AR, Thielke SM, Beaver KA, Trivedi RB, Williams EC, Fan VS. Racial and ethnic differences in receipt of antidepressants and psychotherapy by veterans with chronic depression. Psychiatr Serv 2014 Feb 01;65(2):193-200 [FREE Full text] [doi: 10.1176/appi.ps.201300057] [Medline: 24178411]
6. Olfson M, Kroenke K, Wang S, Blanco C. Trends in office-based mental health care provided by psychiatrists and primary care physicians. J Clin Psychiatry 2014 Mar 15;75(03):247-253. [doi: 10.4088/jcp.13m08834]
7. Rones M, Hoagwood K. School-based mental health services: a research review. Clin Child Fam Psychol Rev 2000 Dec;3(4):223-241. [doi: 10.1023/A:1026425104386] [Medline: 11225738]
8. Owens J, Lyon A, Brandt N, Warner C, Nadeem E, Spiel C, et al. Implementation science in school mental health: key constructs in a developing research agenda. School Ment Health 2014 May 01;6(2):99-111 [FREE Full text] [doi: 10.1007/s12310-013-9115-3] [Medline: 26413173]
9. Duong MT, Bruns EJ, Lee K, Cox S, Coifman J, Mayworm A, et al. Rates of mental health service utilization by children and adolescents in schools and other common service settings: a systematic review and meta-analysis. Adm Policy Ment Health 2021 May;48(3):420-439. [doi: 10.1007/s10488-020-01080-9] [Medline: 32940884]

XSL•FO
RenderX

10. Lyon A, Ludwig K, Romano E, Koltracht J, Vander Stoep A, McCauley E. Using modular psychotherapy in school mental health: provider perspectives on intervention-setting fit. J Clin Child Adolesc Psychol 2014;43(6):890-901 [FREE Full text] [doi: 10.1080/15374416.2013.843460] [Medline: 24134063]
11. Stewart R, Chambless D, Stirman. Decision making and the use of evidence based practice: is the three-legged stool balanced? Pract Innov (Washington, DC) 2018 Mar;3(1):56-67 [FREE Full text] [doi: 10.1037/pri0000063] [Medline: 32219174]
12. Psychosocial Interventions for Mental and Substance Use Disorders: A Framework for Establishing Evidence-based Standards. Washington, DC: National Academies Press; 2015.
13. Lyon AR, Koerner K. User-centered design for psychosocial intervention development and implementation. Clin Psychol (New York) 2016 Jun 17;23(2):180-200 [FREE Full text] [doi: 10.1111/cpsp.12154] [Medline: 29456295]
14. Proctor EK, Powell BJ, McMillen JC. Implementation strategies: recommendations for specifying and reporting. Implement Sci 2013 Dec 01;8:139 [FREE Full text] [doi: 10.1186/1748-5908-8-139] [Medline: 24289295]
15. Lewis CC, Klasnja P, Powell BJ, Lyon AR, Tuzzio L, Jones S, et al. From classification to causality: advancing understanding of mechanisms of change in implementation science. Front Public Health 2018 May;6:136 [FREE Full text] [doi: 10.3389/fpubh.2018.00136] [Medline: 29868544]
16. Lyon AR, Munson SA, Renn BN, Atkins DC, Pullmann MD, Friedman E, et al. Use of human-centered design to improve implementation of evidence-based psychotherapies in low-resource communities: protocol for studies applying a framework to assess usability. JMIR Res Protoc 2019 Oct 09;8(10):e14990 [FREE Full text] [doi: 10.2196/14990] [Medline: 31599736]
17. ISO 9241-420:2011 - Ergonomics of human-system interaction — Part 420: selection of physical input devices. ISO. URL: https://www.iso.org/standard/52938.html [accessed 2019-10-09]
18. Novins DK, Green AE, Legha RK, Aarons GA. Dissemination and implementation of evidence-based practices for child and adolescent mental health: a systematic review. J Am Acad Child Adolesc Psychiatry 2013 Oct;52(10):1009-25.e18 [FREE Full text] [doi: 10.1016/j.jaac.2013.07.012] [Medline: 24074468]
19. 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. [doi: 10.1037/a0016032] [Medline: 19634954]
20. Becker EM, Smith AM, Jensen-Doss A. Who's using treatment manuals? A national survey of practicing therapists. Behav Res Ther 2013 Oct;51(10):706-710. [doi: 10.1016/j.brat.2013.07.008] [Medline: 23973815]
21. Garland A, Hawley K, Brookman-frazee L, Hurlburt M. Identifying common elements of evidence-based psychosocial treatments for children's disruptive behavior problems. J Am Academy Child Adolescent Psychiatry 2008 May;47(5):505-514. [doi: 10.1097/chi.0b013e31816765c2]
22. Shelton RC, Cooper BR, Stirman SW. The sustainability of evidence-based interventions and practices in public health and health care. Annu Rev Public Health 2018 Apr 01;39(1):55-76. [doi: 10.1146/annurev-publhealth-040617-014731] [Medline: 29328872]
23. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
24. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care 2012 Mar;50(3):217-226 [FREE Full text] [doi: 10.1097/MLR.0b013e3182408812] [Medline: 22310560]
25. Lyon AR, Coifman J, Cook H, McRee E, Liu FF, Ludwig K, et al. The Cognitive Walkthrough for Implementation Strategies (CWIS): a pragmatic method for assessing implementation strategy usability. Implement Sci Commun 2021 Jul 17;2(1):78 [FREE Full text] [doi: 10.1186/s43058-021-00183-0] [Medline: 34274027]
26. Cook CR, Lyon AR, Locke J, Waltz T, Powell BJ. Adapting a compilation of implementation strategies to advance school-based implementation research and practice. Prev Sci 2019 Aug 31;20(6):914-935 [FREE Full text] [doi: 10.1007/s11121-019-01017-1] [Medline: 31152328]
27. 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]
28. 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 16;104(12):510-520 [FREE Full text] [doi: 10.1258/jrsm.2011.110180] [Medline: 22179294]
29. Cabassa L. Implementation science: why it matters for the future of social work. J Soc Work Educ 2016;52(Suppl 1):S38-S50 [FREE Full text] [Medline: 28216992]
30. Lyon AR, Dopp AR, Brewer SK, Kientz JA, Munson SA. Designing the future of children's mental health services. Adm Policy Ment Health 2020 Sep 06;47(5):735-751 [FREE Full text] [doi: 10.1007/s10488-020-01038-x] [Medline: 32253634]
31. Lyon A, Comtois K, Kerns S, Landes S, Lewis C. Closing the science–practice gap in implementation before it widens. In: Implementation Science 3.0. New York, NY, USA: Spinger; 2020.
32. Giacomin J. What is human centred design? Design J 2015 Apr 28;17(4):606-623. [doi: 10.2752/175630614X14056185480186]
33. Nielsen J. Usability Engineering. San Diego CA, USA: Academic Press; 1994.

XSL•FO
RenderX

34. Rieman J, Franzke M, Redmiles D. Usability evaluation with the cognitive walkthrough. In: Proceedings of the Conference Companion on Human Factors in Computing Systems. 1995 Presented at: CHI95: Conference on Human Factor in Computing Systems; May 7 - 11, 1995; Denver, Colorado, USA. [doi: 10.1145/223355.223735]
35. Doherty G, Coyle D, Matthews M. Design and evaluation guidelines for mental health technologies. Interact Comput 2010 Jul;22(4):243-252. [doi: 10.1016/j.intcom.2010.02.006]
36. Beidas R, Cross W, Dorsey S. Show me, don't tell me: behavioral rehearsal as a training and analogue fidelity tool. Cogn Behav Pract 2014 Feb;21(1):1-11 [FREE Full text] [doi: 10.1016/j.cbpra.2013.04.002] [Medline: 25382963]
37. MacLeod H, Jelen B, Prabhakar A, Oehlberg L, Siek K, Connelly K. A guide to using Asynchronous Remote Communities (ARC) for researching distributed populations. EAI Endorsed Transact Pervasive Health Technol 2017 Jul 18;3(11):152898-152898. [doi: 10.4108/eai.18-7-2017.152898]
38. Lyles CR, Sarkar U, Osborn CY. Getting a technology-based diabetes intervention ready for prime time: a review of usability testing studies. Curr Diab Rep 2014 Oct;14(10):534 [FREE Full text] [doi: 10.1007/s11892-014-0534-9] [Medline: 25173689]
39. Lyon A, Pullmann M, Jacobson J, Osterhage K, Al Achkar M, Renn B, et al. Assessing the usability of complex psychosocial interventions: The Intervention Usability Scale. Implement Res Pract 2021 Feb 08;2:2633489520987829. [doi: 10.1177/2633489520987828]
40. Lewis J, Sauro J. The factor structure of the system usability scale. In: Human Centered Design. Berlin, Heidelberg: Springer; 2009.
41. Suh H, Shahriaree N, Hekler E, Kientz J. Developing and validating the user burden scale: a tool for assessing user burden in computing systems. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 Presented at: 2016 CHI Conference on Human Factors in Computing Systems; May 7-12, 2016; San Jose, CA, USA. [doi: 10.1145/2858036.2858448]
42. Weiner BJ, Lewis CC, Stanick C, Powell BJ, Dorsey CN, Clary AS, et al. Psychometric assessment of three newly developed implementation outcome measures. Implement Sci 2017 Dec 29;12(1):108 [FREE Full text] [doi: 10.1186/s13012-017-0635-3] [Medline: 28851459]
43. Lewinsohn PM. A behavioral approach to depression. In: The Psychology of Depression: Contemporary Theory and Research. Hoboken, New Jersey, United States: John Wiley & Sons; 1974.
44. Jacobson NS, Dobson KS, Truax PA, Addis ME, Koerner K, Gollan JK, et al. A component analysis of cognitive-behavioral treatment for depression. J Consulting Clin Psychol 1996;64(2):295-304. [doi: 10.1037/0022-006x.64.2.295]
45. Jacobson NS, Martell CR, Dimidjian S. Behavioral activation treatment for depression: Returning to contextual roots. Clinical Psychology: Science and Practice 2006 May 11;8(3):255-270. [doi: 10.1093/clipsy.8.3.255]
46. D'Zurilla TJ, Nezu AM. Problem-solving therapy. In: Handbook of Cognitive-Behavioral Therapies. New York, USA: Guildford Press; 2010.
47. Bhattacharya A, Nagar R, Jenness J, Munson S, Kientz J. Designing asynchronous remote support for behavioral activation in teenagers with depression: formative study. JMIR Form Res 2021 Jul 13;5(7):e20969 [FREE Full text] [doi: 10.2196/20969] [Medline: 34255665]
48. Jenness JL, Bhattacharya A, Kientz JA, Munson SA, Nagar R. Lessons learned from designing an asynchronous remote community approach for behavioral activation intervention for teens. Behav Res Ther 2022 Apr;151:104065. [doi: 10.1016/j.brat.2022.104065] [Medline: 35248749]
49. Ben-Zeev D, Meller S, Snyder J, Attah D, Albright L, Le H, et al. A digital toolkit (m-healer) to improve care and reduce human rights abuses against people with mental illness in West Africa: user-centered design, development, and usability study. JMIR Ment Health 2021 Jul 02;8(7):e28526 [FREE Full text] [doi: 10.2196/28526] [Medline: 34255712]
50. Bearss K, Tagavi D, Lyon AR, Locke J. Iterative redesign of a caregiver-mediated intervention for use in educational settings. Autism 2022 Apr 06;26(3):666-677. [doi: 10.1177/13623613211066644] [Medline: 34991353]
51. Increasing the usability and cultural responsiveness of a suicide-specific treatment for high schools. School Mental Health Assessment Research & Training Center. URL: https://smartcenter.uw.edu/research/projects/increasing-the-usability-and-cultural-responsiveness-of-a-suicide-specific-treatment-for-high-schools/ [accessed 2020-05-08]
52. Agapie E, Areán P, Hsieh G, Munson S. Longitudinal goal setting: a holistic, longitudinal process of transformative understanding in mental health. In: Proceedings of the 25th ACM Conference On Computer- Supported Cooperative Work And Social Computing. 2022 Presented at: The 25th ACM Conference On Computer- Supported Cooperative Work And Social Computing; Nov 12-16, 2022; Taipei, Taiwan.
53. Graham AK, Lattie EG, Powell BJ, Lyon AR, Smith JD, Schueller SM, et al. Implementation strategies for digital mental health interventions in health care settings. Am Psychol 2020 Nov;75(8):1080-1092. [doi: 10.1037/amp0000686] [Medline: 33252946]
54. Dumas J, Redish J. A Practical Guide to Usability Testing. Bristol, UK: Intellect books; 1999.
55. Lavery D, Cockton G, Atkinson M. Comparison of evaluation methods using structured usability problem reports. Behav Inf Technol 1997 Jan;16(4-5):246-266. [doi: 10.1080/014492997119824]
56. Holtzblatt K, Wendell JB, Wood S. Rapid Contextual Design: A How-to Guide to Key Techniques for User-Centered Design. Amsterdam, NL: Elsevier; 2004.

XSL•FO
RenderX

57. Hill C, Knox S, Thompson B, Williams E, Hess S, Ladany N. Consensual qualitative research: an update. J Counsel Psychol 2005 Apr;52(2):196-205. [doi: 10.1037/0022-0167.52.2.196]
58. Downe L. Good Services. London, UK: Laurence King Publishing; 2020.
59. Lyon A, Koerner K, Chung J. Usability Evaluation for Evidence-Based Psychosocial Interventions (USE-EBPI): a methodology for assessing complex intervention implementability. Implement Res Pract 2020 Sep 21;1:2633489520932924. [doi: 10.1177/2633489520932924]
60. Eisman A, Kilbourne A, Greene D, Walton M, Cunningham R. The user-program interaction: how teacher experience shapes the relationship between intervention packaging and fidelity to a state-adopted health curriculum. Prev Sci 2020 Aug;21(6):820-829 [FREE Full text] [doi: 10.1007/s11121-020-01120-8] [Medline: 32307625]
61. Sittig DF, Singh H. A new socio-technical model for studying health information technology in complex adaptive healthcare systems. In: Cognitive Informatics for Biomedicine: Human Computer Interaction in Healthcare. Switzerland: Springer; 2015.
62. Woods DD. The price of flexibility. In: Proceedings of the 1st international conference on intelligent user interfaces. 1993 Presented at: IUI93: ACM 1993 International Workshop on Intelligent User Interfaces; Jan 4-7, 1993; Orlando, Florida, USA. [doi: 10.1145/169891.169894]
63. Or C, Dohan M, Tan J. Understanding critical barriers to implementing a clinical information system in a nursing home through the lens of a socio-technical perspective. J Med Syst 2014 Sep;38(9):99. [doi: 10.1007/s10916-014-0099-9] [Medline: 25047519]
64. Holden RJ, Or CK, Alper SJ, Joy Rivera A, Karsh B. A change management framework for macroergonomic field research. Appl Ergon 2008 Jul;39(4):459-474. [doi: 10.1016/j.apergo.2008.02.016] [Medline: 18417095]
65. Chen E, Neta G, Roberts M. Complementary approaches to problem solving in healthcare and public health: implementation science and human-centered design. Transl Behav Med 2021 May 25;11(5):1115-1121 [FREE Full text] [doi: 10.1093/tbm/ibaa079] [Medline: 32986098]
66. Dopp AR, Parisi KE, Munson SA, Lyon AR. A glossary of user-centered design strategies for implementation experts. Transl Behav Med 2019 Nov 25;9(6):1057-1064. [doi: 10.1093/tbm/iby119] [Medline: 30535343]
67. Dopp AR, Parisi KE, Munson SA, Lyon AR. Aligning implementation and user-centered design strategies to enhance the impact of health services: results from a concept mapping study. Implement Sci Commun 2020 Feb 26;1(1):17 [FREE Full text] [doi: 10.1186/s43058-020-00020-w] [Medline: 32885179]
68. Goodman E, Kuniavsky M. Observing the User Experience: a Practitioner's Guide to User Research. Amsterdam, NL: Elsevier; 2012.
69. Virzi R. What can you learn from a low-fidelity prototype? Proc Human Factors Society Annual Meeting 2016 Aug 09;33(4):224-228. [doi: 10.1177/154193128903300405]
70. Rudd J, Stern K, Isensee S. Low vs high-fidelity prototyping debate. Interactions 1996 Jan 02;3(1):76-85. [doi: 10.1145/223500.223514]
71. Catani M, Biers D. Usability evaluation and prototype fidelity: users and usability professionals. Proc Human Factors Ergonomics Society Annual Meeting 2016 Nov 05;42(19):1331-1335. [doi: 10.1177/154193129804201901]
72. Houde S, Hill C. What do prototypes prototype? In: Handbook of Human-Computer Interaction (2nd Edition). Amsterdam, NL: Elsevier Science; 1997.
73. Sauer J, Seibel K, Rüttinger B. The influence of user expertise and prototype fidelity in usability tests. Appl Ergon 2010 Jan;41(1):130-140. [doi: 10.1016/j.apergo.2009.06.003] [Medline: 19632666]
74. Nielsen J, Molich R. Heuristic evaluation of user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1990 Presented at: CHI90: Conference on Human Factors in Computing; Apr 1 - 5, 1990; Seattle Washington USA URL: https://dl.acm.org/doi/proceedings/10.1145/97243 [doi: 10.1145/97243.97281]
75. Sanchez-Adame L, Mendoza S, Urquiza J, Rodriguez J, Meneses-Viveros A. Towards a set of heuristics for evaluating chatbots. IEEE Latin Am Trans 2021 Dec;19(12):2037-2045. [doi: 10.1109/tla.2021.9480145]
76. Somervell JP, Wahid S, McCrickard DS. Usability Heuristics for Large Screen Information Exhibits. In: Proceedings of the International Conference on Human-Computer Interaction. 2003 Presented at: International Conference on Human-Computer Interaction; Sep 1-5, 2003; Zurich, Switzerland.
77. Maguire M. Development of a heuristic evaluation tool for voice user interfaces. In: Design, User Experience, and Usability. Cham: Springer; 2019.
78. Nielsen J. Heuristic evaluation. In: Usability Inspection Methods. Hoboken, New Jersey, USA: John Wiley; 1994.
79. Jordan PW, McClelland IL, Thomas B, Weerdmeeste BA. Usability Evaluation In Industry. Boca Raton, Florida, USA: CRC Press; 1996.
80. Duh HB, Tan GC, Chen VH. Usability evaluation for mobile device: a comparison of laboratory and field tests. In: Proceedings of the 8th conference on Human-computer interaction with mobile devices and services. 2006 Presented at: MobileHCI06: Human-Computer Interaction with Mobile Devices and Services; Sep 12 - 15, 2006; Helsinki Finland URL: https://dl.acm.org/doi/10.1145/1152215.1152254 [doi: 10.1145/1152215.1152254]
81. Litaker D, Tomolo A, Liberatore V, Stange K, Aron D. Using complexity theory to build interventions that improve health care delivery in primary care. J Gen Intern Med 2006 Feb;21(S2):S30-S34. [doi: 10.1007/s11606-006-0272-z]

XSL•FO
RenderX

82. Han CJ, Korde LA, Reding S, Allott K, Van Doren M, Schwarz Y, et al. Investigation of a lifestyle intervention in women at high risk of breast cancer. West J Nurs Res 2018 Jul 23;40(7):976-996. [doi: 10.1177/0193945917697227] [Medline: 28335697]

Abbreviations

ALACRITY: Advanced Laboratories for Accelerating the Reach and Impact of Treatments for Youth and Adults with Mental Illness
BA: behavioral activation
CAMS: Collaborative Assessment and Management of Suicidality
DDBT: Discover, Design/Build, and Test
EBPI: evidence-based psychosocial intervention
HCD: human-centered design
IS: Implementation Science
PST: problem-solving therapy
SDM: shared decision-making
UW: University of Washington
VA: Veterans Affairs

Edited by R Kukafka; submitted 28.02.22; peer-reviewed by L Weinert, C Or; comments to author 17.03.22; revised version received 20.04.22; accepted 21.04.22; published 14.06.22.

Please cite as:
Munson SA, Friedman EC, Osterhage K, Allred R, Pullmann MD, Areán PA, Lyon AR, UW ALACRITY Center Researchers
Usability Issues in Evidence-Based Psychosocial Interventions and Implementation Strategies: Cross-project Analysis
J Med Internet Res 2022;24(6):e37585
URL: https://www.jmir.org/2022/6/e37585
doi:10.2196/37585
PMID:35700016

©Sean A Munson, Emily C Friedman, Katie Osterhage, Ryan Allred, Michael D Pullmann, Patricia A Areán, Aaron R Lyon, UW ALACRITY Center Researchers. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.06.2022. 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

Excess Google Searches for Child Abuse and Intimate Partner Violence During the COVID-19 Pandemic: Infoveillance Approach

Corinne A Riddell[1,2], MSc, PhD; Krista Neumann[2], BMath; N Jeanie Santaularia[3], PhD; Kriszta Farkas[2,3], PhD; Jennifer Ahern[2], PhD; Susan M Mason[3], PhD

[1]Division of Biostatistics, School of Public Health, University of California, Berkeley, CA, United States

[2]Division of Epidemiology, School of Public Health, University of California, Berkeley, CA, United States

[3]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, United States

Corresponding Author:
Corinne A Riddell, MSc, PhD
Division of Epidemiology
School of Public Health
University of California
2121 Berkeley Way West
Room #5404
Berkeley, CA, 94720
United States
Phone: 1 5106424618
Fax: 1 5106424618
Email: c.riddell@berkeley.edu

Abstract

Background: The COVID-19 pandemic has created environments with increased risk factors for household violence, such as unemployment and financial uncertainty. At the same time, it led to the introduction of policies to mitigate financial uncertainty. Further, it hindered traditional measurements of household violence.

Objective: Using an infoveillance approach, our goal was to determine if there were excess Google searches related to exposure to child abuse, intimate partner violence (IPV), and child-witnessed IPV during the COVID-19 pandemic and if any excesses are temporally related to shelter-in-place and economic policies.

Methods: Data on relative search volume for each violence measure was extracted using the Google Health Trends application programming interface for each week from 2017 to 2020 for the United States. Using linear regression with restricted cubic splines, we analyzed data from 2017 to 2019 to characterize the seasonal variation shared across prepandemic years. Parameters from prepandemic years were used to predict the expected number of Google searches and 95% prediction intervals (PI) for each week in 2020. Weeks with searches above the upper bound of the PI are in excess of the model's prediction.

Results: Relative search volume for exposure to child abuse was greater than expected in 2020, with 19% (10/52) of the weeks falling above the upper bound of the PI. These excesses in searches began a month after the Pandemic Unemployment Compensation program ended. Relative search volume was also heightened in 2020 for child-witnessed IPV, with 33% (17/52) of the weeks falling above the upper bound of the PI. This increase occurred after the introduction of shelter-in-place policies.

Conclusions: Social and financial disruptions, which are common consequences of major disasters such as the COVID-19 pandemic, may increase risks for child abuse and child-witnessed IPV.

(J Med Internet Res 2022;24(6):e36445) doi:10.2196/36445

KEYWORDS

child abuse; household violence; infoveillance; violence; domestic violence; abuse; Google; COVID-19

Introduction

Child abuse and intimate partner violence (IPV) are common. In the United States, 37% of children will be involved in an official investigation by Child Protective Services, whereas 25% of women and 11% of men experience IPV [1,2]. The downstream effects of abuse are profound; compared to adults not reporting histories of abuse, adults with histories of abuse

XSL•FO
RenderX

are 60% more likely to abuse drugs, 60% more likely to develop cardiovascular disease, and 3 times more likely to attempt suicide, demonstrating the wide-ranging effects on health across the life course [3,4]. These increased risks of adverse adult health outcomes are hypothesized to be mediated through several pathways such as increased high-risk behaviors (eg, substance abuse, smoking, and exercise avoidance), dysregulated immune functioning, and psychiatric disorders [5].

The risk factors for perpetrating child abuse and IPV include (but are not limited to) undergoing economic stress, feelings of isolation and disconnection, and parental stress (for child abuse) [6,7]. These risk factors were magnified during the first year of the COVID-19 pandemic through increased unemployment, shelter-in-place (SIP) policies, and remote schooling. Previous studies found that calls to the US hotline *Childhelp* increased during the first year of the pandemic, as did arrests, calls, and reports to police departments related to domestic violence [8,9]. At the same time, policy responses to mitigate financial uncertainty in the United States were substantial. For example, the Pandemic Unemployment Compensation (PUC) program increased unemployment payments by US $600 per week for 4 months, which offered an opportunity to explore the potential protective impacts of policies mitigating financial uncertainty. A challenge, however, is that the pandemic hindered the measurement of violence through traditional measures, for example, by reducing interactions with mandated reporters of child maltreatment [10]. A previous study found that during the Great Recession, places with decreases in Child Protective Services referrals had increases in both child mortality and Google searches for child abuse [11]. The divergence of reports from other measures of abuse suggests that abuse surveillance based on referrals may be hindered during periods of economic upheaval and that Google searches may help overcome this limitation.

We considered a broad set of Google searches based on the terms that individuals experiencing or witnessing child abuse or IPV would use as a measure of the incidence of household violence. This approach to monitoring epidemiologic trends falls under the field of “infoveillance,” where user-generated data collected from the internet and social media sites are used for surveillance [12,13]. Peaks in Google searches related to domestic violence were found to occur in the same months as peaks in police calls for domestic violence, suggesting that Google searches may offer a promising way to measure household violence outcomes [14]. The use of Google searches to measure epidemiologic outcomes has varied; searches related to influenza did not track well with the incidence of influenza-like illnesses [15-17], but searches related to the loss of smell correlated strongly with COVID-19 cases and deaths in the first months of the first wave of the pandemic [18]. Given that the pandemic hindered the measurement of violence in conjunction with similar trends of domestic violence with Google searches, the infoveillance approach is well-suited to study violence during a time of uncertainty.

The objective of this study was to establish whether Google searches for child abuse and IPV, which are nontraditional violence measures, increased during the pandemic and consider the timing of the increases in relation to SIP and economic policies that may affect violence risk factors. The findings will have implications for future policy responses to major crises.

Methods

Data Collection

To measure exposure to child abuse, child-witnessed IPV, and exposure to IPV using search data, we created 3 lists of queries that individuals who experience or witness abuse may search for on the internet (see Multimedia Appendix 1). Our methodology has been used in a previous study (Neumann et al, unpublished data, 2022). We conducted a review of the literature to determine how children and adults discuss these experiences (eg, Foster and Hagedorn [19]). We also considered the language used in validated scales measuring violence. We then tested the sensitivity and specificity of these search phrases by searching for them using a Google Incognito browser to ensure that the results were consistent with those experiencing or witnessing abuse and discarded the search phrases that did not appear relevant. We settled on 3 final search terms, each of which combined phrases specific to an abuse subtype (ie, exposure to child abuse, child-witnessed IPV, and exposure to IPV).

The Google Health Trends application programming interface (API) was used to obtain the Google search volume for 3 separate violence measures: exposure to child abuse, child-witnessed IPV, and exposure to IPV. To obtain the search data from the API, the researcher must first apply for an API key. Search terms, geographic region, and the time period of interest must be entered by the researcher, and the API will then return the probability of the search terms for the specified geo-time period. The returned results are based on a random sample of all Google searches and then, for readability, scaled by 10 million (2020 *Google Health Trends API Getting Started Guide*, unpublished document provided with API key). The API output must be interpreted as a relative search volume with an unknown denominator as the total number of searches used to calculate the returned probability is unknown to researchers. For this study, we obtained national-level weekly search volumes for each of our 3 search terms from 2017 to 2020 in the United States. We chose this geo-temporal resolution so that we could assess trends relative to important federal policy changes. All data were retrieved from the API between July 6 and 24, 2021. Since the returned values are probabilities based on a random sample of all Google searches, it is also important to account for sampling variability. To obtain more stable search volume estimates, 10 samples of each search were extracted. We computed both the mean and median of the estimates; their difference was very small, so we used the mean in the model.

Statistical Model

We first built a prediction model using data from 2017 to 2019. Using a linear regression fit using ordinary least squares, we modeled weekly Google search volume based on date, entered with a main effect term (to control for linear increases [or decreases] in Google search volume over the prediction period) and a restricted cubic spline for the week of year (with interior knots at the 10th, 50th, and 90th percentiles) to capture seasonal

XSL•FO
RenderX

patterning. We report the adjusted r-squared value to quantify the amount of outcome variation that is captured by the model.

We then used the model to predict the expected Google search volume for each week in 2020 alongside its 95% prediction intervals (PI). PIs place bounds on where observed individual values are expected to fall [20]. Thus, observations from 5% (2-3 weeks) of the 52 weeks in 2020 are expected to fall outside of the bounds of the 95% PI, and any more than that is considered a notable finding that is not predicted given the previous trends in the search volume.

We plotted weekly search volume, overlayed with the predicted searches and 95% PI. We annotated these plots with information about policies and payments that may ameliorate or accentuate risk factors for abuse, including the introduction of state-specific SIP policies (starting March 19, 2020 [21,22]), the date when individuals started receiving one-time Economic Impact Payments (April 17, 2020 [23]), and the end date of the PUC program (July 31, 2020 [23]), which provided an additional US $600 per week to claimants on top of usual unemployment benefits.

The R statistical software (version 4.1.0; R Foundation for Statistical Computing) was used to conduct this analysis. All code can be found in the GitHub repository [24].

Ethical Considerations

No personal information is available to researchers through the Google Health Trends API (2020 Google Health Trends API Getting Started Guide, unpublished document provided with API key). Google search volumes below an unspecified lower bound are suppressed by Google and not made available to researchers.

Results

Google Search Volumes

Yearly average Google search volumes for the abuse outcomes ranged between volumes of 58.3 (child-witnessed IPV in 2019) and 87.0 (exposure to child abuse in 2020; Table 1 and Figures S1-3 in Multimedia Appendix 2). All models met the assumptions required for linear regression (Figures S4-6 in Multimedia Appendix 2).

Table 1. Yearly average Google search volume and SD for exposure to child abuse, child-witnessed intimate partner violence (IPV), and exposure to IPV.

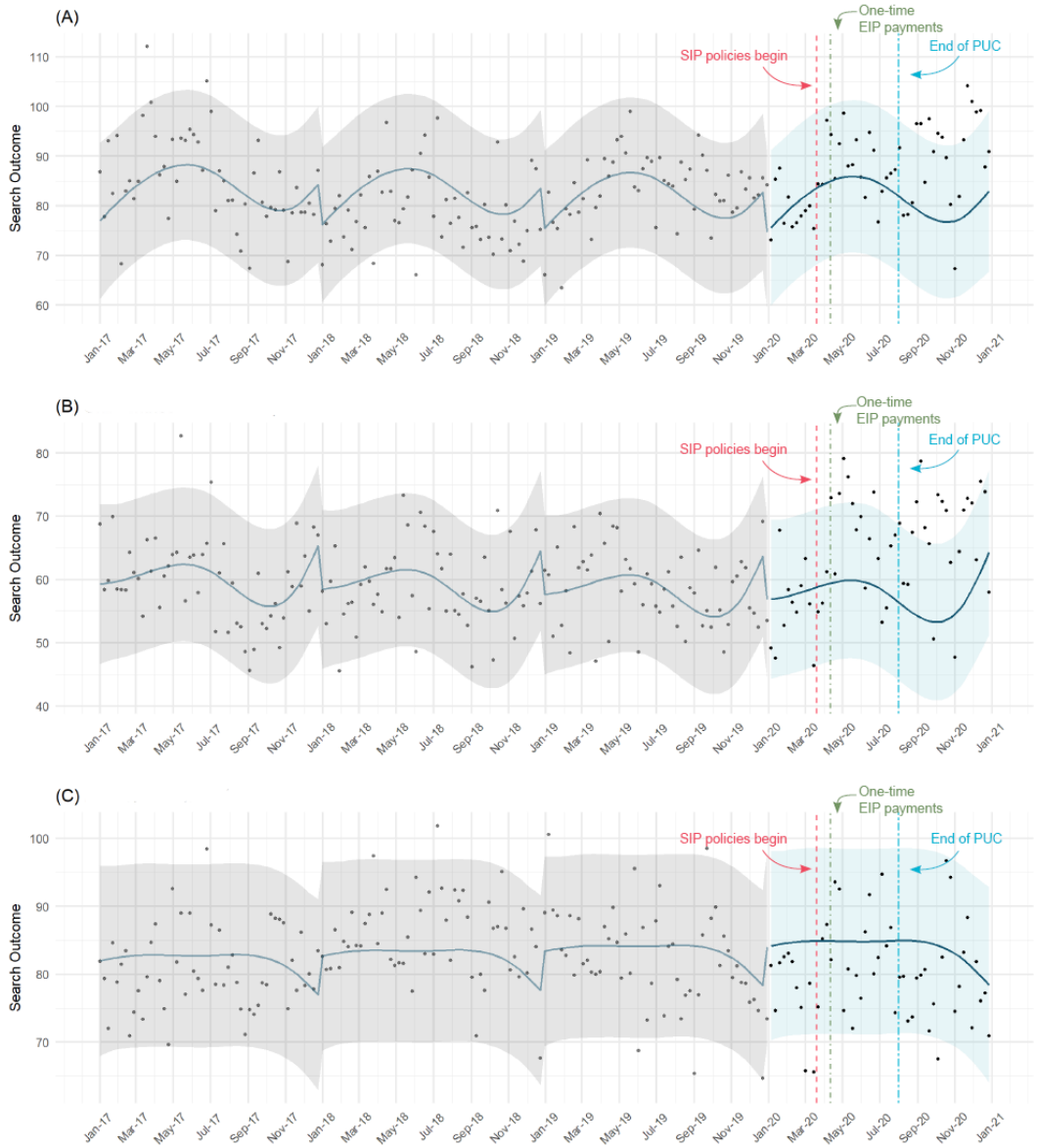
Domain of abuse	Google search volume, mean (SD)			
	2017	2018	2019	2020
Exposure to child abuse	85.0 (9.4)	79.0 (7.5)	83.8 (6.3)	87.0 (8.2)
Child-witnessed IPV	60.0 (7.0)	58.7 (6.4)	58.3 (6.0)	64.0 (8.7)
Exposure to IPV	80.7 (5.9)	85.6 (6.2)	82.1 (7.3)	80.1 (7.2)

Exposure to Child Abuse

Over the time period from 2017 to 2019, the Google search volume for child abuse was consistently highest in June and lowest in January, and the search volume decreased slightly year-to-year (Figure 1A). The model explained 15.3% of the variation in child abuse searches. From 2017 to 2019 inclusive, 5 out of 157 (3.2%) weeks fell outside of the PI, as expected, with 2 points above and 3 points below the upper and lower bounds, respectively. In 2020, 10 out of 52 (19%) weeks fell above the PI, all above the upper bound, suggesting an increase in child abuse searches. These increases were detected beginning August 30, 2020, about 4 weeks after the end of the PUC program.

XSL•FO
RenderX

Figure 1. Average weekly Google search volume (points) alongside predicted Google search volume (curve) and 95% prediction intervals (grey and blue bands) for (A) exposure to child abuse, (B) child-witnessed intimate partner violence, and (C) exposure to intimate partner violence, United States, 2017-2020. State-specific shelter-in-place (SIP) policies began on March 19, 2020, with California's SIP order, shortly after the national emergency was declared on March 13, 2020. The Coronavirus Aid, Relief, and Economic Security (CARES) Act was enacted on March, 27, 2020, and one-time Economic Impact Payments (EIP) were sent to nearly 90 million individuals by April 17, 2020, as part of the CARES Act. The Pandemic Unemployment Compensation (PUC) program, which was also part of the CARES act and provided an additional US $600 per week to claimants on top of usual unemployment benefits, expired on July 31, 2020.



Child-Witnessed IPV

From 2017 to 2019, the Google search volume was consistently highest in December and lowest in October, and the search volume declined slightly year-to-year (Figure 1B). The model explained 11.4% of the variation in child-witnessed IPV searches. From 2017 to 2019 inclusive, 6 out of 157 (3.8%) weeks occurred outside of the PI (3 points each above and below the upper and lower bounds), as expected, while in 2020, 17 out of 52 (33%) fell above the PI, suggesting an increase in child-witnessed IPV searches. All increases occurred after SIP policies began and continued after the end of the PUC program.

Exposure to IPV

From 2017 to 2019, IPV searches showed dips in November and December, with an otherwise flat yearly trend (Figure 1C). The model explained only 1.8% of the variation in the Google search volume for exposure to IPV. The model detected 3 out of 52 (6%) weeks with lower-than-expected search volumes in 2020, which was similar to prepandemic years (5%, 8/157).

Discussion

Principal Findings

Overall, we found that following the start of the COVID-19 pandemic, child abuse and child-witnessed IPV searches were elevated beyond that predicted by search history (from 2017 to 2019) for a substantial fraction of months/weeks. Child abuse searches increased a month after the PUC program ended. This pattern would be consistent with the hypothesis that the substantial loss in income from the end of the PUC program

XSL•FO
RenderX

may have led to an increase in child abuse; this would be valuable to examine in future research. These findings are consistent with previous literature linking decreased family income downstream of policy changes to increased reports to Child Protective Services [25-27]. Child-witnessed IPV searches, but not exposure to IPV searches, increased at the time of SIP policies. This might suggest greater opportunities for children to witness IPV rather than an increase in IPV itself, although searches for IPV itself might have been impacted by less privacy with household members spending more time together. The findings of increases in child abuse and child-witnessed IPV align with documented increases in calls to *ChildHelp* and police reports for domestic violence [8,9]. In the first study, calls to the hotline *Childhelp* increased 14% during 2020 compared to 2019 [8]. In the second study, increases between 10% to 27% were reported in the number of arrests, calls, or reports to police departments related to domestic violence [9]. This study is important because Google search data are a promising alternative to traditional measures of child abuse and IPV, which have well-documented reporting biases. Surveys estimate that less than 10% of child abuse [28] and less than half of domestic abuse is reported [29], implying that the majority of abuse goes undetected using traditional measures. The consistency of our findings with research that used hotline and police report data adds to the limited evidence on the impacts of the pandemic on child abuse and IPV and supports the promise of this approach to measuring abuse. Google search data may become particularly salient for this purpose at times when traditional detection approaches may be disrupted.

Limitations

Our study has limitations. First, we did not directly measure child abuse or IPV, and this study assumes that Google searches for child abuse and IPV track with the underlying incidence of the outcomes. A previous study found that Google searches for domestic abuse were associated with police calls for domestic violence in Finland [14], but no studies have examined this link in the United States. A second limitation is that Google searches can only be performed by individuals with access to the internet. Thus, the results may not generalize to households with no internet access, especially if the effect of the pandemic on abuse was larger or smaller compared to households with internet access. These results also do not reflect the experiences of children who do not use the internet, and thus may correspond more to the experiences of older children. Although some studies have found that Google searches can be affected by mass media related to the topic [30], we attempted to overcome this by limiting searches to those made by individuals experiencing or witnessing abuse, rather than focusing on broad searches like "child abuse" that may track with high-profile cases of abuse. We also removed terms that returned Google search results that were not relevant to exposure to abuse or child-witnessed IPV as part of a process we developed to use Google searches to measure epidemiologic constructs (Neumann et al, unpublished data, 2022). Lastly, the analytic approach we used can be hindered by multiple testing, since we deemed that a week of Google search volume was notable if it fell outside of the PI and we do this for each week in 2020. However, we found that multiple weeks—serially located in time—fall above the PI, which does not seem to suggest that we were detecting a "one-off" that happens to fall outside the PI. Thus, we do not believe that multiple testing played a role in these findings.

Conclusions

Social and financial disruptions, which are common consequences of major disasters, may increase the risks for child abuse and child-witnessed IPV. The increase in child abuse searches after the abrupt loss of income when PUC payments ceased suggests that economic mitigation strategies may be protective if sustained, though this study did not establish causation. Public health responses to future disasters should incorporate strategies to mitigate household violence.

Acknowledgments

We would like to thank Eli Michaels and Jackson Michaels for the Python code they adapted and shared to process the data returned by the Google Health Trends application programming interface. The results reported herein were funded by 2 pilot grants: one from the National Institute for Child Health and Human Development via the Berkeley Population Center (grant P2CHD073964) and the other from The Hawley Family Fund through the Division of Epidemiology and Community Health at the University of Minnesota. This work was also supported by the National Institute for Child Health and Human Development via grants from the Minnesota Population Center and the Interdisciplinary Population Health Science Training Program at the University of Minnesota (T32HD095134 and P2C HD041023).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms.
[DOCX File , 37 KB - jmir_v24i6e36445_app1.docx]

Multimedia Appendix 2
Maps of average search volume and regression model assumptions.
[DOCX File , 426 KB - jmir_v24i6e36445_app2.docx]

XSL•FO
RenderX

References

1. Kim H, Wildeman C, Jonson-Reid M, Drake B. Lifetime prevalence of investigating child maltreatment among US children. Am J Public Health 2017 Feb;107(2):274-280. [doi: 10.2105/AJPH.2016.303545] [Medline: 27997240]
2. Smith SG, Zhang X, Basile KC, Merrick MT, Wang J, Kresnow MJ, et al. National Intimate Partner and Sexual Violence Survey: 2015 data brief - updated release. National Center for Injury Prevention and Control, Centers for Disease Control and Prevention. 2018 Nov. URL: https://www.cdc.gov/violenceprevention/pdf/2015data-brief508.pdf [accessed 2021-08-11]
3. Norman RE, Byambaa M, De R, Butchart A, Scott J, Vos T. The long-term health consequences of child physical abuse, emotional abuse, and neglect: a systematic review and meta-analysis. PLoS Med 2012;9(11):e1001349 [FREE Full text] [doi: 10.1371/journal.pmed.1001349] [Medline: 23209385]
4. Rich-Edwards JW, Mason S, Rexrode K, Spiegelman D, Hibert E, Kawachi I, et al. Physical and sexual abuse in childhood as predictors of early-onset cardiovascular events in women. Circulation 2012 Aug 21;126(8):920-927 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.111.076877] [Medline: 22787111]
5. Sachs-Ericsson N, Cromer K, Hernandez A, Kendall-Tackett K. A review of childhood abuse, health, and pain-related problems: the role of psychiatric disorders and current life stress. J Trauma Dissociation 2009 Apr 03;10(2):170-188. [doi: 10.1080/15299730802624585] [Medline: 19333847]
6. Risk and protective factors for perpetration. Centers for Disease Control and Prevention. 2021 Nov 02. URL: https://www.cdc.gov/violenceprevention/intimatepartnerviolence/riskprotectivefactors.html [accessed 2022-04-13]
7. Risk and protective factors. Centers for Disease Control and Prevention. 2022 Apr 06. URL: https://www.cdc.gov/violenceprevention/childabuseandneglect/riskprotectivefactors.html [accessed 2022-04-13]
8. Ortiz R, Kishton R, Sinko L, Fingerman M, Moreland D, Wood J, et al. Assessing child abuse hotline inquiries in the wake of COVID-19: answering the call. JAMA Pediatr 2021 Aug 01;175(8):859-861 [FREE Full text] [doi: 10.1001/jamapediatrics.2021.0525] [Medline: 33938944]
9. Boserup B, McKenney M, Elkbuli A. Alarming trends in US domestic violence during the COVID-19 pandemic. Am J Emerg Med 2020 Dec;38(12):2753-2755 [FREE Full text] [doi: 10.1016/j.ajem.2020.04.077] [Medline: 32402499]
10. Welch M, Haskins R. What COVID-19 means for America's child welfare system. Brookings. 2020 Apr 30. URL: https://www.brookings.edu/research/what-covid-19-means-for-americas-child-welfare-system/ [accessed 2021-01-20]
11. Stephens-Davidowitz S. Unreported victims of an economic downturn. Squarespace. 2013 Jul 13. URL: https://static1.squarespace.com/static/51d894bee4b01caf88ccb4f3/t/51e22f38e4b0502fe211fab7/1373777720363/childabusepaper13.pdf [accessed 2021-05-24]
12. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
13. 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]
14. Koutaniemi EM, Einiö E. Seasonal variation in seeking help for domestic violence based on Google search data and Finnish police calls in 2017. Scand J Public Health 2021 May 11;49(3):254-259. [doi: 10.1177/1403494819834098] [Medline: 30973072]
15. Lazer D, Kennedy R, King G, Vespignani A. The parable of Google Flu: traps in big data analysis. Science 2014 Mar 14;343(6176):1203-1205. [doi: 10.1126/science.1248506] [Medline: 24626916]
16. Pollett S, Boscardin WJ, Azziz-Baumgartner E, Tinoco YO, Soto G, Romero C, et al. Evaluating Google flu trends in Latin America: important lessons for the next phase of digital disease detection. Clin Infect Dis 2017 Jan 01;64(1):34-41 [FREE Full text] [doi: 10.1093/cid/ciw657] [Medline: 27678084]
17. Lohr S. Google flu trends: the limits of big data. Bits Blog. 2014 Mar 28. URL: https://bits.blogs.nytimes.com/2014/03/28/google-flu-trends-the-limits-of-big-data/ [accessed 2021-08-25]
18. Walker A, Hopkins C, Surda P. Use of Google Trends to investigate loss-of-smell-related searches during the COVID-19 outbreak. Int Forum Allergy Rhinol 2020 Jul 15;10(7):839-847 [FREE Full text] [doi: 10.1002/alr.22580] [Medline: 32279437]
19. Foster JM, Hagedorn WB. Through the eyes of the wounded: a narrative analysis of children's sexual abuse experiences and recovery process. J Child Sex Abus 2014 Jul 29;23(5):538-557. [doi: 10.1080/10538712.2014.918072] [Medline: 24819252]
20. Kleinbaum DG, Kupper LL, Nizam A, Rosenberg ES. 5.10 Prediction of a new value of Y at X0. In: Applied Regression Analysis and Other Multivariable Models. 5th ed. Belmont, CA: Duxbury Press; 2014:68-72.
21. Mervosh S, Lu D, Swales V. See which states and cities have told residents to stay at home. New York Times. 2020 Apr 20. URL: https://www.nytimes.com/interactive/2020/us/coronavirus-stay-at-home-order.html [accessed 2021-01-20]
22. Raifman J, Nocka K, Jones D, Bor J, Lipson S, Jay J. COVID-19 US state policy database (CUSP). 2021 May. URL: https://statepolicies.com/ [accessed 2022-05-31]
23. Han J, Meyer BD, Sullivan JX. Income and poverty in the COVID-19 pandemic. National Bureau of Economic Research. 2020 Aug. URL: http://www.nber.org/papers/w27729.pdf [accessed 2022-05-31]
24. Github. URL: https://github.com/corinne-riddell/excess-abuse-covid [accessed 2022-05-31]

XSL•FO
RenderX

25. McLaughlin M. Less money, more problems: how changes in disposable income affect child maltreatment. Child Abuse Negl 2017 May;67:315-321. [doi: 10.1016/j.chiabu.2017.03.006] [Medline: 28340424]
26. Brooks-Gunn J, Schneider W, Waldfogel J. The Great Recession and the risk for child maltreatment. Child Abuse Negl 2013 Oct;37(10):721-729 [FREE Full text] [doi: 10.1016/j.chiabu.2013.08.004] [Medline: 24045057]
27. Kovski NL, Hill HD, Mooney SJ, Rivara FP, Morgan ER, Rowhani-Rahbar A. Association of state-level earned income tax credits with rates of reported child maltreatment, 2004-2017. Child Maltreat 2021 Jan 19:1077559520987302. [doi: 10.1177/1077559520987302] [Medline: 33464121]
28. MacMillan HL, Jamieson E, Walsh CA. Reported contact with child protection services among those reporting child physical and sexual abuse: results from a community survey. Child Abuse Negl 2003 Dec;27(12):1397-1408. [doi: 10.1016/j.chiabu.2003.06.003] [Medline: 14644057]
29. Morgan RE, Truman JL. Criminal victimization, 2019. U.S. Department of Justice. 2020 Sep. URL: https://bjs.ojp.gov/content/pub/pdf/cv19.pdf [accessed 2022-05-03]
30. Cervellin G, Comelli I, Lippi G. Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. J Epidemiol Glob Health 2017 Sep;7(3):185-189 [FREE Full text] [doi: 10.1016/j.jegh.2017.06.001] [Medline: 28756828]

Abbreviations

API: application programming interface
IPV: intimate partner violence
PI: prediction intervals
PUC: Pandemic Unemployment Compensation
SIP: shelter-in-place

Edited by A Mavragani; submitted 17.01.22; peer-reviewed by E Omaki, A Rovetta; comments to author 06.04.22; revised version received 09.05.22; accepted 24.05.22; published 13.06.22.

Please cite as:
Riddell CA, Neumann K, Santaularia NJ, Farkas K, Ahern J, Mason SM
Excess Google Searches for Child Abuse and Intimate Partner Violence During the COVID-19 Pandemic: Infoveillance Approach
J Med Internet Res 2022;24(6):e36445
URL: https://www.jmir.org/2022/6/e36445
doi:10.2196/36445
PMID:35700024

©Corinne A Riddell, Krista Neumann, N Jeanie Santaularia, Kriszta Farkas, Jennifer Ahern, Susan M Mason. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.06.2022. 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

Emergency Departments' Uptake of Telehealth for Stroke Versus Pediatric Care: Observational Study

Kori S Zachrison[1], MD, MSc; Emily M Hayden[1], MD, MHPE; Krislyn M Boggs[1], MPH; Tehnaz P Boyle[2], MD, PhD; Jingya Gao[1], MS; Margaret E Samuels-Kalow[1], MD, MPhil; James P Marcin[3], MD, MPH; Carlos A Camargo Jr[1], MD, DrPH

[1]Department of Emergency Medicine, Massachusetts General Hospital, Boston, MA, United States

[2]Department of Pediatrics, Boston Medical Center, Boston, MA, United States

[3]Department of Pediatrics, University of California Davis School of Medicine, University of California, Sacramento, CA, United States

Corresponding Author:
Kori S Zachrison, MD, MSc
Department of Emergency Medicine
Massachusetts General Hospital
55 Fruit Street
Boston, MA, 02114
United States
Phone: 1 6177266636
Email: kzachrison@partners.org

Abstract

Background: Telehealth for emergency stroke care delivery (telestroke) has had widespread adoption, enabling many hospitals to obtain stroke center certification. Telehealth for pediatric emergency care has been less widely adopted.

Objective: Our primary objective was to determine whether differences in policy or certification requirements contributed to differential uptake of telestroke versus pediatric telehealth. We hypothesized that differences in financial incentives, based on differences in patient volume, prehospital routing policy, and certification requirements, contributed to differential emergency department (ED) adoption of telestroke versus pediatric telehealth.

Methods: We used the 2016 National Emergency Department Inventory–USA to identify EDs that were using telestroke and pediatric telehealth services. We surveyed all EDs using pediatric telehealth services (n=339) and a convenience sample of the 1758 EDs with telestroke services (n=366). The surveys characterized ED staffing, transfer patterns, reasons for adoption, and frequency of use. We used bivariate comparisons to examine differences in reasons for adoption and use between EDs with only telestroke services, only pediatric telehealth services, or both.

Results: Of the 442 EDs surveyed, 378 (85.5%) indicated use of telestroke, pediatric telehealth, or both. EDs with both services were smaller in bed size, volume, and ED attending coverage than those with only telestroke services or only pediatric telehealth services. EDs with telestroke services reported more frequent use, overall, than EDs with pediatric telehealth services: 14.1% (45/320) of EDs with telestroke services reported weekly use versus 2.9% (8/272) of EDs with pediatric telehealth services (P<.001). In addition, 37 out of 272 (13.6%) EDs with pediatric telehealth services reported no consults in the past year. Across applications, the most frequently selected reason for adoption was "improving level of clinical care." Policy-related reasons (ie, for compliance with outside certification or standards or for improving ED performance on quality metrics) were rarely indicated as the most important, but these reasons were indicated slightly more often for telestroke adoption (12/320, 3.8%) than for pediatric telehealth adoption (1/272, 0.4%; P=.003).

Conclusions: In 2016, more US EDs had telestroke services than pediatric telehealth services; among EDs with the technology, consults were more frequently made for stroke than for pediatric patients. The most frequently indicated reason for adoption among all EDs was related to clinical care.

(J Med Internet Res 2022;24(6):e33981) doi:10.2196/33981

KEYWORDS

telehealth; telemedicine; emergency care; stroke; pediatric care

XSL•FO
RenderX

Introduction

Resource availability in US emergency departments (EDs) varies substantially, with major disparities in access to specialists [1]. With increasing regionalization of care, consultants may become less available in smaller or more rural EDs [2,3]. Telehealth has been increasingly acknowledged as a tool that may mitigate these disparities in access. With rapid expansion in the use of telehealth in emergency care delivery during the COVID-19 pandemic [4-6], it is possible that growth in technological infrastructure may be harnessed for longer-term solutions.

Telehealth for emergency stroke care delivery (telestroke) has a history of successful implementation, with a large body of work demonstrating improved delivery of stroke care [7-10]. One possible explanation for the extensive adoption of telestroke may be that it can provide a more cost-effective way for hospitals to achieve certification requirements that would have otherwise been difficult, if not impossible, to attain. For example, by providing 24/7 access to neurology consultation, telestroke has likely enabled many hospitals to achieve stroke center status without the expense of fully staffing in-person neurologist coverage. By achieving this designation, hospitals may then advertise themselves as stroke centers, may receive more stroke patient transports from prehospital emergency medical services (EMS), and may have the ability to admit stroke patients who may have otherwise been transferred to another facility. Given the generally favorable billing associated with diagnosis-related groups (DRGs) for patients admitted with stroke diagnoses, hospitals may readily see the financial advantages of investing in telestroke services.

In contrast to stroke, telehealth in pediatric emergency care (ie, pediatric telehealth) is infrequently used [11]. Many studies have demonstrated the relationship between telehealth and improved care and decision-making in the care of critically ill children in rural EDs [12,13], to help avoid unnecessary transfers [14], and to improve patient satisfaction [12]. Yet even when pediatric telehealth programs exist, low consult volumes often lead to discontinuation [15]. While designations for pediatric EDs exist, unlike stroke center certification, there is not a widely advertised national pediatric emergency care certification program that would enable hospitals to tout their certification status or to admit patients with advantageous billing. Overall, the financial and policy incentives for telestroke adoption are largely absent with respect to pediatric telehealth. We hypothesized that this contrast between stroke and pediatric emergency care might be an important factor driving differential uptake of telehealth for stroke versus for pediatric care.

To better understand barriers and motivators of telehealth adoption in EDs, we surveyed US EDs with telestroke and pediatric telehealth prior to the COVID-19 pandemic. Our primary objective was to determine whether differences in policy or certification requirements contributed to differential uptake of telestroke versus pediatric telehealth. We hypothesized that differences in hospital financial incentives, based on differences in patient volume, prehospital patient routing policy, and certification requirements, contributed to differential ED adoption of telestroke versus pediatric telehealth. These findings may have implications for health system leaders or policy makers interested in increasing uptake of pediatric telehealth.

Methods

Study Design, Selection of Participants, Survey, and Administration

We used data from the 2016 National Emergency Department Inventory–USA (NEDI-USA) survey responses to classify all responding EDs based on use of telehealth for stroke and for pediatric emergency care, and we targeted these EDs for a follow-up survey. The NEDI-USA survey is a brief, one-page survey that collects basic ED characteristics, including staffing and telehealth use, from EDs nationally (n=5404). The survey was administered in 2017 to characterize US EDs in 2016. The NEDI-USA survey is included in Multimedia Appendix 1 and was coordinated by the Emergency Medicine Network (EMNet) [16]; methods have been previously reported, including details of the telehealth component of the survey [11]. In 2018, as part of a study focused on understanding barriers and facilitators to ED adoption of telehealth, we administered a set of follow-up surveys to EDs using telehealth for stroke and for pediatric emergency care; this was done to understand the differential motivators of telehealth adoption between EDs using telehealth for stroke versus for pediatric emergency care. The surveys characterized details of the ED and clinical care, barriers to use for nonusers, and details of telehealth use in the preceding year for users.

Based on our a priori sample size calculations, we determined that we would need 453 EDs with telestroke services and 453 EDs with pediatric telehealth services for our follow-up survey in order to detect a 10% difference in the proportion of EDs indicating a policy-based motivation for adoption, assuming an α value of .05 and power of 0.80. There were more than 453 EDs with telestroke services (n=1758) but fewer than 453 EDs with pediatric telehealth services (n=339). Among these, there were 259 EDs that reported both pediatric telehealth and telestroke services. We identified a random sample of 366 EDs with telestroke services but not pediatric telehealth services, and all EDs that reported only pediatric telehealth services (n=339). Thus, the final population of EDs receiving the second survey included 76 EDs with pediatric telehealth but not telestroke services, 263 EDs with telestroke and pediatric telehealth services, and 103 EDs with telestroke but not pediatric telehealth services; this generated a total of 442 EDs for the follow-up survey. This study was conducted as part of a larger grant-funded study with other aims, related to understanding barriers and facilitators of ED adoption and use of telehealth [17,18]; this included a separate survey to rural EDs that did not receive telestroke or pediatric telehealth certification. On that survey, some responding EDs subsequently clarified that they did have telestroke or pediatric telehealth services; these EDs were then included in this analysis.

The follow-up survey varied slightly based on the nature of telehealth use in the surveyed EDs. The survey included additional questions characterizing ED staffing and transfer patterns, as these may influence telehealth adoption; provider perceptions of reasons for telehealth adoption; and estimated

XSL•FO
RenderX

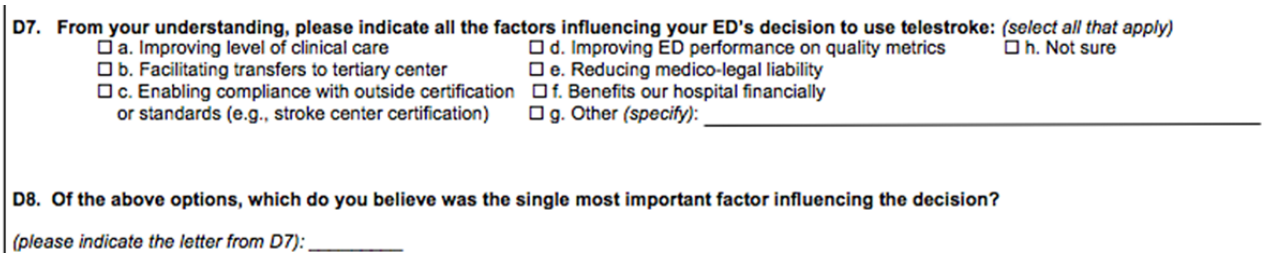
frequency of telehealth use. This survey included a combination of questions from prior research [19], as well as questions specifically developed for the aims of this study. The newly added questions were developed with input from several telehealth researchers, as well as emergency medicine researchers and nonresearch faculty. An example of the version of the survey including questions for EDs with telestroke and pediatric telehealth services is included in Multimedia Appendix 1.

We mailed the follow-up surveys by post to ED directors twice over a 3-month period and included a link to a web-based version of the survey in each mailing. We also followed up with nonresponsive and partially responding sites via telephone. Survey data were managed using REDCap (Research Electronic Data Capture; Vanderbilt University).

Outcomes

The primary outcome was reason for telehealth adoption. We dichotomized responses as motivated by policy or certification requirements (“yes” or ”no”). This was based on the response to the question asking about the single most important factor influencing the decision to adopt telehealth (Figure 1). Response options included (1) improving level of clinical care, (2) facilitating transfers to tertiary center, (3) enabling compliance with outside certification or standards, (4) improving ED performance on quality metrics, (5) reducing medicolegal liability, (6) benefits our hospital financially, (7) other (specify), and (8) not sure. Responses classified as policy or certification motivated were “enabling compliance with outside certification or standards” and “improving ED performance on quality metrics.” Any free-text responses included in the “other” section were independently reviewed and coded by two authors (KSZ and EMH) as policy or certification motivated or not.

Figure 1. Screenshot of the survey question regarding the reason for telehealth adoption. ED: emergency department.



Other Variables of Interest

The full survey is included in Multimedia Appendix 1. We also collected data on ED volume and characteristics of the ED space and staffing. We collected stroke-related variables, including certification status, typical treatment and stroke patient dispositions, availability of neurologists, and frequency of telestroke use in 2016. We collected pediatric emergency care–related variables, including who typically cares for a child presenting to the ED, availability of in-person pediatric consultation, and estimated frequency of pediatric telehealth consultation in 2016.

We identified academic EDs as those that were the primary site for an emergency medicine residency [20]. We identified rural EDs as those located outside of a core-based statistical area [21]. We used data from the Center for Connected Health Policy [22] and the American Telehealth Association 2016 Gaps Analysis [23] to identify states’ telehealth policy environment based on state policy in 2016. States were categorized as having no coverage parity (ie, no requirement for payors to reimburse telehealth care), a partial or conditional mandate for payment parity, or full payment parity.

Analysis

Data analysis was performed using SAS software (version 9.4; SAS Institute Inc). Our analysis focused on EDs indicating that they had telestroke services, pediatric telehealth services, or both in 2016. We compared EDs by telehealth usage using the Kruskal-Wallis test for continuous variables, the chi-square test for categorical variables, and the Fisher exact test for small-sized categorical variables of interest (ie, >20% of cells with expected frequencies of <5). For simplicity, we report *P* values only for the key comparisons. We addressed our research hypothesis by determining the proportion of telestroke EDs for which the reason for adoption was policy motivated, and the proportion of pediatric telehealth EDs for which the reason for adoption was policy motivated.

Ethics Approval

This study was approved by the Mass General Brigham Institutional Review Board (protocol No. 2017P000130).

Results

Overview

The 2016 NEDI-USA survey yielded responses from 4506 out of 5404 (83.38%) EDs; 4410 out of 5404 (81.61%) EDs responded to the telehealth question asking them to report presence or absence of telehealth in the ED [11]. Based on the responses to the telehealth questions on the NEDI-USA, we identified EDs using telestroke and pediatric telehealth for our follow-up survey. Details of the sampling strategy are included in the Methods.

Of the 442 EDs sampled for our follow-up survey, 378 (85.5%) responded; this included 106 (28.0%) EDs with telestroke but not pediatric telehealth, 214 (56.6%) EDs with telestroke and pediatric telehealth, and 58 (15.3%) EDs with pediatric telehealth but not telestroke.

ED Characteristics

Characteristics of the 378 EDs in our sample are provided in Table 1. EDs had a median annual volume of 9959 (IQR

XSL•FO
RenderX

2475-30,000) visits and a median annual pediatric volume of 1800 (IQR 429-5163) visits. Very few were academic EDs (n=6, 1.6%), but nearly half were rural (n=179, 47.4%). More EDs were in the Midwest (n=144, 38.1%) and the South (n=102, 27.0%) than in the West (n=79, 20.9%) and the Northeast (n=49, 13.0%). Most EDs were in states without any payment parity policy (n=237, 62.7%), though 107 (28.3%) were in states with partial payment parity, and 34 (9.0%) were in states with full payment parity. Other frequently reported applications of telehealth in these EDs included psychiatry (n=173, 45.8%) and trauma (n=159, 42.1%).

Table 1. Emergency department characteristics overall and by type of telehealth used.

ED[a] characteristics	All EDs in sample (N=378)	EDs with telestroke only (n=106)	EDs with telestroke and pediatric telehealth (n=214)	EDs with pediatric telehealth only (n=58)
ED volume (visits), median (IQR)	9959 (2475-30,000)	20,945 (8605-40,033)	4783 (1278-20,000)	14,733 (6005-35,400)
ED pediatric volume (visits), median (IQR)	1800 (429-5163)	3664 (1596-6590)	860 (240-3211)	2993 (766-6000)
Pediatric space in ED, n (%)	39 (10.3)	15 (14.2)	15 (7.0)	9 (15.5)
PECC[b], n (%)	46 (12.1)	13 (12.3)	23 (10.7)	10 (17.2)
Total number of beds (adult and pediatric), median (IQR)	9 (4-19)	13 (8-24)	6 (3-14)	12 (6-20)
Number of FTE[c] attendings, median (IQR)	4 (2-8)	6 (4-12)	4 (1-6)	5 (4-8)
Proportion of attending emergency physicians BC/BE[d] by ABEM[e], AOBEM[f], or ABP[g] in pediatric emergency medicine (%), n (%)				
<20	104 (27.5)	21 (19.8)	71 (33.2)	12 (20.7)
20-49	32 (8.5)	10 (9.4)	16 (7.5)	6 (10.3)
50-79	34 (9.0)	7 (6.6)	17 (7.9)	10 (17.2)
80-100	156 (41.3)	55 (49.1)	77(36.0)	24 (41.4)
Missing	52 (13.8)	13 (12.3)	33 (15.4)	6 (10.3)
Academic, n (%)	6 (1.6)	3 (2.8)	1 (0.5)	2 (3.4)
Rural location, n (%)	179 (47.4)	32 (30.2)	125 (58.4)	22 (37.9)
Region, n (%)				
Northeast	51 (13.5)	17 (16.0)	18 (8.4)	16 (27.6)
Midwest	143 (37.8)	30 (28.3)	101 (47.2)	12 (20.7)
South	103 (27.2)	37 (34.9)	46 (21.5)	20 (34.5)
West	81 (21.1)	22 (20.8)	49 (22.8)	10 (17.2)
State payment policy, n (%)				
Full parity	34 (9.0)	6 (5.7)	21 (9.8)	7 (12.1)
Partial parity	107 (28.3)	35 (33.0)	52 (24.3)	20 (34.5)
None	237 (62.7)	65 (61.3)	141 (65.9)	31 (53.4)
Other specialties for which ED receives telehealth, n (%)				
Psychiatry	173 (45.8)	30 (28.3)	125 (58.4)	18 (31.0)
Trauma	159 (42.1)	21 (19.8)	126 (58.9)	12 (20.7)
Dermatology	54 (14.3)	5 (4.7)	44 (20.6)	5 (8.6)
Radiology	59 (15.6)	8 (7.5)	48 (22.5)	3 (5.2)

[a]ED: emergency department.

[b]PECC: pediatric emergency care coordinator.

[c]FTE: full-time equivalent.

[d]BC/BE: board certified or board eligible.

[e]ABEM: American Board of Emergency Medicine.

[f]AOBEM: American Osteopathic Board of Emergency Medicine.

[g]ABP: American Board of Pediatrics.

XSL•FO
RenderX

Stroke Care and Telestroke Use

EDs with telestroke services only were more frequently Joint Commission–certified stroke centers relative to those with telestroke and pediatric telehealth services or pediatric telehealth services only (Table 2). With respect to availability of an in-person neurologist, EDs with telestroke and pediatric telehealth had the least availability. Of all telestroke EDs (n=320), 45 (14.1%) reported weekly use and another 44 (13.8%) reported using telestroke services every 1 to 2 weeks during 2016. Fewer than one-third of EDs with telestroke reported administering alteplase without a telestroke consultation (n=94, 29.4%). There were no significant differences in admission practices between groups, with the exception of admission of alteplase-treated patients. EDs with only telestroke services reported capacity to admit alteplase-treated stroke patients more frequently (36/106, 34.0%) relative to EDs with both telestroke and pediatric telehealth services (42/214, 19.6%) or EDs with only pediatric telehealth services (13/58, 22%; *P*=.02).

XSL•FO
RenderX

Table 2. Telestroke use and the clinical care of stroke patients.

ED[a] characteristics	All EDs in sample (N=378), n (%)	EDs with telestroke only (n=106), n (%)	EDs with telestroke and pediatric telehealth (n=214), n (%)	EDs with pediatric telehealth only (n=58), n (%)
Joint Commission certification	117 (31.0)	45 (42.5)	56 (26.2)	16 (27.6)
If no Joint Commission certification, alternative stroke certification status	59/261 (22.6)	16/61 (26.2)	37/158 (23.4)	6/42 (14.3)
Neurologist available in person in the ED	63 (16.7)	27 (25.5)	25 (11.7)	11 (19.0)
If neurologist available in person, timing of arrival (minutes)[b]				
0-29	39 (61.9)	18 (66.7)	14 (56.0)	7 (63.6)
30-59	16 (25.4)	8 (29.6)	5 (20.0)	3 (27.3)
≥60	5 (7.9)	0 (0)	4 (16.0)	1 (9.1)
If neurologist available, is available 24/7[b]	38 (60.3)	17 (63.0)	12 (48.0)	9 (81.8)
Approximate number of telestroke consultations in 2016				
None	37 (9.8)	10 (9.4)	27 (12.6)	N/A[c]
<12 (<1/month)	133 (35.2)	33 (31.1)	100 (46.7)	N/A
12-25 (every 3-4 weeks)	41 (10.8)	18 (17.0)	23 (10.7)	N/A
26-52 (every 1-2 weeks)	44 (11.6)	15 (14.2)	29 (13.6)	N/A
>52 (>1/week)	45 (11.9)	22 (20.8)	23 (10.7)	N/A
Missing	20 (5.3)	8 (7.5)	12 (5.6)	N/A
In 2016, was alteplase ever administered to a stroke patient in the ED without a telestroke consultation?				
Yes	94 (24.9)	30 (28.3)	64 (30.0)	N/A
No	167 (44.2)	55 (51.9)	112 (52.3)	N/A
Not sure	46 (12.2)	16 (15.1)	30 (14.0)	N/A
In 2016, approximately how many stroke patients were treated with alteplase in your ED?				
0	29 (7.7)	5 (4.7)	21 (9.8)	3 (5.0)
1-3	129 (34.1)	28 (26.4)	83 (38.8)	18 (31.0)
≥4	172 (49.4)	56 (52.8)	84 (39.3)	32 (55.2)
Not sure	36 (9.5)	14 (13.2)	20 (9.3)	2 (3.4)
Patients typically admitted by hospital				
Patients who experienced TIA[d]	292 (77.2)	85 (80.2)	164 (76.6)	43 (74.1)
Patients who experienced stroke, without alteplase	215 (56.9)	64 (60.4)	113 (52.8)	38 (65.5)
Patients who experienced stroke, treated with alteplase	91 (24.1)	36 (34.0)	42 (19.6)	13 (22.4)

[a]ED: emergency department.

[b]These values are based on the number of neurologists available in person in the ED—all EDs: n=63; EDs with telestroke: n=27; EDs with both: n=25; EDs with pediatric telehealth: n=11.

[c]N/A: not applicable; no telestroke services.

[d]TIA: transient ischemic attack.

Pediatric Care and Pediatric Telehealth Use

The vast majority of EDs in the sample reported that children were generally cared for by a general emergency physician (289/378, 76.5% overall); however, this did vary by category of telehealth use (Table 3). Among all EDs with pediatric telehealth services, frequency of use was relatively low, with 8 out of 272 (2.9%) reporting weekly use, and 12 (4.4%) reporting use every 1 to 2 weeks. Most (164/272, 60.3%) reported use fewer than 12 times over the year. Many EDs with pediatric telehealth did report using telehealth for pediatric mental health consultation (82/272, 30.1%).

XSL•FO
RenderX

Table 3. Use of telehealth for pediatric emergency care and the clinical care of children.

ED[a] characteristics	All EDs in sample (N=378), n (%)	EDs with telestroke only (n=106), n (%)	EDs with telestroke and pediatric telehealth (n=214), n (%)	EDs with pediatric telehealth only (n=58), n (%)
Who typically cares for a child presenting to the ED at 6 PM on a typical day?				
Pediatric emergency physician	19 (5.0)	8 (7.5)	6 (2.8)	5 (8.6)
General emergency physician	289 (76.5)	94 (88.7)	145 (67.8)	50 (86.2)
General pediatrician	23 (6.1)	9 (8.5)	9 (4.2)	5 (8.6)
Physician of another specialty	90 (23.8)	17 (16.0)	62 (29.0)	11 (19.0)
Physician assistant or nurse practitioner	252 (66.7)	70 (66.0)	148 (69.2)	34 (58.6)
Professional available for in-person pediatric consultation				
Pediatrics attending	109 (28.8)	45 (42.4)	43 (20.1)	21 (36.2)
Pediatrics trainee	9 (2.4)	3 (2.8)	4 (1.9)	2 (3.4)
Family medicine attending	128 (33.9)	34 (32.1)	76 (35.5)	18 (31.0)
Family medicine trainee	11 (2.9)	3 (2.8)	5 (2.3)	3 (5.2)
Other	40 (10.6)	10 (9.4)	24 (11.2)	6 (10.3)
None	144 (38.1)	32 (30.2)	92 (43.0)	20 (34.5)
Does a physician assistant or nurse practitioner ever provide care for a child in the ED? (yes)	283 (74.9)	77 (72.6)	164 (76.6)	42 (72.4)
If yes to above, are they supervised by the on-site attending? (yes)	174/283 (61.5)	56/77 (72.7)	81/164 (49.4)	37/42 (88.1)
In 2016, approximate number of telehealth consultations for pediatric emergency care				
None	37 (9.8)	N/A[b]	29 (13.6)	8 (13.8)
<12 (<1/month)	164 (43.4)	N/A	127 (59.2)	37 (63.8)
12-25 (every 3-4 weeks)	40 (10.6)	N/A	32 (15.0)	8 (13.8)
26-52 (every 1-2 weeks)	12 (3.2)	N/A	10 (4.7)	2 (3.5)
>52 (>1/week)	8 (2.1)	N/A	8 (3.7)	0 (0)
In 2016, did your ED ever use telehealth for pediatric mental health consultation? (yes)	82 (21.7)	N/A	69 (32.2)	13 (22.4)

[a]ED: emergency department.

[b]N/A: not applicable; no pediatric telehealth services.

Policy- Versus Nonpolicy-Motivated Adoption of Telehealth

Among all EDs with telestroke services, 213 out of 320 (66.6%) reported a policy-motivated reason for adoption, whereas among EDs with pediatric telehealth services, 138 out of 272 (50.7%) did so (Table 4). When asked to select the *single most important* factor influencing the decision for adoption, policy-motivated reasons were rarely selected, but they were selected slightly more frequently by EDs with telestroke services (12/320, 3.8%) than by EDs with pediatric telehealth services (1/272, 0.4%; *P*=.003). Reasons specified when "other" was selected are included in Table S1 in Multimedia Appendix 1.

XSL•FO
RenderX

Table 4. Factors influencing emergency department use of telehealth.

Factor	EDs[a] selecting factors influencing use of telestroke (n=320), n (%)		EDs selecting factors influencing use of pediatric telehealth (n=272)	
	When selecting *all that apply*	When selecting the s*ingle most important* factor	When selecting *all that apply*	When selecting the s*ingle most important* factor
Improving level of clinical care	276 (86.3)	223 (69.7)	231 (84.9)	187 (68.8)
Facilitating transfer to tertiary center	236 (73.8)	34 (10.6)	218 (80.1)	56 (20.6)
Enabling compliance with outside certification or standards	141 (44.1)	5 (1.6)	75 (27.6)	0 (0)
Improving ED performance on quality metrics	198 (61.9)	7 (2.2)	128 (47.1)	1 (0.4)
Reducing medicolegal liability	136 (42.5)	2 (0.6)	107 (39.3)	2 (0.7)
Benefits our hospital financially	57 (17.8)	2 (0.6)	50 (18.4)	1 (0.4)
Other	27 (8.4)	18 (5.6)	20 (7.4)	8 (2.9)
Not sure	11 (3.4)	11 (3.4)	4 (1.5)	3 (1.1)
Missing	15 (4.7)	18 (5.6)	13 (4.8)	14 (5.1)
Policy motivated[b]	213 (66.6)	12 (3.8)	138 (50.7)	1 (0.4)

[a]ED: emergency department.

[b]This response was based on the following two responses: "enabling compliance with outside certification or standards" and "improving ED performance on quality metrics."

Discussion

Principal Findings

In this study, we surveyed a national sample of EDs with telestroke services and all EDs with pediatric telehealth services prior to the COVID-19 pandemic. Among these EDs, whether using telehealth for telestroke, pediatric telehealth, or both, the single most commonly reported factor driving telehealth use was for the purpose of improving clinical care. Policy- or certification-related reasons were selected as a motivator by many EDs, more often for telestroke services than for pediatric telehealth services. However, when asked about the single most important reason, the vast majority of all EDs indicated that telehealth was used for improving clinical care.

Comparison to Prior Work

There has been little previous work focusing specifically on EDs' reasons for adoption of particular lines of telehealth services. A previous mixed methods study of 17 programs providing pediatric telehealth services reported a number of barriers and facilitators to adoption and successful maintenance of telehealth programs [15]. The investigators suggested that particular policy-related solutions may be effective for realigning incentives and enabling more widespread adoption. One suggested solution is particularly underscored by our results. Specifically, the investigators found that insufficient consult volume was a problem that contributed to program closure, and noted that in the setting of inadequate volume it may be difficult to maintain competency with technology and may also be difficult to justify the investment [15]. Likewise, we also found that EDs with pediatric telehealth services reported infrequent use in the majority of cases, with 77% of these EDs reporting use that was less than one time per month, on average, during the previous year. This contrasted with EDs with telestroke services where fewer than half reported such infrequent use. It is not surprising that suspected strokes are more common than sick children requiring telehealth consultation. Further, the framework of the technology acceptance model points to perceived usefulness as an important driver of telehealth's acceptance [24]. However, it is the very nature of the rarity of a critically ill pediatric patient that makes telehealth such a potentially effective tool. If an emergency physician in a relatively low-volume ED sees a critically ill child as an exceedingly rare event, then having the ability to connect with an expert consultant becomes that much more valuable. This is particularly true given that many EDs have been found to have critical deficiencies in pediatric emergency services [25-27]. Benefits may also be realized in the use of pediatric telehealth for less critically ill children. One recent study demonstrated a successful pediatric telehealth program in which the implementation was supported, adoption and use increased over time, and efficiency of health care resource use improved [28]. Our surveys did not capture how many times an ED reached out to either a pediatric critical care physician or a pediatric emergency medicine physician outside of a formal telehealth program. It may be possible in pediatric telehealth that EDs may not feel that the volume of sick children is enough for a pediatric telehealth subscription. It is likely that there was less of a desire to subscribe to such a program if prior to such a program offering they were able to connect with well-meaning pediatric acute care specialists who would advise the ED provider on the management of the patient over the phone.

Implications and Future Directions

Despite compelling examples of successful pediatric telehealth programs [13,29] and the endorsement of pediatric telehealth by the National Academy of Medicine as a solution to address

XSL•FO
RenderX

disparities in access to care [30], our results underscore the relatively infrequent use of pediatric telehealth services relative to telestroke services by EDs nationally prior to the COVID-19 pandemic. In 2016, only 339 EDs reported having pediatric telehealth services, as compared to 1758 EDs with telestroke services. We had hypothesized that the ability to obtain external certification or to improve performance on national quality metrics for stroke may have been an important driver in the significantly higher prevalence of telestroke services as compared to pediatric telehealth services. However, when examining the single most important reason for adoption, our results do not fully support that hypothesis. It may be that ED directors were not the appropriate source of this information and that a hospital-level financial administrator may have had more insight into the decision. An alternative explanation may be related to the nature of the typical telehealth consult for these conditions. For example, the typically envisioned telestroke consultation may be a patient with a potential stroke and an emergency physician looking for additional guidance, expertise, or shared liability in the decision to treat with thrombolytic therapy. In contrast, in the setting of a critically ill child in a remote ED, the emergency physician is often hoping to transfer the child as quickly as possible, rather than to delay transfer with a telehealth consultation [15]. In particular, given that small EDs may not have the ability to admit children to their hospital, many have no choice but to transfer these pediatric patients, and a pediatric consultation may be considered of lower value if patients are inevitably transferred.

In addition to differences in ED clinical care, the comparison of telestroke to pediatric telehealth services in EDs is an "apples-to-oranges" comparison in other ways as well. Indeed, these differences are an important part of what interested us in the comparison and in better understanding differential drivers of adoption. This includes differences in volume, in prehospital considerations, and in hospital financial motivations. Whereas EDs likely see consistent volumes of patients with stroke-like symptoms, a critically ill pediatric patient is a much more infrequent event, and such differences in patient volume may be an important factor contributing to different perceptions of the need for telestroke services as compared to pediatric telehealth services. Prehospital EMS patient triage may also contribute. An ED that requires telehealth services to connect with pediatric expertise would likely prefer that a critically ill child be transported directly to a higher-resourced center when possible. In contrast, prehospital considerations for patients with suspected stroke are much different. The time-dependent benefit of acute stroke interventions means that the closest capable hospital is considered the optimal transport destination [31]. Prehospital stroke triage policies vary by region; however, typically, in order to be considered a "capable" hospital, an ED needs to have access to neurologist expertise, generally reflected as stroke center certification. Often, the most cost-effective way to achieve 24/7 neurologist access and stroke center certification is through a telestroke program. In addition, while EDs may also receive a pediatric readiness score [26,32,33], this is not a certification process, per se. Some states do have their own pediatric emergency facility recognition programs [34,35]; however, these are not universal and we did not explicitly study whether the existence of these standards motivated pediatric telehealth adoption. Finally, once the patient with suspected stroke is transported to an ED, the hospital continues to have financial incentives to admit the patient rather than to transfer, as DRG billing for stroke care is generally favorable.

These differences in patient volume, in prehospital triage decisions, and in hospital financial incentives for disposition decisions were what motivated our research question and hypothesis. Nevertheless, they are also important reasons for telehealth adoption that our approach did not fully capture. Given previous research showing that cost was a major barrier for EDs without telehealth services [17], these differences in hospitals' anticipated return on investment for telestroke versus pediatric telehealth programs are important. One potential solution for these smaller or lower-volume EDs is to capitalize on economies of scale to facilitate implementation of telehealth. If the technology can be in place and shared among various applications, then the expense of implementation may be more justifiable, and the providers may be more able to maintain a baseline level of comfort and competence with the technology as well. Indeed, we found that relative to EDs with telestroke services alone or pediatric telehealth services alone, those EDs with both telestroke and pediatric telehealth services were smaller, lower-volume sites with fewer full-time equivalent attendings on staff. It may be that these economies of scale are already being realized by these EDs. This may be a good target for future research or future grant programs for smaller or less resourced hospitals.

Finally, in our recent experience during the COVID-19 pandemic, we have seen tremendous growth in the use of telehealth in health care, specifically emergency medicine [4-6]. Many of these changes were stimulated by clinical need in combination with changes in telehealth legislation and reimbursement policy. This highlights the interconnected relationship in which clinical need and policy changes worked together to increase adoption and use of the technology. This also underscores the importance of understanding drivers of telehealth adoption and use moving forward. By characterizing barriers and facilitators of telehealth in emergency care and understanding how these factors vary between clinical indications, we may be better equipped to ensure that EDs that benefit from the technology are able to continue to use telehealth to provide optimal clinical care. Future qualitative work may be of particular value to better understand these motivators.

Limitations

Our results predate the dramatic changes in telehealth that occurred during the pandemic and are likely not representative of current use or of what lies ahead. However, in many ways, the COVID-19 experience has generated new equivalents for the "geographic distance" that telehealth was previously bridging. With new limitations to access in more urban settings as well as clinical demands related to the pandemic, it will continue to be important to understand drivers of and barriers to adoption of telehealth use in emergency care. We believe that our findings remain relevant to that question. Our study has other limitations as well. Our results were self-reported by ED directors, and these individuals may not have had full insight into their hospitals' decisions to implement telehealth

XSL•FO
RenderX

technology; a survey targeted to a group with a different role may have produced different results. Survey responders may also have been confused about the definition of telehealth. While we were able to confirm and clarify this with the EDs responding by phone, those EDs responding by postal mail did not have this opportunity. It is also possible that our results were subject to selection bias related to nonresponders. However, response rates for our survey were generally high, and the characteristics of responding and nonresponding EDs were similar (Table S2 in Multimedia Appendix 1). It is also worth noting that EDs in our sample had a variety of staffing models, and the nature of pediatric telehealth use may vary based on the training of the clinician seeing the patient. For example, most emergency and family physicians are comfortable with routine pediatric conditions; however, physician assistants, nurse practitioners, or adult-trained physicians may be less so. Future work may further explore these differences. Finally, by asking about the single most important factor driving telehealth adoption, we cannot fully comment on whether policy is an important driver to motivate adoption. It may be necessary but not sufficient, or it may not be considered the single most important factor when the other response options also include "improving clinical care."

Conclusions

Prior to the COVID-19 pandemic, pediatric telehealth services were less common than telestroke services in US EDs. For both applications, the most frequently reported single most important reason for adoption was related to improving clinical care. Notably, EDs with pediatric telehealth services used the technology much less frequently than EDs with telestroke services. There may be value for smaller EDs to benefit from economies of scale in telehealth implementation in order to address disparities in access to care.

Acknowledgments

The authors would like to thank Ashley F Sullivan, MS, MPH, for her leadership of the EMNet Coordinating Center, Boston, Massachusetts, and Rachel D Freid, MPH; Rain E Freeman, MPH; and Darya M Herscovici for their help with data collection and management.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The 2016 National Emergency Department Inventory–USA (NEDI-USA), example follow-up survey, and supplemental tables.
[DOCX File , 1194 KB - jmir_v24i6e33981_app1.docx]

References

1. Purnell TS, Calhoun EA, Golden SH, Halladay JR, Krok-Schoen JL, Appelhans BM, et al. Achieving health equity: Closing the gaps in health care disparities, interventions, and research. Health Aff (Millwood) 2016 Aug 01;35(8):1410-1415. [doi: 10.1377/hlthaff.2016.0158] [Medline: 27503965]
2. Marcin JP, Shaikh U, Steinhorn RH. Addressing health disparities in rural communities using telehealth. Pediatr Res 2016 Jan;79(1-2):169-176. [doi: 10.1038/pr.2015.192] [Medline: 26466080]
3. Sanders JL, Raja AS, Hasegawa K, Bittner J, Espinola JA, Olamiju B, et al. Decline in consultant availability in Massachusetts emergency departments: 2005 to 2014. Ann Emerg Med 2016 Oct;68(4):461-466. [doi: 10.1016/j.annemergmed.2016.06.013] [Medline: 27569109]
4. Turer R, Jones I, Rosenbloom S, Slovis C, Ward M. Electronic personal protective equipment: A strategy to protect emergency department providers in the age of COVID-19. J Am Med Inform Assoc 2020 Jun 01;27(6):967-971 [FREE Full text] [doi: 10.1093/jamia/ocaa048] [Medline: 32240303]
5. Joshi AU, Lewiss RE. Telehealth in the time of COVID-19. Emerg Med J 2020 Oct;37(10):637-638. [doi: 10.1136/emermed-2020-209846] [Medline: 32753392]
6. Hollander J, Kvedar JC, Larson D. Handling the COVID-19 crisis: The wake-up call to revolutionize telemedicine [webinar]. WorkCast. New Rochelle, NY: Mary Ann Liebert, Inc, publishers; 2020 May 12. URL: https://www.workcast.com/register?cpak=6478263095456584 [accessed 2022-05-24]
7. Sauser-Zachrison K, Shen E, Sangha N, Ajani, Z, Neil WP, Gould MK, et al. Safe and effective implementation of telestroke in a US community hospital setting. Perm J 2016;20(4):15-217 [FREE Full text] [doi: 10.7812/tpp/15-217]
8. Wilcock AD, Schwamm LH, Zubizarreta JR, Zachrison KS, Uscher-Pines L, Richard JV, et al. Reperfusion treatment and stroke outcomes in hospitals with telestroke capacity. JAMA Neurol 2021 May 01;78(5):527-535 [FREE Full text] [doi: 10.1001/jamaneurol.2021.0023] [Medline: 33646272]
9. Adcock A, Choi J, Alvi M, Murray A, Seachrist E, Smith M, et al. Expanding acute stroke care in rural America: A model for statewide success. Telemed J E Health 2020 Jul;26(7):865-871 [FREE Full text] [doi: 10.1089/tmj.2019.0087] [Medline: 31596679]

XSL•FO
RenderX

10. Wechsler LR, Demaerschalk BM, Schwamm LH, Adeoye OM, Audebert HJ, Fanale CV, American Heart Association Stroke Council, Council on Epidemiology and Prevention, Council on Quality of Care and Outcomes Research. Telemedicine quality and outcomes in stroke: A scientific statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2017 Jan;48(1):e3-e25. [doi: 10.1161/STR.0000000000000114] [Medline: 27811332]
11. Zachrison KS, Boggs KM, Hayden EM, Espinola JA, Camargo CA. A national survey of telemedicine use by US emergency departments. J Telemed Telecare 2018 Dec 17;26(5):278-284. [doi: 10.1177/1357633x18816112]
12. Dharmar M, Romano PS, Kuppermann N, Nesbitt TS, Cole SL, Andrada ER, et al. Impact of critical care telemedicine consultations on children in rural emergency departments. Crit Care Med 2013 Oct;41(10):2388-2395. [doi: 10.1097/CCM.0b013e31828e9824] [Medline: 23921273]
13. Dharmar M, Kuppermann N, Romano PS, Yang NH, Nesbitt TS, Phan J, et al. Telemedicine consultations and medication errors in rural emergency departments. Pediatrics 2013 Dec;132(6):1090-1097. [doi: 10.1542/peds.2013-1374] [Medline: 24276844]
14. Lin CB, Peterson ED, Smith EE, Saver JL, Liang L, Xian Y, et al. Patterns, predictors, variations, and temporal trends in emergency medical service hospital prenotification for acute ischemic stroke. J Am Heart Assoc 2012 Aug 06;1(4):e002345. [doi: 10.1161/jaha.112.002345]
15. Uscher-Pines L, Kahn JM. Barriers and facilitators to pediatric emergency telemedicine in the United States. Telemed J E Health 2014 Nov;20(11):990-996 [FREE Full text] [doi: 10.1089/tmj.2014.0015] [Medline: 25238565]
16. Emergency Medicine Network (EMNet). URL: http://www.emnet-usa.org/ [accessed 2018-06-29]
17. Zachrison KS, Boggs KM, Hayden EM, Espinola JA, Camargo CA. Understanding barriers to telemedicine implementation in rural emergency departments. Ann Emerg Med 2020 Mar;75(3):392-399. [doi: 10.1016/j.annemergmed.2019.06.026] [Medline: 31474481]
18. Hayden EM, Boggs KM, Espinola JA, Camargo CA, Zachrison KS. Telemedicine facilitation of transfer coordination from emergency departments. Ann Emerg Med 2020 Nov;76(5):602-608 [FREE Full text] [doi: 10.1016/j.annemergmed.2020.04.027] [Medline: 32534835]
19. Weiner SG, Raja AS, Bittner JC, Curtis KM, Weimersheimer P, Hasegawa K, et al. Opioid-related policies in New England emergency departments. Acad Emerg Med 2016 Sep;23(9):1086-1090 [FREE Full text] [doi: 10.1111/acem.12992] [Medline: 27098615]
20. Directories. Society for Academic Emergency Medicine. URL: http://www.saem.org/resources/directories/residency-directory [accessed 2018-06-29]
21. Housing patterns and core-based statistical areas. United States Census Bureau. URL: https://www.census.gov/topics/housing/housing-patterns/about/core-based-statistical-areas.html [accessed 2020-06-08]
22. Trout KE, Rampa S, Wilson FA, Stimpson JP. Legal mapping analysis of state telehealth reimbursement policies. Telemed J E Health 2017 Oct;23(10):805-814. [doi: 10.1089/tmj.2017.0016] [Medline: 28430029]
23. Report archive. American Telemedicine Association. URL: https://www.americantelemed.org/resource_categories/report/ [accessed 2020-07-09]
24. Hu PJ, Chau PY, Sheng ORL, Tam KY. Examining the technology acceptance model using physician acceptance of telemedicine technology. J Manag Inf Syst 2015 Dec 02;16(2):91-112. [doi: 10.1080/07421222.1999.11518247]
25. Camargo CA, Boggs KM, Sullivan AF, Gutierrez CE, Petrack EM. Grassroots intervention to increase appointment of pediatric emergency care coordinators in Massachusetts emergency departments. Acad Emerg Med 2018 Dec;25(12):1442-1446 [FREE Full text] [doi: 10.1111/acem.13630] [Medline: 30311325]
26. Gausche-Hill M, Ely M, Schmuhl P, Telford R, Remick KE, Edgerton EA, et al. A national assessment of pediatric readiness of emergency departments. JAMA Pediatr 2015 Jun;169(6):527-534. [doi: 10.1001/jamapediatrics.2015.138] [Medline: 25867088]
27. Sullivan AF, Rudders SA, Gonsalves AL, Steptoe AP, Espinola JA, Camargo CA. National survey of pediatric services available in US emergency departments. Int J Emerg Med 2013 Apr 24;6(1):13 [FREE Full text] [doi: 10.1186/1865-1380-6-13] [Medline: 23618163]
28. Foster CC, Macy ML, Simon N, Stephen R, Lehnig K, Bohling K, et al. Emergency Care Connect: Extending pediatric emergency care expertise to general emergency departments through telemedicine. Acad Pediatr 2020 Jul;20(5):577-584. [doi: 10.1016/j.acap.2020.02.028] [Medline: 32112864]
29. Li J, Monuteaux M, Bachur R. Interfacility transfers of noncritically ill children to academic pediatric emergency departments. Pediatrics 2012 Jul;130(1):83-92. [doi: 10.1542/peds.2011-1819] [Medline: 22665410]
30. Board on Health Care Services, Institute of Medicine. The Role of Telehealth in an Evolving Health Care Environment: Workshop Summary. Washington, DC: National Academies Press; 2012.
31. Powers WJ, Rabinstein AA, Ackerson T, Adeoye OM, Bambakidis NC, Becker K, et al. Guidelines for the Early Management of Patients With Acute Ischemic Stroke: 2019 update to the 2018 Guidelines for the Early Management of Acute Ischemic Stroke: A guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2019 Dec;50(12):e344-e418 [FREE Full text] [doi: 10.1161/STR.0000000000000211] [Medline: 31662037]

XSL•FO
RenderX

32. Remick K, Kaji AH, Olson L, Ely M, Schmuhl P, McGrath N, et al. Pediatric readiness and facility verification. Ann Emerg Med 2016 Mar;67(3):320-328.e1. [doi: 10.1016/j.annemergmed.2015.07.500] [Medline: 26320519]
33. Rice A, Dudek J, Gross T, St Mars T, Woolridge D. The impact of a pediatric emergency department facility verification system on pediatric mortality rates in Arizona. J Emerg Med 2017 Jun;52(6):894-901. [doi: 10.1016/j.jemermed.2017.02.011] [Medline: 28341087]
34. Cicero MX, Walsh B, Solad Y, Whitfill T, Paesano G, Kim K, et al. Do you see what I see? Insights from using Google Glass for disaster telemedicine triage. Prehosp Disaster Med 2015 Jan 09;30(1):4-8. [doi: 10.1017/s1049023x1400140x]
35. Ball JW, Sanddal ND, Mann NC, Esposito T, Nadkarni M, Wilkins G, et al. Emergency department recognition program for pediatric services: Does it make a difference? Pediatr Emerg Care 2014 Sep;30(9):608-612. [doi: 10.1097/PEC.0000000000000205] [Medline: 25162686]

Abbreviations

DRG: diagnosis-related group
ED: emergency department
EMNet: Emergency Medicine Network
EMS: emergency medical services
NEDI-USA: National Emergency Department Inventory–USA
REDCap: Research Electronic Data Capture
telestroke: telehealth for emergency stroke care delivery

Edited by T Leung; submitted 01.10.21; peer-reviewed by T Ndabu, J Sorra, J Colquitt, Jr, H Ayatollahi; comments to author 05.01.22; revised version received 25.03.22; accepted 09.05.22; published 20.06.22.

Please cite as:
Zachrison KS, Hayden EM, Boggs KM, Boyle TP, Gao J, Samuels-Kalow ME, Marcin JP, Camargo Jr CA
Emergency Departments' Uptake of Telehealth for Stroke Versus Pediatric Care: Observational Study
J Med Internet Res 2022;24(6):e33981
URL: https://www.jmir.org/2022/6/e33981
doi:10.2196/33981
PMID:35723927

©Kori S Zachrison, Emily M Hayden, Krislyn M Boggs, Tehnaz P Boyle, Jingya Gao, Margaret E Samuels-Kalow, James P Marcin, Carlos A Camargo Jr. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.06.2022. 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 Indicators as Measures of Individual Health Status and Their Public Perspectives: Cross-sectional Survey Study

Temiloluwa Sokoya[1], MPH; Yuchun Zhou[2], PhD; Sebastian Diaz[3], PhD, JD; Timothy Law[4], DO; Lina Himawan[5], MSc; Francisca Lekey[6], MPH; Lu Shi[7], PhD; Ronald W Gimbel[7], PhD; Xia Jing[7], MD, PhD

[1]Department of Behavioral and Environmental Health, College of Health Sciences, Jackson State University, Jackson, MS, United States

[2]Gladys W and David H Patton College of Education, Ohio University, Athens, OH, United States

[3]College of Medicine, Northeast Ohio Medical University, Rootstown, OH, United States

[4]Ohio Musculoskeletal and Neurological Institute, Ohio University, Athens, OH, United States

[5]Department of Psychology, College of Arts and Sciences, Ohio University, Athens, OH, United States

[6]College of Health Sciences and Professions, Ohio University, Athens, OH, United States

[7]Department of Public Health Sciences, College of Behavioral, Social, and Health Sciences, Clemson University, Clemson, SC, United States

Corresponding Author:
Xia Jing, MD, PhD
Department of Public Health Sciences
College of Behavioral, Social, and Health Sciences
Clemson University
511 Edwards Hall
Clemson, SC, 29634
United States
Phone: 1 8646563347
Email: xia.xjing@gmail.com

Abstract

Background: Disease status (eg, cancer stage) has been used in routine clinical practice to determine more accurate treatment plans. Health-related indicators, such as mortality, morbidity, and population group life expectancy, have also been used. However, few studies have specifically focused on the comprehensive and objective measures of individual health status.

Objective: The aim of this study was to analyze the perspectives of the public toward 29 health indicators obtained from a literature review to provide evidence for further prioritization of the indicators. The difference between health status and disease status should be considered.

Methods: This study used a cross-sectional design. Online surveys were administered through Ohio University, ResearchMatch, and Clemson University, resulting in three samples. Participants aged 18 years or older rated the importance of the 29 health indicators. The rating results were aggregated and analyzed as follows (in each case, the dependent variables were the individual survey responses): (1) to determine the agreement among the three samples regarding the importance of each indicator, where the independent variables (IVs) were the three samples; (2) to examine the mean differences between the retained indicators with agreement across the three samples, where the IVs were the identified indicators; and (3) to rank the groups of indicators into various levels after grouping the indicators with no mean differences, where the IVs were the groups of indicators.

Results: In total, 1153 valid responses were analyzed. Descriptive statistics revealed that the top five–rated indicators were drug or substance abuse, smoking or tobacco use, alcohol abuse, major depression, and diet and nutrition. Among the 29 health indicators, the three samples agreed upon the importance of 13 indicators. Inferential statistical analysis indicated that some of the 13 indicators held equal importance. Therefore, the 13 indicators were categorized by rank into seven levels: level 1 included blood sugar level and immunization and vaccination; level 2 included LDL cholesterol; level 3 included HDL cholesterol, blood triglycerides, cancer screening detection, and total cholesterol; level 4 included health literacy rate; level 5 included personal care needs and air quality index greater than 100; level 6 included self-rated health status and HIV testing; and level 7 included the supply of dentists. Levels 1 to 3 were rated significantly higher than levels 4 to 7.

Conclusions: This study provides a baseline for prioritizing 29 health indicators, which can be used by electronic health record or personal health record system designers or developers to determine what can be included in the systems to capture an individual's health status. Currently, self-rated health status is the predominantly used health indicator. Additionally, this study provides a

XSL•FO
RenderX

foundation for tracking and measuring preventive health care services more accurately and for developing an individual health status index.

(J Med Internet Res 2022;24(6):e38099) doi:10.2196/38099

KEYWORDS

health status measurement; individual health indicators; public perspectives; surveys and questionnaires

Introduction

Disease status, such as cancer stage, has frequently been used in routine clinical practice to determine more accurate treatment plans. Health-related indicators, such as mortality, morbidity, and life expectancy for a given population group, have also been used. Few studies, however, have focused on more comprehensive and objective measures of an individual's health status. Self-rated health status has previously been identified as a reliable indicator of an individual's overall health status [1,2], but this is subjective and the rating criteria are unclear. Although there is research on health indicators used for the measurement of care quality [3], as well as social and behavioral measures in electronic health record (EHR) systems [4,5], more comprehensive and objective health indicators of an individual's health status are lacking. These must be developed and used to measure health status more accurately, objectively, and consistently, as well as to evaluate the outcomes of preventive medicine services [1,6]. As the health care paradigm shifts from treatment to prevention [7,8], the accurate, objective, and convenient measurement of preventive services and their long-term outcomes becomes an urgent and growing need.

Individual health status refers to a person's overall physical, mental, and social well-being, as well as freedom from illness or injury. In contrast, individual disease status refers to a person's physical or mental symptoms with or without diagnosis [9]. Accurate individual health status measures can guide customized preventive medicine services and lifestyle suggestions and be applied to population health programs by aggregating an individual's health data into meaningful groups. Chronic diseases are increasingly costly to both patients and society, and most can be prevented or delayed via preventive medicine services. These services need to be provided in a routine and consistent manner [7,8], thus maximizing the potential to control health care costs.

The Institute of Medicine reviewed social and behavioral domain measures, as seen in EHR systems [4,5]. They identified 17 domains, and these measurements were used as a foundation for the Office of the National Coordinator for Health Information Technology's Meaningful Use of EHRs reporting requirements [4,5]. In 2015, the Centers for Disease Control and Prevention's National Center for Health Statistics released 15 selected health indicators based on the National Health Interview Survey [10]. Other research [1,2,6,11] also considered health indicators, although none focused on comprehensive, objective measures of an individual's health status.

Although preventive medicine has been recognized for its critical role in health care, such services are not provided consistently to the majority of the American population [12]. Because chronic diseases represent a large portion of health care expenditures, it is critical to prevent or delay the onset of chronic diseases via preventive services [13]. The tracking of health indicators has been reported to help policy makers note changes needed in coverage and to influence policy decisions [14]. Such tracking also enables governments to better allocate health resources [14]. Nevertheless, accurate measurements of preventive services are inadequate or lacking.

We conducted a literature review of existing health indicators [1,2,6,11,12,15-17]. We consolidated the described health indicators and determined that 29 health indicators could be used to measure an individual's health. We then examined four commercial EHR systems used in rural, primary care, and ambulatory settings to explore the availability and presentation of these indicators as a pilot study. None of these systems were found to capture *all* of the indicators [9], but each system provided data on some available health indicators, and all four systems had preventive medicine panels. The pilot study indicated that no established group of health indicators existed for individuals, nor were indicators incorporated or used consistently and routinely across different EHR systems. Therefore, there is at least a need to provide more evidence for these health indicators. This would constitute what should be included among the individual health indicators used by EHR systems and whether these indicators can be prioritized based on their importance. Additionally, we recognize that these health indicators have much broader potential use beyond incorporating them into EHR [9].

This study aimed to examine public perspectives on the importance of 29 health indicators to inform their relative perceived priority. This would, for example, allow the separation of the health indicators into core and secondary sets, which could be incorporated into the EHR or similar systems [18]. Such health indicators could capture an individual's health status, thus informing and enabling preventive services to make them more accurate, consistent, and convenient without overburdening providers' data collection workload. These public perspectives could also provide a foundation for developing an individual health index, which could be used to stratify healthy populations into subgroups based on the corresponding study requirements. There is no established list or ranking of health indicators according to importance, nor is there a well-established methodology to develop such a list. Therefore, we attempted to use public perspective surveys as a starting point in this study. We plan to validate the results quantitatively through longitudinal health record analysis in a future study. The assumption is that an individual's perception of the importance of each health indicator may be associated with their conscious or unconscious behaviors, which would ultimately affect health outcomes. This paper focuses on the public

XSL•FO
RenderX

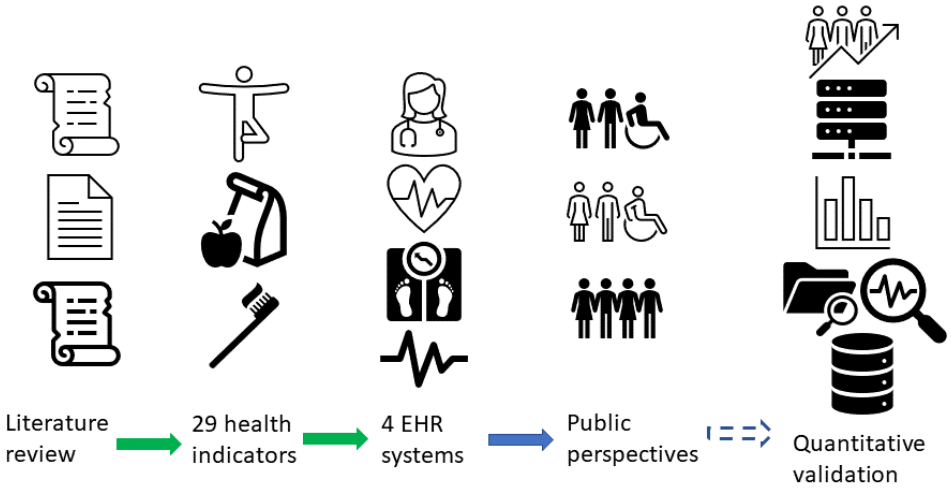
perspectives, the approach, the results, and the analysis of the results.

Methods

General Study Design

The overall design of this project is illustrated in Figure 1 to provide context for this paper. The work reported here focuses on public perspectives, the methods used, and the results. The first three steps in Figure 1 have been completed and published [9,18]. The section on the far-right side of the figure, regarding quantitative validation, illustrates directions we plan to explore in future studies.

Figure 1. The overall design of the project; public perspectives are the focus of this paper. The three sections connected via green arrows have been completed, and the far-right section is for future work. EHR: electronic health record.



Data Collection

An online survey (Multimedia Appendix 1) was administered through Ohio University in the summer of 2017, through ResearchMatch [19] in the summer of 2018, and through Clemson University in the summer of 2020, providing three samples. The inclusion criterion for participation in the survey was being 18 years of age or older. The participants were allowed to share the survey's URL link, and all respondents acknowledged informed consent.

The survey included seven demographic questions and rating items related to the importance of the 29 health indicators. Definitions of these health indicators were provided within the survey (Multimedia Appendix 2). In the survey, the 29 health indicators were separated into five subscales [1,2,17]:

1. Health risks and behaviors, with eight indicators: alcohol abuse, BMI, diet and nutrition, drug or substance abuse, family history of cancer, physical inactivity, smoking or tobacco use, and sun protection.
2. Health care, with three indicators: immunization and vaccination, insurance coverage, and personal care needs.
3. Health care provider supply, with three indicators: cancer screening detection, hypertension screening, and HIV testing.
4. Blood tests in physical exams, with five indicators: blood sugar level, blood triglycerides, high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, and total cholesterol.
5. Other health indicators, with 10 indicators: self-rated health status, high school diploma, air quality index greater than 100, supply of dentists, engagement in life, health literacy rate, major depression, having a sense of purpose in one's life, race and ethnicity, and being unemployed.

After removing invalid data, the final sample yielded 362 responses from Ohio University, 694 from ResearchMatch, and 97 from Clemson University (Multimedia Appendix 3). Survey items used by Ohio University and Clemson University were rated on a scale of 0 to 10, where 0 refers to "not at all important" and 10 refers to "extremely important." Survey items in the ResearchMatch sample were measured using a scale of 0 to 100, where 0 refers to "not at all important" and 100 refers to "extremely important." Therefore, as part of the data cleaning process, the data from ResearchMatch were converted to a scale of 0 to 10 (Multimedia Appendix 4 contains the codebook). In the Ohio University survey, there were five health indicators—blood sugar, blood triglycerides, HDL, LDL, and total cholesterol—for which a scale of 0 to 11, instead of 0 to 10, was used. Due to this error, data for these five indicators were removed from the Ohio University data set. As a result, the total sample size of these five indicators was 791 respondents, whereas the total sample size of the other indicators was 1153 respondents.

The internal reliability of the survey instruments was calculated for the overall set and the three institutional subsets using Cronbach α [20].

Data Analytic Strategies

Data analyses included rating the 29 health indicators based on their perceived importance. After aggregating data from the three samples with descriptive statistics, a three-step analysis was conducted. The first step of the analysis involved determining whether the three samples had unanimous

XSL•FO
RenderX

agreement on the importance of each indicator. A one-way analysis of variance (ANOVA) with a post hoc test was conducted using SPSS software (version 27; IBM Corp) for each indicator to examine any group mean difference, where the independent variables were the three samples and the dependent variables were the individual survey responses. A Levene test was used to test the homogeneity of variance for each indicator before running an ANOVA. The indicators with no group mean differences across samples were retained for the following analysis step.

The second step of the analysis examined the mean differences between the retained indicators via a one-way ANOVA, where the independent variables were the identified indicators and the dependent variables were the individual survey responses. Any indicators with no significant mean differences were grouped together because they could not be ranked.

The third step of the analysis ranked the groups of indicators into various levels after grouping the indicators with no mean differences. A one-way ANOVA with a post hoc test was conducted to examine the mean differences between the levels of indicators, where the independent variables were the levels of indicators and the dependent variables were the individual survey responses. Any significant mean difference between any two levels of indicators indicated the ranking order of the two levels.

Ethics Approval

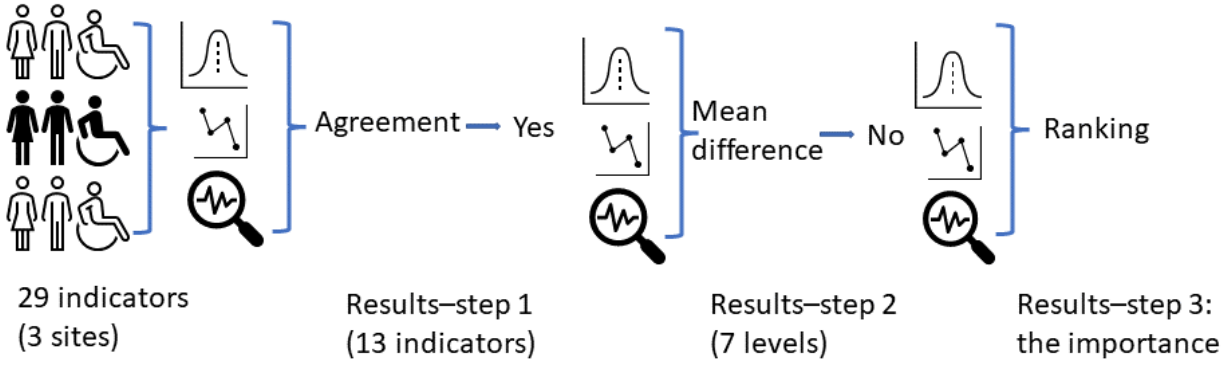
This study was approved by the Institutional Review Boards of Ohio University (17-X-142) and Clemson University (IRB2019-441).

Results

The Overall Layout of the Findings

The primary purpose of this study was to identify the public perspectives on the importance of the 29 selected health indicators, whether they agreed with each other, and, if so, what were the importance rankings of the health indicators that agreed with each other. Figure 2 summarizes the analytic strategies and the main results of each step. The following paragraphs elaborate on the detailed results for each step.

Figure 2. The primary analytic strategies and overall results of each step.



Results of Descriptive Statistics

Descriptive statistics for the 29 health indicators are reported in Table 1. The descriptive statistics for the demographic information for all respondents are reported in Multimedia Appendix 5. Descriptive analyses show that drug and substance abuse, smoking and tobacco use, alcohol abuse, major depression, and diet and nutrition were found to be the five most important health indicators, as rated by the study participants. Additionally, race and ethnicity, possession of a high school diploma, engagement in life, unemployment status, and sun protection were the five least important health indicators. Self-rated health status, the most-used health indicator to assess an individual's health status, was ranked in the 20th position.

XSL•FO
RenderX

Table 1. Descriptive statistics for all 29 health indicators.

Health indicators	ResearchMatch (n=694)		Ohio University (n=362)		Clemson University (n=97)		Total (N=1153)	
	Score[a], mean (SD)	n (%)	Score[a], mean (SD)	n (%)	Score[a], mean (SD)	n (%)	Score[a], mean (SD)	n (%)
Drug or substance abuse	8.75 (1.5)	694 (100)	8.13 (1.96)	362 (100)	8.36 (1.87)	97 (100)	8.53 (1.71)	1153 (100)
Smoking and tobacco use	8.8 (1.52)	694 (100)	8.02 (2.06)	362 (100)	8.18 (1.84)	97 (100)	8.5 (1.77)	1153 (100)
Alcohol abuse	8.34 (1.71)	694 (100)	7.56 (2.03)	362 (100)	8.06 (1.64)	97 (100)	8.07 (1.84)	1153 (100)
Major depression	8.1 (1.6)	685 (98.7)	7.79 (1.94)	362 (100)	8.03 (1.57)	97 (100)	7.99 (1.72)	1144 (99.2)
Diet and nutrition	8.01 (1.58)	694 (100)	7.8 (1.93)	362 (100)	8.36 (1.65)	97 (100)	7.97 (1.71)	1153 (100)
Blood sugar level	7.76 (1.63)	694 (100)	N/A[b]	N/A	7.59 (1.75)	97 (100)	7.74 (1.65)	791 (68.6)
Physical inactivity	7.9 (1.68)	694 (100)	7.41 (2.13)	362 (100)	7.68 (1.77)	97 (100)	7.73 (1.85)	1153 (100)
Immunization and vaccination	7.49 (2.12)	694 (100)	7.67 (2.3)	362 (100)	7.72 (2.4)	97 (100)	7.57 (2.2)	1153 (100)
Hypertension screening	7.59 (1.91)	694 (100)	7.17 (2.29)	362 (100)	7.42 (2.03)	97 (100)	7.45 (2.05)	1153 (100)
LDL[c] cholesterol	7.43 (1.85)	694 (100)	N/A	N/A	7.56 (1.91)	97 (100)	7.45 (1.86)	791 (68.6)
Blood triglycerides	7.32 (1.78)	694 (100)	N/A	N/A	7.34 (1.95)	97 (100)	7.32 (1.80)	791 (68.6)
HDL[d] cholesterol	7.31 (1.83)	694 (100)	N/A	N/A	7.43 (1.91)	97 (100)	7.32 (1.84)	791 (68.6)
Having a sense of purpose in one's life	7.59 (1.94)	685 (98.7)	6.67 (2.53)	362 (100)	7.88 (1.93)	97 (100)	7.32 (2.19)	1144 (99.2)
Cancer screening detection	7.22 (2.06)	694 (100)	7.26 (2.3)	362 (100)	7.49 (2.09)	97 (100)	7.25 (2.14)	1153 (100)
Total cholesterol	7.2 (2.02)	694 (100)	N/A	N/A	7.6 (1.85)	97 (100)	7.25 (2.00)	791 (68.6)
Health literacy rate	6.99 (2.02)	685 (98.7)	7.06 (2.26)	362 (100)	7.34 (2.01)	97 (100)	7.04 (2.10)	1144 (99.2)
Personal care needs	6.82 (2.08)	694 (100)	7.01 (2.3)	362 (100)	7.21 (2.1)	97 (100)	6.91 (2.16)	1153 (100)
Air quality index >100	6.74 (1.92)	685 (98.7)	6.76 (2.13)	362 (100)	6.89 (1.93)	97 (100)	6.76 (1.99)	1144 (99.2)
Family history of cancer	6.98 (2.06)	694 (100)	6.37 (2.24)	362 (100)	6.25 (1.98)	97 (100)	6.73 (2.13)	1153 (100)
Self-rated health status	6.63 (2.2)	694 (100)	6.62 (2.15)	362 (100)	6.92 (1.89)	97 (100)	6.65 (2.16)	1153 (100)
HIV testing	6.62 (2.36)	694 (100)	6.62 (2.64)	362 (100)	6.84 (2.37)	97 (100)	6.64 (2.45)	1153 (100)
Insurance coverage	6.4 (2.88)	694 (100)	6.79 (2.91)	362 (100)	7.26 (2.51)	97 (100)	6.6 (2.87)	1153 (100)
BMI	6.86 (2.28)	694 (100)	5.8 (2.54)	362 (100)	6.64 (2.45)	97 (100)	6.51 (2.42)	1153 (100)
Supply of dentists	6.53 (2.02)	685 (98.7)	6.34 (2.26)	362 (100)	6.04 (1.99)	97 (100)	6.43 (2.10)	1144 (99.2)
Sun protection	6.63 (2)	694 (100)	5.73 (2.3)	362 (100)	5.54 (2.18)	97 (100)	6.25 (2.16)	1153 (100)
Unemployed individual	6.07 (2.34)	685 (98.7)	5.52 (2.68)	362 (100)	6.2 (2.69)	97 (100)	5.91 (2.49)	1144 (99.2)
Engagement in life	6.38 (2.18)	685 (98.7)	4.82 (2.87)	362 (100)	6.4 (2.33)	97 (100)	5.89 (2.54)	1144 (99.2)
High school diploma as a health indicator	5.02 (2.57)	694 (100)	6.04 (3.07)	362 (100)	5.56 (2.75)	97 (100)	5.38 (2.79)	1153 (100)
Race and ethnicity	5.28 (2.53)	685 (98.7)	4.32 (2.76)	362 (100)	5.02 (2.85)	97 (100)	4.96 (2.67)	1144 (99.2)

[a]Items were rated on a scale of 0 to 10, where 0 refers to "not at all important" and 10 refers to "extremely important."

[b]N/A: not applicable; data for this indicator were removed from the Ohio University data set because a scale of 0 to 11 vs 0 to 10 was used.

[c]LDL: low-density lipoprotein.

[d]HDL: high-density lipoprotein.

Results of Inferential Statistics

A Levene test was conducted to test the homogeneity of variance for each indicator before running an ANOVA. This resulted in nine health indicators with homogenous variance: blood sugar level, HDL cholesterol, LDL cholesterol, total cholesterol, immunization and vaccination, insurance coverage, cancer screening detection, air quality index greater than 100, and self-rated health status (Multimedia Appendix 6). A total of 20 health indicators were found to have heterogeneous variance. These included the following indicators: blood triglycerides, alcohol abuse, BMI, diet and nutrition, drug or substance abuse, family history of cancer, physical inactivity, smoking and

XSL•FO
RenderX

tobacco use, sun protection, personal care needs, hypertension screening, HIV testing, high school diploma as a health indicator, supply of dentists, engagement in life, health literacy rate, major depression, having a sense of purpose in one's life, race and ethnicity, and unemployment (Multimedia Appendix 7).

For the nine indicators with homogeneous variance, a one-way ANOVA was used. A one-way ANOVA Welch test was used for the 20 indicators with heterogeneous variance. As a result, 13 indicators were found to have no statistically significant mean differences among the three samples. This indicates that survey participants generally agreed on the relative level of importance of these indicators (Table 2). The 13 indicators were blood sugar level, blood triglycerides, HDL cholesterol, LDL cholesterol, total cholesterol, personal care needs, HIV testing, self-rated health status, supply of dentists, health literacy rate, immunization and vaccination, cancer screening detection, and air quality index greater than 100. The means and SDs of their ratings are presented in Table 2. These 13 indicators were retained for the second step of the analysis. Significant mean differences were found among the other 16 indicators, which indicates that survey participants disagreed on their level of importance (Multimedia Appendix 8 contains the post hoc results).

In the second step of the analysis, a one-way ANOVA was run for the 13 retained indicators, where the independent variables were the 13 indicators and the dependent variables were the individual survey responses. The indicators with no mean differences were grouped into the same level (Table 3) because they were rated as equally important and could not be ranked within a level. As a result, seven levels were formed (Table 3). Level 1 to level 7 rankings were organized based on the mean importance of the health indicators from high to low within and between levels. Level 1 included blood sugar level and immunization and vaccination; level 2 included LDL cholesterol; level 3 included HDL cholesterol, blood triglycerides, cancer screening detection, and total cholesterol; level 4 included health literacy rate; level 5 included personal care needs and air quality index greater than 100; level 6 included self-rated health status and HIV testing; and level 7 included the supply of dentists.

In the third step of the analysis, a one-way ANOVA was used to rank the seven levels of indicators, where the independent variables were the seven levels and the dependent variables were the individual survey responses. There were seven indicators in levels 1 to 3: blood sugar level, immunization and vaccination, LDL cholesterol, HDL cholesterol, blood triglycerides, cancer screening detection, and total cholesterol. These indicators were found to be significantly more important to the survey participants than the six indicators ranked in levels 4 to 7: health literacy rate, personal care needs, air quality index greater than 100, self-rated health status, HIV testing, and supply of dentists (Table 4).

Among the more important indicators, the two indicators in level 1 (ie, blood sugar level and immunization and vaccination) were rated as significantly more important than the four indicators in level 3 (ie, HDL cholesterol, blood triglycerides, cancer screening detection, and total cholesterol). Therefore, based on the surveys and our analysis results, among these 13 agreeable health indicators, blood sugar level, and immunization and vaccination were the most important, and the perspectives of the participants were agreed upon across all three samples.

Furthermore, among the less important indicators, the indicator assigned to level 4 (ie, health literacy rate) was found to be significantly more important than the two indicators in level 6 (ie, self-rated health status and HIV testing) and the indicator assigned to level 7 (ie, supply of dentists). Additionally, the two indicators assigned to level 5 (ie, air quality index >100 and personal care needs) were found to be significantly more important than the indicator assigned to level 7 (ie, supply of dentists). Therefore, the survey and analysis results showed that the supply of dentists was the least important among the 13 health indicators, with agreed-upon perspectives across the three samples. Additionally, the inferential statistical test results among the levels provided more confidence in ranking the seven levels from the most important (ie, blood sugar level and immunization and vaccination) to the least important (ie, supply of dentists). The statistical significance test results among the levels provided evidence for prioritizing the 13 health indicators.

XSL•FO
RenderX

Table 2. The 13 indicators with nonsignificant mean differences across the three samples.

Health indicator and sources	Responses[a], n (%)	Score[b], mean (SD)	*P* value[c]
Blood sugar level[d]			
ResearchMatch (n=694)	694 (100)	7.756 (1.6303)	.35
Clemson University (n=97)	97 (100)	7.588 (1.7485)	
Blood triglycerides[d]			
ResearchMatch (n=694)	694 (100)	7.318 (1.7786)	.91
Clemson University (n=97)	97 (100)	7.34 (1.952)	
HDL[e] cholesterol[d]			
ResearchMatch (n=694)	694 (100)	7.307 (1.8264)	.53
Clemson University (n=97)	97 (100)	7.433 (1.9143)	
LDL[f] cholesterol[d]			
ResearchMatch (n=694)	694 (100)	7.43 (1.8489)	.53
Clemson University (n=97)	97 (100)	7.557 (1.9147)	
Total cholesterol[d]			
ResearchMatch (n=694)	694 (100)	7.203 (2.0177)	.07
Clemson University (n=97)	97 (100)	7.598 (1.8465)	
Personal care needs			
ResearchMatch (n=694)	694 (100)	6.816 (2.0786)	.14
Ohio University (n=362)	362 (100)	7.011 (2.3026)	
Clemson University (n=97)	97 (100)	7.206 (2.1013)	
HIV testing			
ResearchMatch (n=694)	694 (100)	6.616 (2.361)	.69
Ohio University (n=362)	362 (100)	6.619 (2.6439)	
Clemson University (n=97)	97 (100)	6.835 (2.3747)	
Self-rated health status			
ResearchMatch (n=694)	694 (100)	6.63 (2.2032)	.45
Ohio University (n=362)	362 (100)	6.619 (2.1504)	
Clemson University (n=97)	97 (100)	6.918 (1.8912)	
Supply of dentists			
ResearchMatch (n=694)	685 (98.7)	6.525 (2.0215)	.07
Ohio University (n=362)	362 (100)	6.34 (2.2572)	
Clemson University (n=97)	97 (100)	6.041 (1.9944)	
Health literacy rate			
ResearchMatch (n=694)	685 (98.7)	6.986 (2.0199)	.26
Ohio University (n=362)	362 (100)	7.061 (2.2617)	
Clemson University (n=97)	97 (100)	7.34 (2.0098)	
Immunization and vaccination			
ResearchMatch (n=694)	694 (100)	7.494 (2.1184)	.37
Ohio University (n=362)	362 (100)	7.666 (2.3041)	
Clemson University (n=97)	97 (100)	7.722 (2.3968)	
Cancer screening detection			
ResearchMatch (n=694)	694 (100)	7.217 (2.0625)	.52

XSL•FO
RenderX

Health indicator and sources	Responses[a], n (%)	Score[b], mean (SD)	*P* value[c]
Ohio University (n=362)	362 (100)	7.257 (2.3045)	
Clemson University (n=97)	97 (100)	7.485 (2.0922)	
Air quality index >100			
ResearchMatch (n=694)	685 (98.7)	6.736 (1.9232)	.78
Ohio University (n=362)	362 (100)	6.76 (2.125)	
Clemson University (n=97)	97 (100)	6.887 (1.9304)	

[a]The independent variables were the three samples and the dependent variables were the individual survey responses.

[b]Items were rated on a scale of 0 to 10, where 0 refers to "not at all important" and 10 refers to "extremely important."

[c]*P* values were based on analysis of variance or *t* test results for each health indicator among three samples; they are reported in the top row for each group.

[d]The Ohio University data set for this indicator was removed because a scale of 0 to 11 vs 0 to 10 was used.

[e]HDL: high-density lipoprotein.

[f]LDL: low-density lipoprotein.

Table 3. The seven levels of health indicators with no significant mean differences within levels.

Level and health indicators with no group mean differences[a]	Individual survey data[b]		
	Responses, n (%)	Score[c], mean (SD)	*P* value[d]
Level 1			
Blood sugar	791 (68.6)	7.74 (1.65)	.053
Immunization and vaccination	1153 (100)	7.57 (2.20)	
Level 2			
LDL[e] cholesterol	791 (68.6)	7.45 (1.86)	N/A[f]
Level 3			
HDL[g] cholesterol	791 (68.6)	7.32 (1.84)	.77
Blood triglycerides	791 (68.6)	7.32 (1.80)	
Cancer screening detection	1153 (100)	7.25 (2.14)	
Total cholesterol	791 (68.6)	7.25 (2.00)	
Level 4			
Health literacy rate	1144 (99.2)	7.04 (2.10)	N/A
Level 5			
Personal care needs	1153 (100)	6.91 (2.16)	.08
Air quality index >100	1144 (99.2)	6.76 (1.99)	
Level 6			
Self-rated health status	1153 (100)	6.65 (2.16)	.87
HIV testing	1153 (100)	6.64 (2.45)	
Level 7			
Supply of dentists	1144 (99.2)	6.43 (2.10)	N/A

[a]Independent variables were the indicators in each level.

[b]Dependent variables were the individual survey responses.

[c]Items were rated on a scale of 0 to 10, where 0 refers to "not at all important" and 10 refers to "extremely important."

[d]*P* values indicate whether mean differences existed among the indicators within each level based on analysis of variance or *t* test results, and are reported in the top row of each group.

[e]LDL: low-density lipoprotein.

[f]N/A: not applicable; no comparison was conducted because the row has only one health indicator.

[g]HDL: high-density lipoprotein.

XSL•FO
RenderX

Table 4. Analysis of variance post hoc test results for the seven levels of indicators.

Health indicator level[a]	Health indicator level	*P* value
Level 1	Level 2	.58
Level 1	Level 3	<.001
Level 1	Level 6	<.001
Level 1	Level 7	<.001
Level 2	Level 4	.006
Level 2	Level 3	.68
Level 2	Level 5	<.001
Level 2	Level 7	<.001
Level 3	Level 4	.06
Level 3	Level 5	<.001
Level 4	Level 5	.27
Level 4	Level 1	<.001
Level 4	Level 6	<.001
Level 5	Level 6	.14
Level 5	Level 1	<.001
Level 5	Level 7	<.001
Level 6	Level 7	.28
Level 6	Level 2	<.001
Level 6	Level 3	<.001
Level 7	Level 2	<.001
Level 7	Level 3	<.001
Level 7	Level 4	<.001

[a]The independent variables were the levels of indicators and the dependent variables were the individual survey responses.

Reliability of Survey Instruments

The 29 items from the survey instruments showed good levels of internal reliability (Cronbach α=.912), as did each of the three subsets related to institutions where the survey was administered (Table 5). Instruments with Cronbach α values equal to or higher than .7 are generally considered to be reliable [20].

Table 5. Reliability of the survey instrument.

Survey components and data analyzed	Cronbach α
Entire survey (all items)	
All three samples	.912
ResearchMatch	.922
Ohio University	.893
Clemson University	.925
Survey subscales	
Health risk and behavior indicators	.795
Health care	.613
Health care provider supply	.831
Blood tests in physical exams	.934
Other health indicators	.823

XSL•FO
RenderX

Discussion

Principal Findings

Among all three samples, the ranking of the importance of 13 out of 29 (45%) health indicators showed agreement (Table 3). However, these health indicators were not necessarily more important than the other 16; instead, participants were observed to have perceived importance more consistently among these 13 health indicators. When we compared the 13 health indicators (Table 3) and their corresponding rankings in Table 1, we noticed that the 13 health indicators were placed between the 6th and 24th rankings in Table 1. This indicated more agreement among participants regarding the middle-ranked health indicators than the higher- or lower-ranked ones. The perspectives were more heterogeneous for the higher- or lower-ranked health indicators. Noticeably, the currently widely used standard individual health indicator, self-rated health status, was ranked 20th based on the results of the descriptive statistics. These results indicate a need for new and improved health indicators.

Among the 13 health indicators found in the seven levels, all levels were not significantly different from their immediate next level (Table 4); that is, there were no significant differences between levels 1 and 2 (ie, between n and n + 1). There were, however, significant differences between level 1 and levels 3 to 7 (ie, between n and any level higher than n + 1). These results pertain to the further prioritization of health indicators.

Given descriptive statistics and inferential test results, our findings among the 13 health indicators can reasonably be generalized to some extent to a broader population beyond our survey respondents. We do not claim the complete generalizability of our results mainly because our respondents were not perfectly representative of the composition of the American population. However, we believe that the 13 health indicators and their importance rankings within and among levels can provide substantial and useful evidence when such indicators need to be prioritized.

Cronbach α is one of the more cited statistics for informing internal consistency for the items of an instrument. If Cronbach α is greater than .7, the instrument is reliable [20]. The Cronbach α for the entire survey among the three samples was between .893 and .925. This indicates that we developed a reliable survey instrument. When examining the subscales, only the health care category, which included vaccination and immunization, insurance coverage, and personal care need, was below .7. The items in this category are among the most discussed topics in health care in the United States. Understandably, the reliability is lower since the respondents have relatively less consistent perspectives regarding these items.

Significance and Comparison With Related Research

This study provides a more comprehensive understanding of the indicators affecting an individual's health status, particularly as compared to self-rated health status, the most commonly used health status measurement [2]. Although there are advantages associated with using a single health indicator during clinical encounters, we believe that the multidimensional measurement of an individual's health status may be more objective and can provide additional insights into the individual's health status, particularly if we are concerned with improving and maximizing the preventive health care services offered. Obtaining these public perspectives is the first step toward a more accurate and effective measurement of individual health status.

This work can be potentially used in two ways: (1) to develop a more comprehensive and objective measurement of an individual's health status and (2) to develop a health index for an individual. Additionally, these results can be used to prioritize various health indicators (eg, to distinguish between core and secondary indicators). They can also be referenced by designers and developers for EHR systems, personal health record (PHR) systems, or other data capture and analysis applications to determine what health indicators to include in the systems. Furthermore, these results can contribute to developing a health index, which can be used to stratify healthy research participants to make them more comparable. This would be analogous to the Charlson Comorbidity Index [21] or propensity scores [22], which are broadly used in clinical epidemiology data analytics, both of which, however, are disease oriented. Although the health indicators reported here are not in a formula format, this will be a focus for future research. These results set the foundation for further weighting, prioritizing, and validating health indicators via additional data resources.

Additionally, these measurements can track overall health status, measure the outcomes of preventive services, or aggregate data to examine community health. Although having more data points provides increased accuracy and specificity for health indicators embedded within an EHR or PHR, it is important to consider clinician burnout [23] when using technology. Therefore, it is necessary to be mindful of the impacts that creating more data capture requirements or expectations of clinical users may have. In this regard, prioritizing health indicators is a necessary step.

Over the years, other systems have been developed to assess various health risks and associations. The Johns Hopkins Adjusted Clinical Group system, developed and maintained by Johns Hopkins University for over 30 years, is a global tool used in population health analytics [24]. This system is focused on chronic conditions and comorbidities, and its goal is, therefore, fundamentally different from ours, which is to measure individual *health* status, versus *disease* status, more accurately. Another system, the Committee on Quality Measures for the Healthy People Leading Health Indicators [3], focuses more on quality measures with an aim to align the measurements within a framework of assessment, improvement, and accountability. The focus, however, is on monitoring and reporting at the population level, not necessarily individual health [3].

There are other health-related surveys broadly used worldwide. For example, the 36-Item Short Form Health Survey, developed by the RAND Corporation [25], measures quality-of-life and health outcomes. Similarly, compared with the related but smaller 12-Item Short Form Health Survey [26], our health indicators provide a more comprehensive measurement beyond physical and mental health. The 9-item Patient Health

XSL•FO
RenderX

Questionnaire [27] is a validated tool that measures depression severity. However, we were looking for more objective indicators to measure an individual's physical and mental health status in our work.

Our health indicators have good but not all-inclusive coverage. The Institute of Medicine's Committee on the Recommended Social and Behavioral Domains and Measures for Electronic Health Records identified measures across the individual and neighborhood levels that involve sociodemographic, psychological, and behavioral data [4,5]. Among the 17 domains identified by the committee [4,5], 10 were included in our 29 health indicators. Healthy People 2030 [28] proposed 22 leading health indicators for different age groups, of which 16 are included in our health indicators.

Limitations of This Study

The main limitation of this study is that it is only the first step in determining the importance of these health indicators and, notably, the results are subjective, as they are based on public perspectives. Further validation of these results via additional objective measures, such as health care expenditure by disease category [29] and the burden of illness estimates for specific disease categories [30], is needed to support these findings. In this study, for each health indicator, the sample size of valid responses ranged from 791 to 1153. We recognize that larger sample sizes may generate more conclusive and generalizable results. Therefore, our results about the 13 health indicators, even though they are inferential statistics, should be treated as preliminary baseline results; future research may be needed to validate these findings in other settings.

Another limitation concerns the survey respondents. Females comprised the majority of survey respondents, making up 72.1%, 77.7%, and 69.0% of the samples from Ohio University, ResearchMatch, and Clemson University, respectively. We noticed a similar phenomenon in other studies conducted via ResearchMatch. While we are pleased with the relatively large sample size, responses may reflect the perspectives of well-educated females more than those of other groups. For example, survey respondents with a college-level education and beyond represented 54.6%, 82.2%, and 74.0% of the respondents from Ohio University, ResearchMatch, and Clemson University, respectively.

In addition to the distribution imbalance in gender and educational background among our respondents, we also noticed that race and ethnicity groups (Multimedia Appendix 5) were not perfectly representative of the composition of the American population. The breakdown by racial groups among respondents of our surveys was as follows: White American, 87.3%; African American, 3.3%; Hispanic and Latino American, 2.2%; Asian American, 1.6%; Native American, 0.4%; and two or more races, 2%. We recognize that our data set's gender and ethnicity imbalances are limitations of our current convenience sampling method. In the future, a stratified random-sampling method based on census-based population demographical data might provide more representative results and be a better option. This is a critical point that should be considered when using the results from this study.

Future Research

We foresee several potential directions in which to continue this project. Our primary goal for future research is to validate the results obtained from the three completed surveys. This can be accomplished in several ways. Because we wish to measure individual health status accurately over time, the use of longitudinal data would be ideal. One data source is a citizen science project initiated by the National Institutes of Health, the All of Us [31] research program. Another source is the UK Biobank initiated in the United Kingdom [32], but the most ideal source would be well-documented longitudinal data of a group of individuals that include not only their EHR data but also other data that correlate with our health indicators. Such ideal data sources would allow for examining the corresponding health indicators and validation of the importance of health indicators via EHR records and additional health-related data. In this way, public perspectives will be considered along with more concrete quantitative evidence to ensure more confidence in prioritizing health indicators and using them for various purposes.

Additionally, to mitigate the effect of the current imbalances seen in respondents regarding gender, race and ethnicity, and other factors, we could explore the possibility of stratified random sampling to proactively select more representative participants. The respondent pool can be more proportionally representative of the composition of the American population. As a potential future project, we may also explore possible correlations between demographic variables and rating results.

Conclusions

Well-designed health indicators are critical tools needed to accurately measure individual health status. They enable the determination of effective preventive services and verify their outcomes. Obtaining the public's perspective on specific health indicators is the first step toward prioritizing them for analytical and clinical use. This study found that the top five–rated health indicators were drug and substance abuse, smoking and tobacco use, alcohol abuse, major depression, and diet and nutrition. Our respondents, however, had heterogeneous views on the top- and bottom-rated health indicators. The middle 13 health indicators were rated more homogeneously among all the respondents. These 13 health indicators were separated into seven levels based on their perceived importance, providing further evidence that was used to prioritize these health indicators. Levels 1 to 7 were organized based on the mean importance of health indicators from high to low within and between each level. Level 1 included blood sugar level and immunization and vaccination; level 2 included LDL cholesterol; level 3 included HDL cholesterol, blood triglycerides, cancer screening detection, and total cholesterol; level 4 included health literacy rate; level 5 included personal care needs and air quality index greater than 100; level 6 included self-rated health status and HIV testing; and level 7 included the supply of dentists. The results of this study can provide evidence to EHR or PHR system designers and developers, which they can then use to select health indicators to incorporate into their systems.

XSL•FO
RenderX

Acknowledgments

We want to thank all the respondents who answered the surveys. Without their input, this study would not have been possible. This work is partially supported by the Department of Public Health Sciences, College of Behavioral, Social, and Health Sciences, Clemson University, which provided start-up funding for XJ.

Data Availability

This paper's data sets and statistical analysis codes are available by request from the corresponding author.

Authors' Contributions

XJ, TS, FL, and SD were responsible for conceptualization and design. XJ, TS, YZ, SD, TL, LH, LS, and RWG were responsible for the acquisition, analysis, or interpretation of data and for critically revising the manuscript. TS, YZ, and XJ were responsible for drafting the manuscript. TS, YZ, SD, LH, and XJ were responsible for statistical analysis. XJ was responsible for obtaining funding. XJ and RWG were responsible for supervision of the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey instrument. Health indicators to measure an individual's health status: a public perspective survey.
[PDF File (Adobe PDF File), 183 KB - jmir_v24i6e38099_app1.pdf]

Multimedia Appendix 2
Definition of the health indicators used in the survey.
[PDF File (Adobe PDF File), 105 KB - jmir_v24i6e38099_app2.pdf]

Multimedia Appendix 3
Records from three data sets before and after data cleaning.
[PDF File (Adobe PDF File), 30 KB - jmir_v24i6e38099_app3.pdf]

Multimedia Appendix 4
Survey codebook.
[PDF File (Adobe PDF File), 86 KB - jmir_v24i6e38099_app4.pdf]

Multimedia Appendix 5
Demographic descriptive statistics for all respondents.
[PDF File (Adobe PDF File), 48 KB - jmir_v24i6e38099_app5.pdf]

Multimedia Appendix 6
Total of nine indicators with homogenous variance.
[PDF File (Adobe PDF File), 30 KB - jmir_v24i6e38099_app6.pdf]

Multimedia Appendix 7
Total of 20 indicators with heterogeneous variance.
[PDF File (Adobe PDF File), 85 KB - jmir_v24i6e38099_app7.pdf]

Multimedia Appendix 8
Total of 16 indicators with significant mean differences: post hoc results for the three samples.
[PDF File (Adobe PDF File), 105 KB - jmir_v24i6e38099_app8.pdf]

References

1. Wold C. Health Indicators: A Review of Reports Currently in Use. Washington, DC: The State of the USA; 2008 Jul. URL: https://www.uc.pt/fluc/gigs/GeoHealthS/doc_apoio/6_Health_Indicators_review_of_reports_currently_in_use_2008.pdf [accessed 2022-06-03]

XSL•FO
RenderX

2. Committee on Core Metrics for Better Health at Lower Cost, Institute of Medicine. In: Blumenthal D, Malphrus E, McGinnis JM, editors. Vital Signs: Core Metrics for Health and Health Care Progress. Washington, DC: National Academies Press; 2015.
3. Committee on Quality Measures for the Healthy People Leading Health Indicators, Board on Population Health and Health Practice, Institute of Medicine. Toward Quality Measures for Population Health and the Leading Health Indicators. Washington, DC: National Academies Press; 2013.
4. Committee on the Recommended Social and Behavioral Domains and Measures for Electronic Health Records, Board on Population Health and Public Health Practice, Institute of Medicine. Capturing Social and Behavioral Domains and Measures in Electronic Health Records. Phase 2. Washington, DC: National Academies Press; 2015.
5. Committee on the Recommended Social and Behavioral Domains and Measures for Electronic Health Records, Board on Population Health and Public Health Practice, Institute of Medicine. Capturing Social and Behavioral Domains in Electronic Health Records. Phase 1. Washington, DC: National Academies Press; 2014.
6. National Health Interview Survey. Centers for Disease Control and Prevention, National Center for Health Statistics. URL: https://www.cdc.gov/nchs/nhis/index.htm [accessed 2021-08-29]
7. Hensrud DD. Clinical preventive medicine in primary care: Background and practice: 1. Rationale and current preventive practices. Mayo Clin Proc 2000 Feb;75(2):165-172. [doi: 10.4065/75.2.165] [Medline: 10683656]
8. Hensrud DD. Clinical preventive medicine in primary care: Background and practice: 2. Delivering primary preventive services. Mayo Clin Proc 2000 Mar;75(3):255-264. [doi: 10.4065/75.3.255] [Medline: 10725952]
9. Jing X, Lekey F, Kacpura A, Jefford K. Health indicators within EHR systems in primary care settings: Availability and presentation. In: Proceedings of the 5th International Conference on Health Information Science. 2016 Presented at: The 5th International Conference on Health Information Science; November 5-7, 2016; Shanghai, China p. 161-167. [doi: 10.1007/978-3-319-48335-1_18]
10. Ward BW, Clarke TC, Nugent CN, Schiller JS. Early Release of Selected Estimates Based on Data From the 2015 National Health Interview Survey. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics; 2016. URL: https://www.cdc.gov/nchs/data/nhis/earlyrelease/earlyrelease201605.pdf [accessed 2022-06-03]
11. Institute of Medicine (US) Committee on the State of the USA Health Indicators. State of the USA Health Indicators: Letter Report. Washington, DC: National Academies Press; 2009.
12. State of Global Well-being: Results of the Gallup-Healthways Global Well-being Index. Washington, DC: Gallup, Inc, and Healthways, Inc; 2014. URL: https://wellbeingindex.sharecare.com/wp-content/uploads/2017/12/State-of-Global-Well-Being-2014.pdf [accessed 2022-06-12]
13. Broussard DL, Sappenfield WB, Fussman C, Kroelinger CD, Grigorescu V. Core state preconception health indicators: A voluntary, multi-state selection process. Matern Child Health J 2011 Feb;15(2):158-168. [doi: 10.1007/s10995-010-0575-x] [Medline: 20225127]
14. Hunter BM, Requejo JH, Pope I, Daelmans B, Murray SF. National health policy-makers' views on the clarity and utility of Countdown to 2015 country profiles and reports: Findings from two exploratory qualitative studies. Health Res Policy Syst 2014 Aug 15;12:40 [FREE Full text] [doi: 10.1186/1478-4505-12-40] [Medline: 25128385]
15. Healthy People 2020 Leading Health Indicators: Progress Update. HealthyPeople.gov. Washington, DC: US Department of Health and Human Services; 2014. URL: https://www.healthypeople.gov/2020/leading-health-indicators/Healthy-People-2020-Leading-Health-Indicators%3A-Progress-Update [accessed 2022-06-03]
16. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Population Health and Public Health Practice, Committee on Informing the Selection of Health Indicators for Healthy People 2030. Leading Health Indicators 2030: Advancing Health, Equity, and Well-Being. Washington, DC: National Academies Press; 2020.
17. Measures codes. Archive-It Wayback Machine. Baltimore, MD: Centers for Medicare & Medicaid Services; 2016. URL: https://wayback.archive-it.org/2744/20160513011238/https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/MeasuresCodes.html [accessed 2022-06-12]
18. Jing X, Ogundeyi T, Lekey F, Law T, Diaz S. Health indicators to measure individual health status: Public perspectives–preliminary results. In: Proceedings of the AMIA 2019 Informatics Summit. 2019 Presented at: The AMIA 2019 Informatics Summit; March 25-28, 2019; San Francisco, CA p. 970-971.
19. ResearchMatch. URL: https://www.researchmatch.org/ [accessed 2021-09-19]
20. Christmann A, Van Aelst S. Robust estimation of Cronbach's alpha. J Multivar Anal 2006 Aug;97(7):1660-1674. [doi: 10.1016/j.jmva.2005.05.012]
21. Charlson ME, Pompei P, Ales KL, MacKenzie C. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. J Chronic Dis 1987 Jan;40(5):373-383. [doi: 10.1016/0021-9681(87)90171-8]
22. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res 2011 May;46(3):399-424 [FREE Full text] [doi: 10.1080/00273171.2011.568786] [Medline: 21818162]

XSL•FO
RenderX

23. McPeek-Hinz E, Boazak M, Sexton JB, Adair KC, West V, Goldstein BA, et al. Clinician burnout associated with sex, clinician type, work culture, and use of electronic health records. JAMA Netw Open 2021 Apr 01;4(4):e215686 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.5686] [Medline: 33877310]
24. Johns Hopkins ACG® System. URL: https://www.hopkinsacg.org/ [accessed 2021-09-19]
25. Ware JE, Sherbourne CD. The MOS 36-ltem Short-Form Health Survey (SF-36). Med Care 1992;30(6):473-483. [doi: 10.1097/00005650-199206000-00002]
26. 12-Item Short Form Survey (SF-12). RAND Corporation. URL: https://www.rand.org/health-care/surveys_tools/mos/12-item-short-form.html [accessed 2021-09-28]
27. 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]
28. Leading Health Indicators. Health.gov. URL: https://health.gov/healthypeople/objectives-and-data/leading-health-indicators [accessed 2021-08-30]
29. Ozieh MN, Bishu KG, Dismuke CE, Egede LE. Trends in healthcare expenditure in United States adults with chronic kidney disease: 2002-2011. BMC Health Serv Res 2017 May 22;17(1):368 [FREE Full text] [doi: 10.1186/s12913-017-2303-3] [Medline: 28532412]
30. Cannon A, Handelsman Y, Heile M, Shannon M. Burden of illness in type 2 diabetes mellitus. J Manag Care Spec Pharm 2018 Sep;24(9-a Suppl):S5-S13. [doi: 10.18553/jmcp.2018.24.9-a.s5] [Medline: 30156443]
31. All of Us Research Program Investigators, Denny JC, Rutter JL, Goldstein DB, Philippakis A, Smoller JW, et al. The "All of Us" research program. N Engl J Med 2019 Aug 15;381(7):668-676 [FREE Full text] [doi: 10.1056/NEJMsr1809937] [Medline: 31412182]
32. UK Biobank. URL: https://www.ukbiobank.ac.uk/ [accessed 2022-06-03]

Abbreviations

ANOVA: analysis of variance
EHR: electronic health record
HDL: high-density lipoprotein
LDL: low-density lipoprotein
PHR: personal health record

Edited by A Mavragani; submitted 18.03.22; peer-reviewed by M Richards, E Beverly, KM Kuo; comments to author 07.04.22; revised version received 10.05.22; accepted 20.05.22; published 21.06.22.

Please cite as:
Sokoya T, Zhou Y, Diaz S, Law T, Himawan L, Lekey F, Shi L, Gimbel RW, Jing X
Health Indicators as Measures of Individual Health Status and Their Public Perspectives: Cross-sectional Survey Study
J Med Internet Res 2022;24(6):e38099
URL: https://www.jmir.org/2022/6/e38099
doi:10.2196/38099
PMID:35623051

©Temiloluwa Sokoya, Yuchun Zhou, Sebastian Diaz, Timothy Law, Lina Himawan, Francisca Lekey, Lu Shi, Ronald W Gimbel, Xia Jing. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.06.2022. 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

Present Situation and the Future Development of Web-Based Prenatal Education in China: Cross-sectional Web-Based Survey

Xinyu Huang[1*], MBBS; Weiwei Sun[1*], MBBS; Renyu Wang[2*], BA; Huailiang Wu[1*], MBBS; Shinning Yu[1*], MBBS; Xuanbi Fang[3*], BSc, MPH; Yiyan Liu[4], MSSC; Babatunde Akinwunmi[5], MMSc, MD; Jian Huang[6], PhD; Wai-kit Ming[7], MD, PhD, MPH, MMSc

[1]International School, Jinan University, Guangzhou, China

[2]School of English for International Business, Guangdong University of Foreign Studies, Guangdong, China

[3]Department of Public Health and Preventive Medicine, School of Medicine, Jinan University, Guangzhou, China

[4]Faculty of Social Science, The Chinese University of Hong Kong, Hong Kong, China (Hong Kong)

[5]Maternal-Fetal Medicine Unit, Department of Obstetrics and Gynecology, Brigham and Women's Hospital, Boston, MA, United States

[6]Singapore Institute for Clinical Sciences, Agency for Science, Technology and Research, Singapore, Singapore

[7]Department of Infectious Diseases and Public Health, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, Hong Kong, China (Hong Kong)

[*]these authors contributed equally

Corresponding Author:
Wai-kit Ming, MD, PhD, MPH, MMSc
Department of Infectious Diseases and Public Health
Jockey Club College of Veterinary Medicine and Life Sciences
City University of Hong Kong
5/F, Block 1, To Yuen Building, 31 To Yuen Street
Tat Chee Avenue, Kowloon
Hong Kong
China (Hong Kong)
Phone: 852 3442 6956
Email: wkming2@cityu.edu.hk

Abstract

Background: Pregnancy serves as an important chapter in the life of women since more attention needs to be paid to both their physical and psychological health during this period. Adequate prenatal knowledge plays a key role in ensuring the health and safety of not only the pregnant women but also their fetuses and the entire family. With the development of information technology, web-based prenatal education has been brought into focus owing to its accessibility to comprehensive information, with high-quality information available to improve the quality of the overall gestation period, labor process, perinatal outcomes, and fetal outcomes.

Objective: This study aims to investigate the present situation of web-based prenatal education and to predict the future research direction of web-based prenatal education in China, thereby providing insights into improving the quality of health care of pregnant women.

Methods: A national cross-sectional study was conducted on 590,912 pregnant women in 31 provincial administrations of mainland China between August 2018 and August 2019. These pregnant women were initially recruited from local hospitals across the nation during antenatal and postnatal periods via a web-based education school. Demographic information and course completion status (including the categories and the number of courses they completed) of all the participants were collected.

Results: A total of 590,912 pregnant women participated in the web-based prenatal education in 2018. Among them, 188,508 (31.90%) participants were excluded because they did not complete any course, while 17,807 (3.01%) actively participated in web-based prenatal education and completed more than 100 courses. There were 5 categories of web-based courses; almost half of the pregnant women attended the courses on first and second trimesters (293,262/590,912, 49.63% and 298,168/590,912, 50.46%, respectively). We found that pregnant women were more concerned about the gestational diet, fetal-related knowledge, and other precautions before the labor.

XSL•FO
RenderX

Conclusions: In the era of digitalization where information is rapidly disseminated, web-based prenatal education could become a more convenient, productive, and effective pathway for pregnant women since it could help them obtain adequate and optimal pregnancy-related information and gain more intellectual awareness about their pregnancy or preparation for pregnancy.

(J Med Internet Res 2022;24(6):e28637) doi:10.2196/28637

KEYWORDS

web-based prenatal education; pregnancy; prenatal; information technology

Introduction

Prenatal information plays a vital role in health decision-making of pregnant women because they come across a series of physiological and psychological changes during pregnancy [1,2]. The quantity and quality of prenatal care information have been reported to be associated with the quality of life of pregnant women and the fetal outcomes in the short term and long term [3-5]. Specifically, high-quality health information is associated with better and safer pregnancy outcomes such as less preterm deliveries, less anxiety problems, lower cesarean section rates, lower maternal and infant mortality, as well as greater prenatal engagement of their partners [6]. Therefore, it is of great significance for pregnant women to gain and understand adequate and high-quality prenatal health information as much as possible.

Adequate and high-quality prenatal education is necessary for pregnant women to obtain reliable information and achieve desirable pregnancy outcomes as well as uneventful postnatal courses. However, the benefits and effectiveness of conventional prenatal education are still limited and inconclusive to some extent owing to the inadequacy in quality and quantity [7,8]. With the development of information technology, web-based prenatal education has received considerable attention. Compared to traditional group prenatal education, structural and well-organized web-based prenatal education programs can provide an easier access to medical services and integrated information, which could promote the utilization of medical sources, lower the overall educational costs, increase family engagement and satisfaction, and thereby improve the quality of daily life for pregnant women [9]. In mainland China, there were more than 900 million smart device users in 2020, and the idea of adopting web-based education for pregnancy health care–related services was increasingly viable [9]. Although there are extensive advantages, the current situation of Chinese pregnant women participating in web-based prenatal education is not known because only few studies have investigated the use of web-based prenatal education programs [10-12]. Therefore, in this study, we aim to provide insights into the present situation and future development of web-based prenatal education in China by screening the data collected from a nationwide web-based survey.

Methods

Recruitment

This study was designed as a nationwide cross-sectional study. Pregnant women from 31 provinces of mainland China were recruited via a web-based prenatal education school (Banmi web-based maternity school). This web-based system was allowed to assess participants' health conditions from local hospitals and provide primary health care education to these participants through the website during their pregnancy. A total of 280 courses were offered by this web-based prenatal school, each of which took about 5 minutes to finish in average. All courses were available at any point during pregnancy. The courses transfer the information passively, but the pregnant women can take some quiz questions for self-examination after finishing each course, or they can consult with their obstetricians when they went to the hospital for antenatal examinations. The courses contained text and video, and they were free of charge to all women. All the results on course participation were based on back-end data from this web-based course system.

Participants

The eligibility inclusion criteria included (1) pregnant women registered with the web-based prenatal education school, (2) pregnant women residing in mainland China, and (3) the expected date of delivery was between August 1, 2018 and August 31, 2019. Informed consent for research was obtained from each participant when they registered with the web-based system.

Ethics Approval

Ethics approval was granted by the Institutional Review Board of the First Affiliated Hospital of Sun Yat-Sen University (ICE-2017-296). All procedures were conducted in accordance with the Declaration of Helsinki.

Variables

Participants' demographic data, including gestational age and residential location, were collected. The web-based prenatal education program categorized these prenatal courses into 5 groups: preparation for pregnancy, first trimester, second trimester, third trimester, and postpartum care. Data on course completion conditions were also collected, including the category and the number of completed courses.

Data Analysis

We performed descriptive analyses on the data collected via the web-based prenatal education school on SPSS Statistics 25.0 for Windows (IBM Corp). We reported mean (SD) and ranges for variables that followed a normal distribution. The number of pregnant women in each province was retrieved from the China Health Statistics Yearbook 2019.

Results

In 2018, 590,912 women were included in this study. Figure 1 shows the heatmap of the proportion of pregnant women

XSL•FO
RenderX

registered with the web-based prenatal education school in 31 Chinese provincial administrations. The region-specific proportion of pregnant women registered with the web-based prenatal education school in mainland China ranged from 0.14% (81/56,622) in Tibet to 10.46% (31,772/303,647) in Shanxi (Table 1). The proportion of pregnant women attending web-based prenatal education was less than 5% (590,912/13,621,475) in 22 (71%) of the 31 provincial administrations.

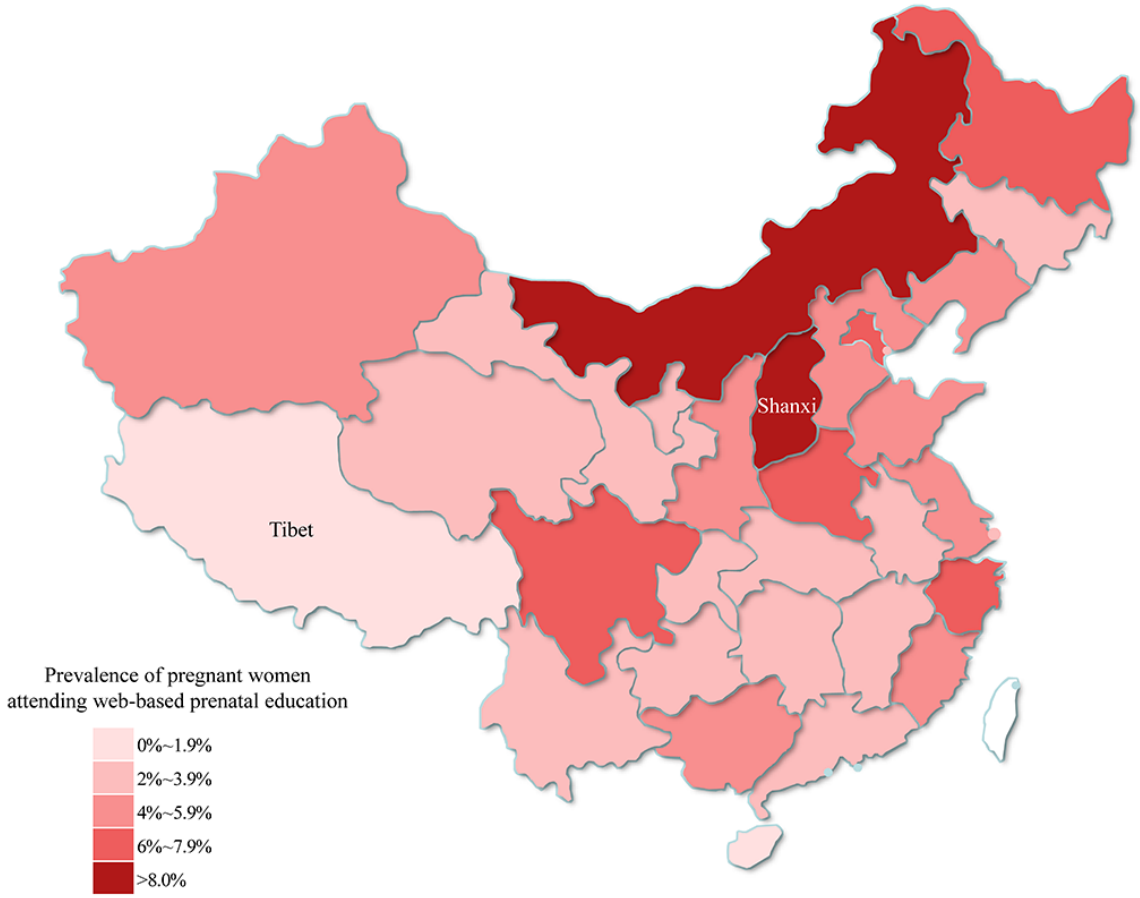
Of the 590,912 pregnant women, 188,508 (31.90%) pregnant women did not complete any course, while 136,938 (23.27%) pregnant women completed 10-100 courses, which we defined as medium participation. There were 247,659 (41.91%) pregnant women who completed only 1-10 courses in web-based prenatal education. We defined this as low participation, and that segment accounted for the largest portion of the population. Only 17,807 (3.01%) pregnant women completed more than 100 courses in the web-based prenatal education school; this was defined as high participation. Among them, 24,761 (4.19%) pregnant women participated in the course on preparation for pregnancy, 293,262 (49.63%) participated in the course on the first trimester, 298,168 (50.46%) participated in the course on the second trimester, 82,726 (14%) participated in the course on the third trimester, and 154,327 (26.12%) took the course on postpartum care.

More than half of the pregnant women attended the courses on gestational diet, fetal-related knowledge, and other precautions before the labor (Table 2). As for dietary educational courses, of the 590,912 pregnant women, 101,890 (17.24%) attended the course on “diet in second trimester” and 81,653 (13.75%) attended the course on calcium supplements, diet restrictions, anemia, and iron supplements during pregnancy. Among these pregnant women, 16.94% (100,109/590,912) were eager to obtain knowledge on fetal kicks and fetal movement counting. Approximately 13.20% (78,009/590,912) attended the courses on screening for Down syndrome, and 12.40% (73,301/590,912) were eager to understand the knowledge of screening for deformity and other congenital malformations. Besides, these pregnant women also attended the courses on behavioral change and precautions during pregnancy, such as body changes, sexual behavior during pregnancy, as well as contraindications and cautions on harmful things for pregnant women.

Apart from the courses with the highest number of participants, we also calculated the 10 courses with the least participants, as shown in Table 3.

From this table, we can observe that the courses related to pregnancy diseases were attended by less number of pregnant women. Less than 1% of the pregnant women attended the courses on gestational diabetes, premature delivery, female infertility, extrauterine pregnancy, and polycystic ovarian syndrome.

Figure 1. Heatmap for the prevalence of pregnant women attending web-based prenatal education in different Chinese provincial administrations.



XSL•FO
RenderX

Table 1. Proportions of pregnant women completing web-based prenatal education in mainland China in 2018.a

Provincial administration	Pregnant women (N)[a]	Pregnant women attending web-based prenatal education, n (%)
North China		
Shanxi	303,647	31,772 (10.46)
Inner Mongolia	181,445	15,311 (8.44)
Beijing	140,304	10,668 (7.60)
Hebei	735,253	34,914 (4.75)
Tianjin	78,072	2765 (3.54)
Northeast China		
Heilongjiang	151,058	9828 (6.51)
Liaoning	253,180	13,405 (5.29)
Jilin	136,396	3399 (2.49)
East China		
Zhejiang	413,967	27,061 (6.54)
Fujian	444,863	21,272 (4.78)
Jiangsu	619,047	26,037 (4.21)
Shandong	1,058,022	42,893 (4.05)
Shanghai	69,734	2505 (3.59)
Jiangxi	558,666	19,513 (3.49)
Anhui	736,530	16,805 (2.28)
Middle China		
Henan	1,126,750	65,987 (5.86)
Hubei	590,923	21,321 (3.61)
Hunan	705,524	18,828 (2.67)
South China		
Guangxi	667,539	36,975 (5.54)
Guangdong	1,317,909	38,879 (2.95)
Hainan	101,307	1076 (1.06)
Southwest China		
Sichuan	800,752	50,241 (6.27)
Yunnan	564,411	16,060 (2.85)
Chongqing	273,940	7689 (2.81)
Qinghai	74,287	1689 (2.27)
Tibet	56,622	81 (0.14)
Northwest China		
Shaanxi	346,083	19,634 (5.67)
Xinjiang	197,954	8497 (4.29)
Gansu	278,912	10,102 (3.62)
Ningxia	73,409	2524 (3.44)
Guizhou	564,969	13,180 (2.33)
Overall	13,621,475	590,912 (4.34)

[a]Source: China Health Statistics Yearbook 2019.

XSL•FO
RenderX

Table 2. Top 10 popular courses among pregnant women (N=590,912).

Course	Participants, n (%)
Diet in the second trimester	101,890 (17.24)
Fetal movement counting	100,109 (16.94)
Body change in the second trimester	99,895 (16.91)
Calcium supplement	81,653 (13.82)
Diet restrictions during pregnancy	81,259 (13.75)
Sexual behavior during pregnancy	81,143 (13.73)
Anemia and iron supplements during pregnancy	78,141 (13.22)
Screening for Down syndrome	78,009 (13.20)
Screening for deformity	73,301 (12.40)
Harmful things for pregnant women	73,179 (12.38)

Table 3. Ten least popular courses among pregnant women (N=590,912).

Course	Participants, n (%)
How to select and use diapers	199 (0.03)
Prevention and prognosis of gestational diabetes	1779 (0.30)
How to prevent premature delivery	1841 (0.31)
Labor sign: amenorrhea	2474 (0.42)
The best age for pregnancy	2583 (0.44)
What is Doula delivery	2661 (0.45)
Pet raise during pregnancy	3149 (0.54)
Female infertility	3174 (0.54)
Reasons for extrauterine pregnancy	3225 (0.55)
Polycystic ovarian syndrome and pregnancy	3757 (0.64)

Discussion

Principal Findings

In this study, we found that the overall proportion of pregnant women registered with the web-based prenatal education school was 4.34% (590,912/13,612,475) in mainland China, while there were substantial differences from region to region. Although about one-third of the participants did not complete any course in the system, each pregnant woman completed an average of about 20 courses. In terms of course preference, half of the pregnant women were enrolled in courses on the first trimester and second trimester of pregnancy. Specifically, dietary supplements, fetal-related knowledge, and predelivery considerations such as how to protect themselves and how to schedule prenatal visits were the most popular topics.

A previous study reported that pregnant women who participated in web-based prenatal education performed better than those who did not, in many aspects during pregnancy, including daily activities and perinatal care [13]. With the prevalence and evolution of the internet and mobile devices, the amount of information available online has increased dramatically. Web-based prenatal education is favored by more pregnant women because of its convenience, accessibility, flexibility, and cost-effectiveness [14]. According to a Chinese study, pregnant women preferred evidence-based information, expert opinions, and tailored advice from the internet [15]. However, in 2018, only 4.34% (590,912/13,612,475) of the pregnant women registered with the web-based prenatal education school in our study. Such results are partly attributed to some unstandardized and unreliable scientific information on websites. In addition, people lack trust in the information available on the internet and fear being misled or misinformed [16]. Findings from a Korean study showed that 39% of internet health information was inaccurate and 42.7% of those misleading messages were irrelevant to health problems and could cause a series of negative impacts [17]. Thus, pregnant women might worry about the adverse effects of following wrong prenatal guidance from the internet [18].

Standard and certified web-based prenatal education contains a wealth of contents that have been professionally screened. Provision of accurate and appropriate prenatal knowledge to pregnant women can improve their self-care ability, and to some extent, improve their safety and physical conditions during their pregnancy [4]. On the contrary, inaccurate information from the internet can mislead pregnant women to perform behaviors that are not conducive to their health and safety. Therefore, web-based prenatal education should be evaluated and improved

XSL•FO
RenderX

by the health care system, medical institutes, and professional experts, as this is a promising pathway to provide higher-quality content for pregnant women and ensuring the accuracy and quality of the web-based prenatal education programs. This will help in achieving the desired aim of protecting pregnant women, thereby improving their quality of life during pregnancy and contributing to better pregnancy outcomes. This will also help in mitigating the anxiety and fears perceived by pregnant women toward inaccurate information and therefore promote the popularization and utilization of the web-based prenatal educational services [3,5,18].

Different use levels of web-based prenatal school in different regions were also found in this study. Internet access and use is closely associated with regional economic and education levels; the educational level and internet use are relatively low in the western provincial regions in mainland China [19,20]. This may well explain our observation that pregnant women in Tibet have the lowest proportion of participation in web-based prenatal education. Furthermore, according to the China Health Statistics Yearbook 2019, the rates of record establishment, prenatal examination, and systemic management of pregnant women in Tibet were the lowest among all provincial administrations in 2018 [21]. These were associated with the poor awareness of importance of prenatal education and examinations and the lack of adequate medical and educational resources for pregnant women in Tibet [22,23]. In contrast, northern and northeastern China were the regions with the highest numbers for these figures. Meanwhile, the gap was deepened by the relatively high rate of pregnant women establishing pregnancy records and prenatal visits in these 2 regions [21].

Only 3.01% (17,807/590,912) of the pregnant women completed more than 100 courses in this study, while about one-third of the pregnant women who had registered for the web-based prenatal school did not complete any course. This indicated a suboptimal level for the completion of participation in the web-based prenatal education. This could be explained by their indifferent attitudes toward web-based prenatal education and insufficient attraction of pregnancy courses. Thus, medical institutions and hospitals should collaborate to improve the contents and services of prenatal education programs. Moreover, courses on the first and second trimesters of pregnancy are more popular among pregnant women owing to the great physical and physiological changes in these phases of pregnancy [1,24,25]. Previous studies have shown that pregnant women normally preferred searching for information online at the beginning of pregnancy [26,27], and pregnant women usually start antenatal care during the first trimester, preferably before the 12th week of gestation [22]. A study found that more than half of the Chinese pregnant women completed their first antenatal care after the first trimester [28]. To sum up the above, these may explain the preference of knowledge on both first and second trimesters, since the pregnant women in our study were generally recruited at their visit to the hospitals for antenatal care. Nevertheless, 4.19% (24,761/590,912) of the pregnant women completed the courses on the preparation for pregnancy, which indicated that some pregnant women were also concerned about the knowledge on the anticipation and preparedness of pregnancy. Therefore, obstetric experts in web-based prenatal education schools should provide web-based lectures and consultation services for women of childbearing age; consequently, there will be an expected increase in safe conception and reduction in pregnancy complications when women of childbearing age seek for help in the web-based prenatal education school prior to pregnancy, with the education and guidance from obstetric experts.

Regarding the courses completed by the participants, the top 10 popular courses mainly focused on gestational diet, fetal-related knowledge, and precautions during pregnancy. Previous studies found that fetal development and nutrition knowledge were often a concern for pregnant women, and our findings were consistent with those reported previously [26,29]. Besides, better quality of maternal diet during pregnancy is positively associated with general maternal health, reproductive outcomes, and child neurodevelopment [24,30]. Furthermore, strong correlations of incidence of gestational diabetes mellitus and type 2 diabetes and energy intake during pregnancy were widely reported [25,31,32]. These pieces of evidence suggest that medical institutes should pay more attention to dietary habit education during pregnancy to improve the quality of life during pregnancy. Besides, fetal-related courses such as fetal movement counting, screening for Down syndrome, and screening for deformity were also popular. Prenatal screening mainly focused on Down syndrome and fetal anomaly scans in the routine second trimester [33]. Thus, it is necessary for hospitals and other medical institutions to provide more education and notices on prenatal screening. At the same time, the courses related to pregnancy diseases were not attractive to pregnant women, and only less than 1% of them attended the courses on gestational diabetes, premature delivery, female infertility, extrauterine pregnancy, and polycystic ovarian syndrome. This condition may be caused by the indifferent attitude toward pregnancy-related diseases, and many of them did not have a positive cognitive attitude toward those diseases. Therefore, obstetrics physicians should inform pregnant women of common pregnancy-related diseases and how to manage them properly in a timely manner, which is also highly relevant to pregnancy health.

With the development of modern information technology, web-based prenatal education can be a good choice for pregnant women to obtain gestational knowledge of qualified quality and adequate quantity, including how to prepare for pregnancy, what should be done during the gestational period, how to effectively recover from parturition, and what challenges they will meet during the postpartum period. The number of internet users in China is nearly 1000 million, of which rural internet users is 309 million. The nationwide network coverage rate reached 70.4% by December 2020, and the network coverage rate in rural areas was 55.9%, while previous data suggested that there is still about 64.5% of Chinese pregnant women who did not attend any prenatal education [34]. The web-based health program is cost-effective to provide information, resources, and education, which could help to fill in the gap in the field of prenatal education [12,35]. For pregnant women in rural areas, web-based prenatal education can help to reduce the financial burden such as the cost of transportation and medical

XSL•FO
RenderX

consultation in hospitals, providing them a more cost-effective way to improve their safety and life quality during pregnancy. Moreover, web-based prenatal education can avoid some unnecessary visits and reduce the risks of cross-infection in hospitals [36,37]. In general, web-based prenatal education provides comprehensive information to enable pregnant women to learn more about pregnancy, equip them with better self-care ability, and protect their own health and safety when facing some physiological and psychological changes during pregnancy. In addition, owing to its convenience and low cost, high-quality medical resources can be availed in any corner of the planet covered with the internet network. Web-based prenatal education provides a certain guarantee of health care for pregnant women both subjectively and objectively.

Web-based prenatal education programs could be improved in the following areas for offering a higher quality of health care for pregnant women [38,39]. First, women of childbearing age need access to log in before pregnancy—the time when they are more likely to search gestational information and gain some preconception knowledge for preparation. Second, the contents of web-based courses, especially knowledge of early pregnancy should be supplemented, as they are necessary for pregnant women to have a better quality of pregnancy. Moreover, obstetricians' involvement and consultation could be considered as a combination strategy with web-based prenatal programs in the future. Third, governments need to build better internet facilities and provide mobile devices with greater discounts in rural areas or for underprivileged people, because these can provide essential requirements for pregnant women to experience and benefit from web-based prenatal education programs.

Limitations

To the best of our knowledge, this is the first study to investigate current situations and provide insights into the development of web-based prenatal education programs in China. However, our study has few limitations. First, this study only included 1 national web-based prenatal school. There are about 10 local web-based prenatal schools in China, but they are mainly used in local hospitals by their own apps or official account. The Banmi web-based prenatal school mentioned in this paper is the largest one and can be accessed by everyone in mainland China. The data from this school as a reference to calculate the proportion of usage of web-based prenatal education for pregnant women might be slightly underestimated, but this result will be more accurate with the development of information technology and popularization of network. Second, all the results of course participation were based on back-end data from the web-based course system; therefore, we did not include any demographics or other maternal characteristics such as age, education, and employment status. When comparing the regional difference of proportions of using web-based prenatal education programs, this study did not account for the aforementioned factors, and these demographic factors might have great influences on the usage of web-based prenatal education programs. Besides, we did not know whether the course participation of pregnant women is driven by women's interests or obstetric providers' recommendations, which may influence the selection of pregnant women.

Comparison With Prior Work

We searched PubMed, Google Scholar, and China National Knowledge Infrastructure for papers about the current development and applications of web-based antenatal care in China. Of the studies we identified, we found that web-based antenatal care currently has a variety of researches and applications in the clinic, such as psychological interventions for depression, childbirth, and breastfeeding education. However, the present situation of Chinese pregnant women participating in web-based prenatal education is not known, and large-scale population-based studies in China focused on the participation rate of pregnant women enrolled in web-based antenatal care are absent. Thus, this is a pioneer national-based study to investigate the participation of pregnant women in web-based prenatal care in China. This study emphasizes on exploring the concerns and preferences of pregnant women, providing evidence and references for future web-based prenatal education development, which can contribute to the improvement of maternal and fetal health, quality of health care for pregnant women, and overall pregnancy outcomes. In this study, we observed the participation, concerns, and preference of pregnant women when they attend web-based prenatal courses. Our study shows that web-based prenatal education has great potential in future development and application, and web-based prenatal education could become a convenient and effective pathway for pregnant women to obtain important and optimal pregnancy-related knowledge and gain more information about the pregnancy while pregnant or preparing to be pregnant.

Conclusions

With the development and popularization of information technology, web-based prenatal education shows promising potential in future development and application. In mainland China, the usage of web-based prenatal education for pregnant women was not very high but increased rapidly. Support from the government and health care professionals can assist in the development of standard and high-quality web-based prenatal programs, which can contribute to the improvement of maternal and fetal health, quality of health care for pregnant women, and the overall pregnancy outcomes.

Acknowledgments

The authors would like to gratefully acknowledge all the participants, our collaborators, and the Banmi web-based maternity school, China.

XSL•FO
RenderX

Authors' Contributions

XH, WS, and HW contributed to the conception and design of work, analysis and interpretation of data, design of tables and figures, drafting and modification of the manuscript, and revision of the final version of the manuscript. SY contributed to the collection, analysis, and interpretation of data and revision of the final version of the manuscript. XF, BA, JH, RW and YL contributed to drafting and modification of the manuscript and revision of the final version of the manuscript. WM contributed to the conception and design of work and revision of the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Soma-Pillay P, Nelson-Piercy C, Tolppanen H, Mebazaa A. Physiological changes in pregnancy. CVJA 2016 May 18;27(2):89-94. [doi: 10.5830/cvja-2016-021]
2. Leifer M. Psychological changes accompanying pregnancy and motherhood. Genet Psychol Monogr 1977 Feb;95(1):55-96. [Medline: 849833]
3. Javanmardi M, Noroozi M, Mostafavi F, Ashrafi-Rizi H. Internet Usage among Pregnant Women for Seeking Health Information: A Review Article. Iran J Nurs Midwifery Res 2018;23(2):79-86 [FREE Full text] [doi: 10.4103/ijnmr.IJNMR_82_17] [Medline: 29628953]
4. Roch G, Borgès Da Silva R, de Montigny F, Witteman HO, Pierce T, Semenic S, et al. Impacts of online and group perinatal education: a mixed methods study protocol for the optimization of perinatal health services. BMC Health Serv Res 2018 May 29;18(1):382 [FREE Full text] [doi: 10.1186/s12913-018-3204-9] [Medline: 29843691]
5. Huberty J, Dinkel D, Beets MW, Coleman J. Describing the use of the internet for health, physical activity, and nutrition information in pregnant women. Matern Child Health J 2013 Oct;17(8):1363-1372. [doi: 10.1007/s10995-012-1160-2] [Medline: 23090284]
6. Ferguson S, Davis D, Browne J. Does antenatal education affect labour and birth? A structured review of the literature. Women Birth 2013 Mar;26(1):e5-e8. [doi: 10.1016/j.wombi.2012.09.003] [Medline: 23063931]
7. Brixval CS, Axelsen SF, Lauemøller SG, Andersen SK, Due P, Koushede V. The effect of antenatal education in small classes on obstetric and psycho-social outcomes - a systematic review. Syst Rev 2015 Feb 28;4:20 [FREE Full text] [doi: 10.1186/s13643-015-0010-x] [Medline: 25875612]
8. Martinez Galiano J, Delgado Rodriguez M. Effectiveness of the professional who carries out the health education program: perinatal outcomes. IJWH 2014 Mar:329-334. [doi: 10.2147/ijwh.s59126]
9. Yao H, Cao L, Gao Y, Wu R, Chen Y, Yan J. Effect of New Media on Health Education and Management of Pregnant Women. Hosp Admin J Chin PLA 2018; (11) 2018 Jul 18:1064-1066. [doi: 10.16770/j.cnki.1008-9985.2018.11.019]
10. Salonen AH, Pridham KF, Brown RL, Kaunonen M. Impact of an internet-based intervention on Finnish mothers' perceptions of parenting satisfaction, infant centrality and depressive symptoms during the postpartum year. Midwifery 2014 Jan;30(1):112-122. [doi: 10.1016/j.midw.2013.02.009] [Medline: 23623471]
11. Huang MZ, Kuo S, Avery MD, Chen W, Lin K, Gau M. Evaluating effects of a prenatal web-based breastfeeding education programme in Taiwan. J Clin Nurs 2007 Aug;16(8):1571-1579. [doi: 10.1111/j.1365-2702.2006.01843.x] [Medline: 17655546]
12. Lagan BM, Sinclair M, Kernohan WG. Internet use in pregnancy informs women's decision making: a web-based survey. Birth 2010 Jun;37(2):106-115. [doi: 10.1111/j.1523-536X.2010.00390.x] [Medline: 20557533]
13. Yin X, Zhang Q. Effect of "Online Pregnant School" on Improving Self-care Ability of Pregnant Women. Chinese Community Doctors 2018 Oct 15;34(28):162-165. [doi: 10.3969/j.issn.1007-614x.2018.28.101]
14. Chedid RA, Terrell RM, Phillips KP. Best practices for online Canadian prenatal health promotion: A public health approach. Women Birth 2018 Aug;31(4):e223-e231. [doi: 10.1016/j.wombi.2017.10.005] [Medline: 29113753]
15. Wang N, Deng Z, Wen LM, Ding Y, He G. Understanding the Use of Smartphone Apps for Health Information Among Pregnant Chinese Women: Mixed Methods Study. JMIR Mhealth Uhealth 2019 Jun 18;7(6):e12631 [FREE Full text] [doi: 10.2196/12631] [Medline: 31215516]
16. Hopf H, Krief A, Mehta G, Matlin SA. Fake science and the knowledge crisis: ignorance can be fatal. R Soc Open Sci 2019 May;6(5):190161 [FREE Full text] [doi: 10.1098/rsos.190161] [Medline: 31218057]
17. Kim J, Kim S. Physicians' perception of the effects of Internet health information on the doctor-patient relationship. Inform Health Soc Care 2009 Sep;34(3):136-148. [doi: 10.1080/17538150903102422] [Medline: 19670004]
18. Zhu C, Zeng R, Zhang W, Evans R, He R. Pregnancy-Related Information Seeking and Sharing in the Social Media Era Among Expectant Mothers: Qualitative Study. J Med Internet Res 2019 Dec 04;21(12):e13694 [FREE Full text] [doi: 10.2196/13694] [Medline: 31799939]
19. Gong W, Li ZG, Stump RL. Global internet use and access: cultural considerations. Asia Pac Jnl of Mrkting & Log 2007 Jan 16;19(1):57-74. [doi: 10.1108/13555850710720902]

XSL•FO
RenderX

20. Fang X, Gao B. A research on the comprehensive development level of education in China. Educational Research 2013 Dec:32-39.
21. China's Health Statistics Yearbook (2019 version) (in Chinese). China: Peking Union Medical College Press; 2019.
22. WHO antenatal care randomised trial for the evaluation of a new model of routine antenatal care. In: World Health Organization. Geneva: World Health Organization; 2002.
23. Lou X. Study on the Distribution of Higher Education Resources in China [masters thesis]. Shanghai, China: East China Normal University; 2014.
24. Borge TC, Aase H, Brantsæter AL, Biele G. The importance of maternal diet quality during pregnancy on cognitive and behavioural outcomes in children: a systematic review and meta-analysis. BMJ Open 2017 Sep 24;7(9):e016777 [FREE Full text] [doi: 10.1136/bmjopen-2017-016777] [Medline: 28947450]
25. Gilbert L, Gross J, Lanzi S, Quansah DY, Puder J, Horsch A. How diet, physical activity and psychosocial well-being interact in women with gestational diabetes mellitus: an integrative review. BMC Pregnancy Childbirth 2019 Feb 07;19(1):60 [FREE Full text] [doi: 10.1186/s12884-019-2185-y] [Medline: 30732571]
26. Gao L, Larsson M, Luo S. Internet use by Chinese women seeking pregnancy-related information. Midwifery 2013 Jul;29(7):730-735. [doi: 10.1016/j.midw.2012.07.003] [Medline: 22958935]
27. De Santis M, De Luca C, Quattrocchi T, Visconti D, Cesari E, Mappa I, et al. Use of the Internet by women seeking information about potentially teratogenic agents. Eur J Obstet Gynecol Reprod Biol 2010 Aug;151(2):154-157. [doi: 10.1016/j.ejogrb.2010.04.018] [Medline: 20478650]
28. Chen L, Dai Y, Zhang Y, Wu Q, Rudan D, Saftić V, et al. A comparison between antenatal care quality in public and private sector in rural Hebei, China. Croat Med J 2013 Apr;54(2):146-156 [FREE Full text] [doi: 10.3325/cmj.2013.54.146] [Medline: 23630142]
29. Lee A, Newton M, Radcliffe J, Belski R. Pregnancy nutrition knowledge and experiences of pregnant women and antenatal care clinicians: A mixed methods approach. Women Birth 2018 Aug;31(4):269-277. [doi: 10.1016/j.wombi.2017.10.010] [Medline: 29126796]
30. Chen X, Zhao D, Mao X, Xia Y, Baker P, Zhang H. Maternal Dietary Patterns and Pregnancy Outcome. Nutrients 2016 Jun 07;8(6):351-378 [FREE Full text] [doi: 10.3390/nu8060351] [Medline: 27338455]
31. Sundarapperuma T, Wijesinghe C, Hettiarachchi P, Wasalathanthri S. Perceptions on Diet and Dietary Modifications during Postpartum Period Aiming at Attenuating Progression of GDM to DM: A Qualitative Study of Mothers and Health Care Workers. J Diabetes Res 2018;2018:6459364 [FREE Full text] [doi: 10.1155/2018/6459364] [Medline: 30225269]
32. Shepherd E, Gomersall J, Tieu J, Han S, Crowther C, Middleton P. Combined diet and exercise interventions for preventing gestational diabetes mellitus. Cochrane Database Syst Rev 2017 Nov 13;11:CD010443 [FREE Full text] [doi: 10.1002/14651858.CD010443.pub3] [Medline: 29129039]
33. Dondorp W, van Lith J. Dynamics of prenatal screening: new developments challenging the ethical framework. Bioethics 2015 Jan;29(1):ii-iv. [doi: 10.1111/bioe.12127] [Medline: 25521974]
34. Liu S, Xu Q, Ge Y. A survey of needs of pregnant women and parturient for pregnant school and its countermeasures. Chinese Nursing Research 2009; (20) 2008 Sep 22:162-165. [doi: 10.3969/j.issn.1009-6493.2009.20.008]
35. Godin KM, Alton GD, Gangodawilage HP, Procter TD, Bourdages NB, Blue SE, et al. Knowledge change associated with participation in prenatal education programs in Ontario: A cohort study. Can J Public Health 2015 Oct 03;106(6):e401-e407 [FREE Full text] [doi: 10.17269/cjph.106.5057] [Medline: 26680432]
36. Ho PL, Tang XP, Seto WH. SARS: hospital infection control and admission strategies. Respirology 2003 Nov;8 Suppl:S41-S45 [FREE Full text] [doi: 10.1046/j.1440-1843.2003.00523.x] [Medline: 15018133]
37. Bayo P, Ochola E, Oleo C, Mwaka AD. High prevalence of hepatitis B virus infection among pregnant women attending antenatal care: a cross-sectional study in two hospitals in northern Uganda. BMJ Open 2014 Nov 11;4(11):e005889 [FREE Full text] [doi: 10.1136/bmjopen-2014-005889] [Medline: 25387757]
38. Pugh MA, Revell MA. Using online materials for prenatal education: The good, the bad and the ugly. International Journal of Childbirth Education 2011;26(4):9-13.
39. Dai Y, Chen X, Liu Y, Lin J. The effect of online education intervention in perinatal care. Journal of Nursing Science 2019;34:26-29. [doi: 10.3870/j.issn.1001-4152.2019.18.026]

Edited by R Kukafka; submitted 09.03.21; peer-reviewed by M Shapiro, M Waring; comments to author 31.05.21; revised version received 31.07.21; accepted 12.08.21; published 30.06.22.

Please cite as:
Huang X, Sun W, Wang R, Wu H, Yu S, Fang X, Liu Y, Akinwunmi B, Huang J, Ming WK
Present Situation and the Future Development of Web-Based Prenatal Education in China: Cross-sectional Web-Based Survey
J Med Internet Res 2022;24(6):e28637
URL: https://www.jmir.org/2022/6/e28637
doi:10.2196/28637
PMID:35771613

XSL•FO
RenderX

©Xinyu Huang, Weiwei Sun, Renyu Wang, Huailiang Wu, Shinning Yu, Xuanbi Fang, Yiyan Liu, Babatunde Akinwunmi, Jian Huang, Wai-kit Ming. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.06.2022. 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

Popular and Scientific Discourse on Autism: Representational Cross-Cultural Analysis of Epistemic Communities to Inform Policy and Practice

Christophe Gauld[1], MD; Julien Maquet[2], MD, MSc; Jean-Arthur Micoulaud-Franchi[3], MD, PhD, HDR; Guillaume Dumas[4,5], PhD, HDR

[1]Department of Child Psychiatry, Université de Lyon, Lyon, France
[2]Department of Internal Medicine, Toulouse University, Toulouse, France
[3]Department of Sleep Medicine, Bordeaux University, Bordeaux, France
[4]Center for Complex Systems and Brain Sciences, Florida Atlantic University, Boca Raton, FL, United States
[5]Center of Research, Centre Hospitalier Universitaire Sainte Justine, Montréal, QC, Canada

Corresponding Author:
Guillaume Dumas, PhD, HDR
Center for Complex Systems and Brain Sciences
Florida Atlantic University
BS-12 777 Glades Road
Boca Raton, FL, 33431
United States
Phone: 1 0014384885
Email: guillaume.dumas@ppsp.team

Abstract

Background: Social media provide a window onto the circulation of ideas in everyday folk psychiatry, revealing the themes and issues discussed both by the public and by various scientific communities.

Objective: This study explores the trends in health information about autism spectrum disorder within popular and scientific communities through the systematic semantic exploration of big data gathered from Twitter and PubMed.

Methods: First, we performed a natural language processing by text-mining analysis and with unsupervised (machine learning) topic modeling on a sample of the last 10,000 tweets in English posted with the term #autism (January 2021). We built a network of words to visualize the main dimensions representing these data. Second, we performed precisely the same analysis with all the articles using the term “autism” in PubMed without time restriction. Lastly, we compared the results of the 2 databases.

Results: We retrieved 121,556 terms related to autism in 10,000 tweets and 5.7x109 terms in 57,121 biomedical scientific articles. The 4 main dimensions extracted from Twitter were as follows: integration and social support, understanding and mental health, child welfare, and daily challenges and difficulties. The 4 main dimensions extracted from PubMed were as follows: diagnostic and skills, research challenges, clinical and therapeutical challenges, and neuropsychology and behavior.

Conclusions: This study provides the first systematic and rigorous comparison between 2 corpora of interests, in terms of lay representations and scientific research, regarding the significant increase in information available on autism spectrum disorder and of the difficulty to connect fragments of knowledge from the general population. The results suggest a clear distinction between the focus of topics used in the social media and that of scientific communities. This distinction highlights the importance of knowledge mobilization and exchange to better align research priorities with personal concerns and to address dimensions of well-being, adaptation, and resilience. Health care professionals and researchers can use these dimensions as a framework in their consultations to engage in discussions on issues that matter to beneficiaries and develop clinical approaches and research policies in line with these interests. Finally, our study can inform policy makers on the health and social needs and concerns of individuals with autism and their caregivers, especially to define health indicators based on important issues for beneficiaries.

(J Med Internet Res 2022;24(6):e32912) doi:10.2196/32912

XSL•FO
RenderX

KEYWORDS

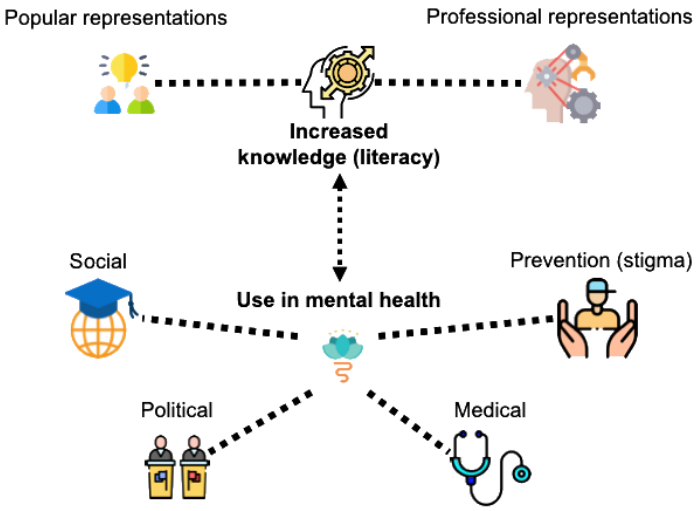
autism spectrum disorder; Twitter; natural language processing; network analysis; popular understanding of illness; knowledge translation; autism; tweets; psychiatry; text mining

Introduction

Autism is characterized in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition by the following core deficits: impairments in social interaction and communication as well as restricted, repetitive behaviors [1]. In the 2010 Global Burden of Disease study, an estimated 52 million people had autism worldwide, equating to a prevalence of 1 in 132 individuals [2]. Autism spectrum disorder (ASD) is a neurodevelopmental condition associated with significant health costs, including medical and health care–related service costs, therapeutic costs, education costs, costs of production loss for adults with ASD, costs of informal care and lost productivity for family or caregivers, and costs of accommodation, respite care, and out-of-pocket expenses. The lifetime cost of supporting an individual with ASD is about US $2.4 million, for an individual with ASD and intellectual disability and US $1.4 million for an individual with ASD without intellectual disability [3]. Support service could considerably reduce these costs, but this depends on understanding popular representations of the condition.

Understanding these lay representations is essential for at least 3 reasons; first, citizens' decisions can influence research and care policies; second, social, political, and medical attitudes toward mental health recipients are motivated by beliefs about the nature of ASD, especially because lay representations interfere with the understanding of diagnosis and treatment [4]; and finally, popular concepts of mental health have their own logic and implications, are not simply pale reflections of professional concepts filtered through the media, and hence are not shallow, incomplete, and outdated [5]. In fact, there may be significant differences between the perspectives of beneficiaries and researchers. While researchers focus on uncovering underlying mechanisms, beneficiaries and their families may be more concerned with interventions that can immediately improve quality of life. Thus, it becomes critical that mental health professionals and decision makers should be aware of popular representations, especially for high-profile topics such as ASD. Studies on "mental health literacy" are evidence of this lack of overlap between expert and popular concepts [6]. An increase in mental health literacy can have a direct impact on the appropriate use of mental health services [7]. To summarize, popular representations of ASD should be taken into account for the purpose of destigmatization, prevention, education, addressing cultural differences, and developing effective health policies [8]. Figure 1 presents a model of cooperation between lay and expert representations that allows the increase in knowledge on these topics.

Figure 1. Collaborative learner model between laypeople and health professionals for the improvement of knowledge about mental health (and literacy).



Methods for exploring popular representations of a disorder or mental condition are evolving rapidly [9]. Social media provide one way to gain access to popular representations of ASD. As social media have become ubiquitous in everyday life, they represent an easily accessible source of large data sets that reflect popular representations on a wide range of health topics [10]. The aggregation of data from social media can provide insights into first-person experience and daily life engagements with various health conditions.

More specifically, Twitter is the most widely used social media in public health and is considered the "Internet radio." An estimated 3.5 million users visit Twitter each month [11], with 336 million monthly active users and 500 million tweets sent per day and a high global adoption rate (77% of its users are located outside the United States). In addition, while most social media data (eg, Facebook) remain private, all Twitter's data are publicly available. A significant portion of Twitter's messages focus on health-related topics [12], and beneficiaries are increasingly turning to it to keep abreast of health developments and better understand their condition [11,12]. Twitter can thus potentially serve as a large-scale interactive platform to reach ASD-affected communities that may be difficult to reach through traditional means.

XSL•FO
RenderX

A brief search of the literature on ASD using the term "mining" shows that only 60 articles have been published on this subject, and that most concern the genetics of autism (19/60, 32%). To our knowledge, the most comprehensive literature review of textual analysis of autism on social media identified 5 articles [13-17]. This study is the first to analyze lay representations of autism and compare them with those of the scientific community.

Methods

We compared the representations of autism in the social media and the scientific literature using the following three steps: (1) textual analysis of data from Twitter; (2) textual analysis of articles on PubMed; and (3) the comparison of these two corpora to highlight the differences between lay representations (by proxy via Twitter) and scientific representations (by proxy via PubMed).

Textual Analysis of Twitter

In the first phase, we conducted an extraction and a textual analysis of mentions of autism on Twitter on January 21, 2021, using a sample of the last 10,000 tweets in English, posted using the term "#autism." We did not apply any geographical limits. We made sure to choose a random day, and especially to avoid any particular day of the year or a specific day related to autism (eg, the entire month of April, aka the "autism month"). This number of tweets (N=10,000) constitutes the maximum number of tweets that can be retrieved by the application programming interface (API) of this platform. All data were obtained through the official Twitter API. We did not use the term "autis*" in order to specifically target tweets related to a theme centered mainly on autism. The term "#autism" can better identify tweets that target a specific message on autism, because the hashtag can help to avoid tweets in which the theme of autism is only secondary (the "hashtag" [#] groups together all the discussions referring to the same theme). No other apparently irrelevant term was deleted, assuming that the relevance of such an analysis lies in reading the topics considered as patterns and not at the level of over-selected results. We extracted the terms most commonly associated with autism using the R software (version 4.0.3, The R Foundation; packages: stringr, rtweet, tidyr, tidytext, rjson, and leaflet) and analyzed the resulting big data on the Béluga supercomputer (*Compute Canada*) in 4 steps.

First, we prepared the data (data munging) by removing the URLs and hyperlinks (high-frequency words) by applying the standard techniques of preprocessing data for text analysis, which comprises (1) punctuation marks, (2) numeric characters, (3) spaces, and (4) special symbols. We associated a unique identifier with each occurrence of a word. Stop words, the most common words in a language that have no interest in the analysis (eg, "and," and "for") were also removed (tidytext package).

Secondly, we carried out a bootstrap analysis with an algorithm specifically created for the study, to analyze words and not tweets, in order to extract the first terms and test the network's stability (NetworkTools package).

Thirdly, we performed an undirected lexical network analysis (qgraph package) to map these terms and dimensions in space and observe their relationships, following the network analysis guidelines for cross-sectional data published by Burger et al [18] (Multimedia Appendix 1). We extracted the 50 main common words to build this network. Thus, the lexical network was constructed by Twitter mining with the term #autism with the first 50 terms. Such a lexical network can be considered as undirected because it does not describe a causal or directional relationship between the nodes. A connection between any 2 terms is represented if these terms are frequently present together in a tweet or an article. In network analysis, the potential importance of nodes within the network can be assessed by the following 4 local measures of network centrality: *strength*, *closeness*, *betweenness* (interval between nodes), and *expected influence* [19]. The *strength* of a node measures the weighted number of connections for a given node, thus showing the degree of involvement of this node in the network. *Closeness* is inversely proportional to the shortest average distance to all other nodes. *Betweenness* measures the degree to which a given node acts as a "bridge" connecting different parts of the network, reflecting the degree by which it controls the flow of information across the network. *Expected influence* is computed to improve the measurement of the centrality of nodes in the network, reflecting the influence on the symptom network based on its positive correlations, negative correlations being corrected by this centrality measure.

Finally, we used an unsupervised machine learning technique capable of scanning tweets to detect word patterns and to automatically cluster word groups, with an algorithm capable of determining the ideal number of clusters, based on the computation of a dissimilarity matrix with Euclidean distance and a cluster analysis by a k-means method (k-means reallocation clustering minimizes within-cluster variances, that is, squared Euclidean distances; it minimizes the sum of distances between the points and their respective cluster centroid; NbClust package). The result is given in the form of a numerical clustering index; in this study, the index used is the C-index: the higher the C-index, the more relevant the number of clusters. This cluster analysis provides a representation of the main semantic dimensions. More specifically, the cluster analysis is based on a method entitled Latent Dirichlet Allocation (LDavis package), which is a generative statistical model explaining sets of observations through unobserved groups, which are themselves defined by data similarities (topicmodels package). The dimensions extracted according by Latent Dirichlet Allocation correspond to the terms that frequently occur together, based on how frequently the word is on an exact topic. This cluster analysis is thus called topic modeling. Each of these topics is labeled on expert opinion according to the lexical fields of the terms found in the clusters. Successive blind iterations were performed until a common agreement was found between the authors. The labels are not methodologically formalized, but they are given according to the terms extracted and presented in the figure. Each of these labels, interpreted qualitatively, should be read according to the sets of terms of the figure.

Textual Analysis of PubMed

In the second phase, we performed an extraction and a textual analysis of mentions of autism using the biomedical database

XSL•FO
RenderX

PubMed, in January 2021 (the same day as Twitter's collection of terms), using the term "autis*" (Multimedia Appendix 1). The analysis of terms included titles, keywords (including MeSH [Medical Subject Headings]), and abstracts. We used for the PubMed analysis the same 4 steps procedure as for the Twitter database (data preparation, bootstrap analysis, undirected lexical network analysis, and clustering with topic modeling).

Comparison of Corpora

The comparison between the 2 corpora has been carried out qualitatively. Despite the absence of the possibility of developing quantitative analyses due to the intrinsically different labels of the networks, the development of similar analysis tools for each database (associated with the parallel extraction of dimensions) will allow room for interpretation in the Discussion section.

Ethical Considerations

Internet-related research raises specific ethical considerations as to whether the obtained data belong to the public or private domain [20], with respect for confidentiality and valid consent [21]. The handles were not kept in the results and therefore all the data remain at the level of the statistical aggregate. Only publicly available data on the web and collected from the Twitter platform were analyzed. We only used data from anonymous users who consented to publicly disclose their data on Twitter (ie, no privacy settings were selected by users; Multimedia Appendix 1).

Results

Data Mining and Bootstrapping

Textual Analysis of Twitter

The request was launched in January 2021, searching for 10,000 unique tweets (not retweeted) using the term "#autism." We retrieved 121,556 terms related to the term #autism. All these tweets (not retweeted) were posted in English on Twitter in less than a month, without geographical limits.

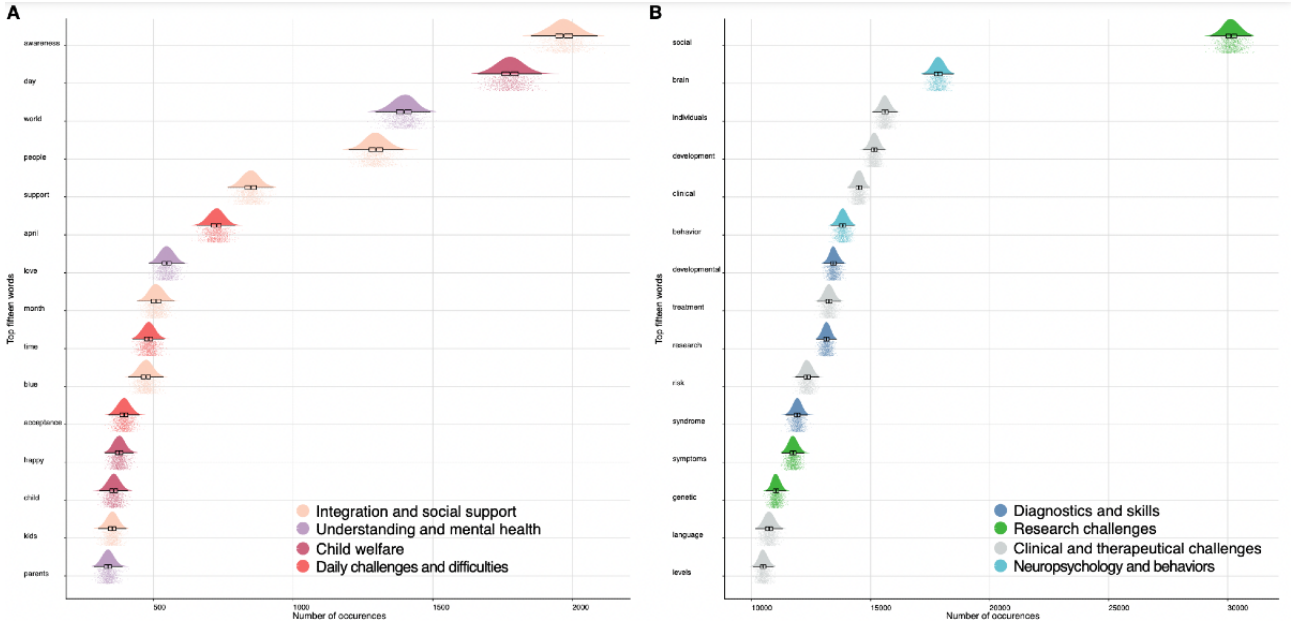
Textual Analysis of PubMed

In January 2021, the PubMed literature search identified 57,121 articles with the term "autis*" and with $5.7x10^9$ terms. The distribution of items over time (timeline) shows an average of 654 items per year smoothed over the last 70 years, with the first articles published in 1946.

The bootstrap algorithm (allowing to analyze words and not tweets) extracts the most frequently used terms and allows to test the network's stability. The confidence intervals (frequency of terms in tweets) of the bootstrap are narrow. Figure 2 shows the 16 terms found most frequently associated with the searched keyword on Twitter (panel A) and PubMed (panel B). The networks were stable at 1000 iterations, as shown by the bootstrap analyses.

Based on globally equivalent distributions, the frequency of the first 15 terms found in the PubMed database is between 10,000 and 18,000 times (ie, a frequency of about $2.5x10^{-4}$). The frequency of the first 15 terms found in the Twitter database is between 250 and 2000 times (ie, a frequency of about 0.92%) with different distributions (Figure 2).

Figure 2. A. 15 first terms found on Twitter by textual analysis on 10,000 tweets in paired analysis using the term #autism (January 2021). The term "autism" itself has been removed from the list for better visibility. B. 15 first terms found on PubMed by textual analysis on 57,121 articles using the term "autis*" (January 2021). The term "autism" itself has been removed from the list for better visibility.



XSL•FO
RenderX

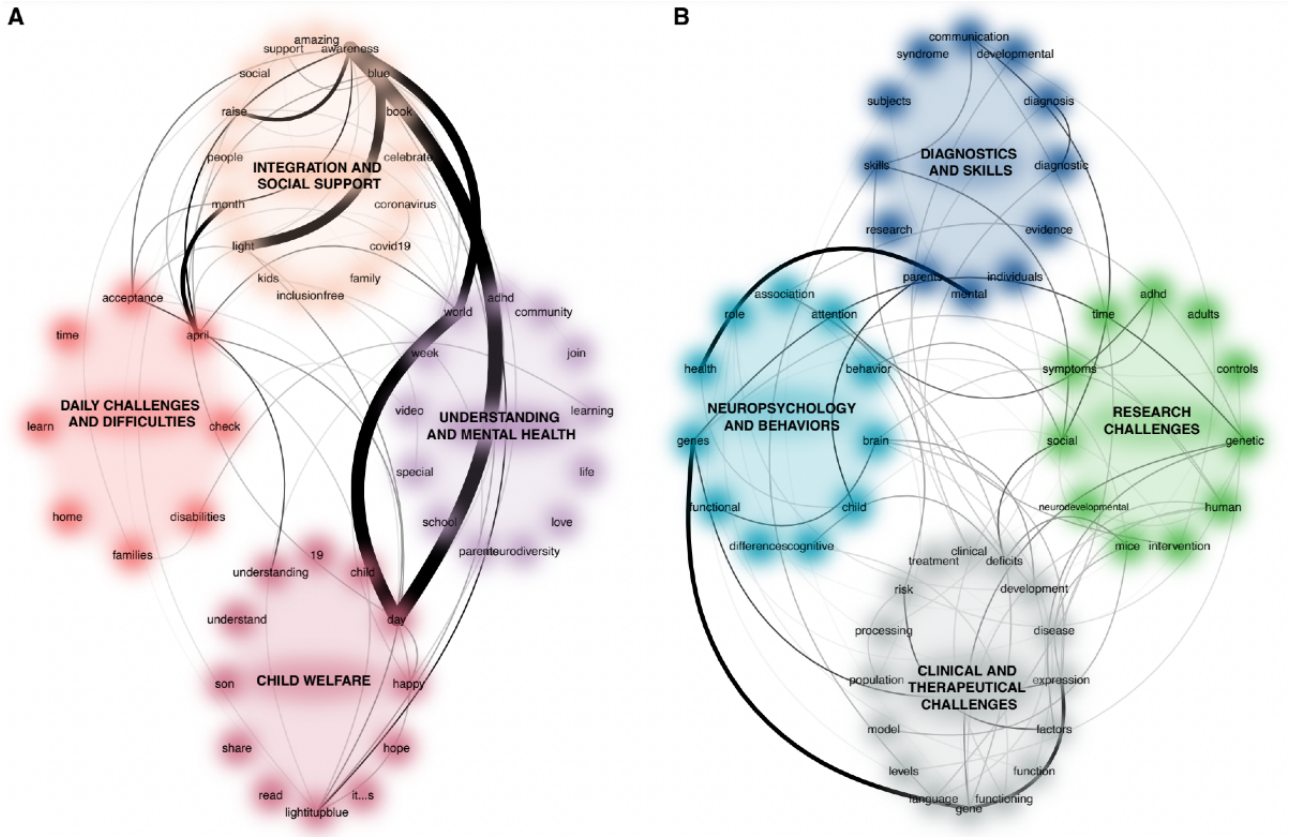
Undirected Lexical Network Analyses

The first lexical network was constructed by Twitter mining using the term #autism with the first 50 terms among the 121,556 terms in the 10,000 unique tweets. The second lexical network was constructed by PubMed mining with the term "autis*" with the first 50 terms (among the $5.7x10^9$ terms in the 57,121 articles). Figure 3 shows these undirected lexical network analyses. Remember that a connection between any 2 terms is represented if these terms are frequently present together in a tweet or an article.

The qualitative analysis of these networks shows marked differences between the terms associated with autism on Twitter and those in PubMed. For example, on Twitter, we found terms such as "Son," "Love," "Happy," or "April" (referring to the World Autism Month) denoting special community attention to the idea of well-being in ASD, but also others such as "Neurodiversity," "ADHD" (for attention deficit hyperactivity disorder), and "Brain" focusing attention on the scientific aspects of ASD. Moreover, terms such as "Acceptance," "Social," or "Understanding" were related. Among the most common terms in the literature, the PubMed analysis revealed words belonging to various scientific lexical fields such as "Gene," "Brain," or "Cognitive," denoting a special scientific focus on physiopathological mechanisms, similar to the terms "Communication," "Skills," or "Social," denoting special focus on the functional issues involved in ASD.

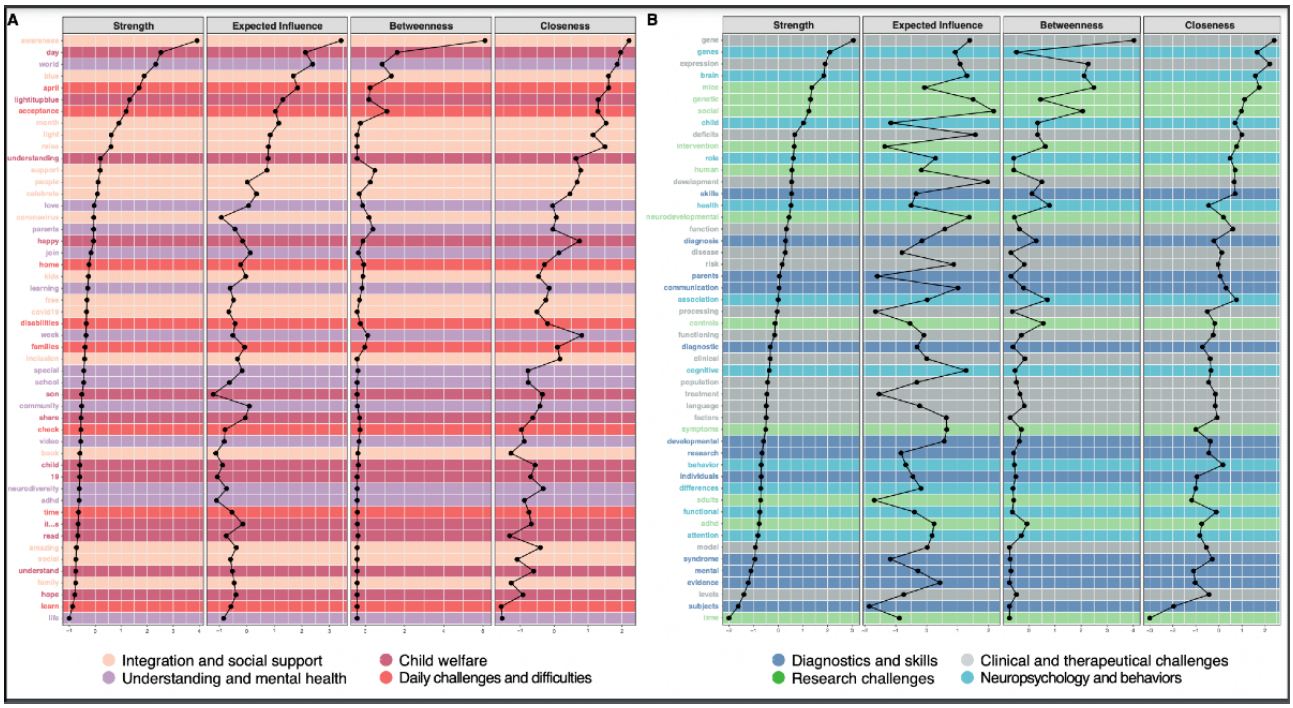
The centrality measures of these networks (Figure 4) showed that in Twitter, the 5 most central terms related to strength, expected influence, and closeness were as follows: "Maths," "Goals," and "Stem." In terms of betweenness, they were related to "Students," "Maths," and "Disabilities." In PubMed, the 3 most central terms related to expected influence were as follows: "Neurodevelopment," "Attention," and "Deficit." In terms of strength, they were "Expression," "Gene," and "Brain." In terms of betweenness, they were also "Expression," "Gene," and "Social." In terms of closeness, they were also "Expression," "Gene," and to "Development."

Figure 3. A. Directed network of 50 first words from Twitter with #autism after textual analysis of 10,000 tweets. B. Directed network of 50 first words from PubMed with "autis*" after textual analysis of all the articles on the database.



XSL•FO
RenderX

Figure 4. Four centrality measures (strength, expected influence, betweenness, and closeness), ordered by expected influence of A. Twitter network and its 4 dimensions; B. PubMed network and its 4 dimensions (details of the different centrality measures are in the Methods section). The most central node in terms of strength, for each database, is located at the top of each of the tables (eg, node "Awareness" for the Twitter network and node "Gene" for the PubMed network). A high centrality node is located to the right of each table, and a low centrality node is located to the left.



Topic Modeling to Extract Main Semantic Dimensions

We determined the number of clusters and identified the best clustering scheme by varying all combinations of the number of clusters, distance measures, and clustering methods. For both Twitter and PubMed, the best clustering schematization algorithm (based on the Euclidean distances and k-mean clustering) proposed 2 dimensions with a C-index of 235.4, and 4 dimensions with a lower C-index of 183.7. We retained this quantitative result but decided to use 4 dimensions for both networks for semantic reasons. After iterative tests with values greater than 4, it seemed more appropriate to use 4 dimensions to interpret and compare the 2 networks, and to compare the 2 corpora more closely. This approach is clinically relevant, since only 2 isolated dimensions would not be sufficiently informative in a study comparing different dimensions within different corpora. The 4 main dimensions extracted are shown in Figure 3. The dimensions in the Twitter network were as follows: (1) integration and social support; (2) understanding and mental health; (3) child welfare; and (4) daily challenges and difficulties. The dimensions in PubMed were as follows: (1) diagnosis and skills; (2) research challenges; (3) clinical and therapeutical challenges; and (4) neuropsychology and behavior.

Discussion

Principal Results

We produced 2 corpora of texts from web-based platforms representing a global social network (Twitter) and international biomedical research (PubMed), reflecting both the interests and representations of the scientific community and mental health professionals. The content of the tweets and scientific articles were explored by searching within these samples for the terms #autism and autis*, respectively. We found a wide divergence in the focus of these corpora, which has implications for how researchers and the public at large understand each other's discourse.

The 4 dimensions extracted from Twitter (integration and social support, understanding and mental health, child welfare, and daily challenges and difficulties) reveal a discourse focusing on the education, evolution, and support for people with ASD that makes relatively little mention of science or technical issues per se. Likewise, the 4 dimensions extracted from PubMed (diagnosis and skills, research challenges, clinical and therapeutical challenges, and neuropsychology and behavior) reflect scientific research practices and reveal an interest in behavioral and linguistic issues, comorbidities ,and neuroscientific topics (eg, neuropsychology, neurogenetics, and neuropharmacology), with very little mention of the issues dear to Twitter users.

Since several scientific communities are involved, the construct of ASD has been fraught with controversy [22,23]. ASD is a multiscale condition, and its scientific study requires different levels of analysis and generates various points of view, with each community providing its own perspective. These specialists include neuroscientists, biologists, pharmaceutical engineers, clinical researchers, and practitioners. Lay communities interested in ASD include parents, parents' and beneficiaries' associations, and associations that support beneficiaries daily. The topics commonly discussed by the scientific community did not appear to overlap with the concerns of the public. This lack of overlap has 3 important implications for research. First, from a scientific perspective, the range of topics covered by Twitter could provide health care professionals and researchers with insights that they would otherwise not have (Figure 1).

XSL•FO
RenderX

Second, from an economic perspective, the study of social media can increase health professional's knowledge of public and patient concerns at low cost, given the relatively free access to personal data that people disseminate about themselves. Third, in terms of public health and interventions, gaining such knowledge of the centers of interest and concerns of the public and patients could help them target social media communities involved with ASD in terms of information and interventions, all at a minute cost. Indeed, the terms found in tweets could reveal trends in the representation of mental disorders or conditions. This would help professionals provide appropriate public health information, set up prevention campaigns, destigmatize the ASD condition, and engage decision-making through an indirect suggestion for significantly changing their representational incentives. In consequence, such divergences may be useful for scientific dynamics, public health information and prevention efforts, psychoeducation, or management of care.

While the representations of ASD found in popular discourse and in biomedicine differ, they do not diverge completely. In fact, the views of professionals regarding psychiatric disorder and conditions, comorbidities, and their implications for families have a considerable influence on social media discussions. However, the differences of focus in the medical literature and the general population are important, which involves implications in terms of naturalization of a disorder and overmedicalization of a condition. In popular representations and therefore in social media, these representations may provoke negative attitudes, encouraging people to consider such a condition as deeply ingrained or constitutional and categorically different from people considered neurotypical [24]; they may also increase stigmatization or increases perceptions of dangerousness and unpredictability about patients [25]. Certainly, ASD poses a particularly challenging issue since laypeople may consider that those with ASD are biologically but not pathologically different, particularly in the context of neurodiversity [26]. However, a uniquely medical understanding of a condition such as ASD may lead to the conviction that those with ASD are unable to function normally, a belief that can lead to pessimism and disengagement if it is held widely by the public [27].

Limitations

This study has several limitations, 5 of which will be discussed here. First, for methodological reasons related to text processing, we limited the searches to the English language and to 2 databases, although both produced large corpora representative of very heterogenous populations. Indeed, texts on social media are known to contain a large proportion of nongrammatical constructions derived from abbreviations and metaphorical uses [28], while scientific texts are known for their specialized vocabulary and characteristic structure that make them a genre per se. Thus, automated comparison of the 2 requires careful annotation. On the other hand, automated methods allow large corpora to be explored, a task that would indeed be challenging if it were to be performed by close textual analysis. Furthermore, they can reveal patterns despite the existence of significant differences in the structure of discourse from various origins [29]. But even if it is possible that the 2 corpora (Twitter and PubMed) do not simply represent 2 different populations or perspectives perfectly distinctive (and are only 2 different modes of communication), this does not change the scope of our results because the claims about differences in concerns regarding the 2 corpora remains; the themes are different in terms of interests. However, we do not wish to exclude individual representations conveyed by the health professionals, who are a significant part of the social landscape.

Second, in our study, first-person representations of individuals with ASD active on Twitter were not considered. This could have been carried out by combining the terms "#autism" and "I." A recent report noted that about 80% of adults living with ASD use social media. Examination of first-person representations would provide a valuable avenue for future research on phenomenology and the mediatized social presentation of self [13]. Third, the use of isolated words and not n-grams (contiguous sequence of terms) could constitute a future perspective, since it would be a question of analyzing the relevance of pertinent contiguous occurrences in ASD (eg, the term "gene" does not have the same meaning if it is associated with "expression" or with "deficit," providing potential important information on the genetics of ASD).

Fourth, this study has the limitations of any study based on social networks and textual data mining, including the sampling bias related to users of this medium (eg, in terms of social class or culture), and the difficulty of checking user profiles [15], search parameters, or the necessary qualitative labeling of clusters. In particular, the use of the hashtag ("#autism") cannot provide the same precision as the MeSH term ("autis*"), with a potential number of false-positive tweets (ie, not specifically related to ASD). In addition, concerning the PubMed corpus, we may have included articles with the term "autis*" in the abstract but not related to autism, even though this eventuality was potentially rare.

Fifth, for computational challenges, we had to limit the number of calculations, so we were unable to perform a series of data extractions. Indeed, the Twitter API limits data extraction for ethical reasons. However, the limitations related to technological access and data processing were offset by the rewriting of specific algorithms for the purpose of this study and by the amount of data that we acquired. The narrowness of the confidence intervals of the frequency of terms obtained by bootstrapping confirms the accuracy and robustness of our data estimation. These limitations are also partially offset by the reduction in social desirability and recall bias compared to traditional survey data collection methods. This observation calls for future studies integrating a variety of data from different sources.

Conclusion

Apart from the stigma that ASD induces, the exposure of ASD also leads to suffering in relatives of people with ASD, who feel a sense of social injustice in that clinical research is still unable to meet their expectations for their loved ones. This paper illustrates the potential to use social media as a proxy for the representations of ASD in the society today. The results suggest a clear distinction between the focus of topics used in the social media and that of scientific communities. This

XSL•FO
RenderX

highlights the importance of knowledge mobilization and exchange to better align research priorities with personal concerns and to address dimensions of well-being, adaptation, and resilience. Similar methods could be used to develop pedagogical or preventive programs, or to help in establishing recommendations for treatment. The analysis of representations prevalent in the social media could also be used to design destigmatization campaigns and to assess their impact over time. As noted by Hacking [30], such representations strongly impact not only care but also research and nosology. The interaction between shared representations and medical research interests could help public health decision makers and mental health professionals to create collaborative learning environments that engage beneficiaries and caregivers.

Acknowledgments

The authors thank Calcul Quebec and Compute Canada for providing access to the Béluga supercomputer.

Authors' Contributions

CG carried out the writing, original draft preparation, conceptualization, and software handling; JM was responsible for methodology, software, and visualization; JAMF was responsible for investigation, writing, methodology, and supervision; DG carried out the supervision, formal analysis, visualization, methodology, validation, and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary files.
[DOCX File , 48 KB - jmir_v24i6e32912_app1.docx]

References

1. Hales RE, editor. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5). Arlington, VA, US: American Psychiatric Association; 2013.
2. Lord C, Brugha TS, Charman T, Cusack J, Dumas G, Frazier T, et al. Autism spectrum disorder. Nat Rev Dis Primers 2020 Jan 16;6(1):5 [FREE Full text] [doi: 10.1038/s41572-019-0138-4] [Medline: 31949163]
3. Rogge N, Janssen J. The Economic Costs of Autism Spectrum Disorder: A Literature Review. J Autism Dev Disord 2019 Jul;49(7):2873-2900. [doi: 10.1007/s10803-019-04014-z] [Medline: 30976961]
4. Hacking I. The looping effects of human kinds. In: Causal Cognition: A Multidisciplinary Debate. New York, US: Clarendon Press/Oxford University Press; 1995.
5. Na S, Ryder AG, Kirmayer LJ. Toward a Culturally Responsive Model of Mental Health Literacy: Facilitating Help-Seeking Among East Asian Immigrants to North America. Am J Community Psychol 2016 Sep;58(1-2):211-225. [doi: 10.1002/ajcp.12085] [Medline: 27596560]
6. Jorm AF. Mental health literacy. Public knowledge and beliefs about mental disorders. Br J Psychiatry 2000 Nov;177:396-401. [doi: 10.1192/bjp.177.5.396] [Medline: 11059991]
7. von Wagner C, Steptoe A, Wolf MS, Wardle J. Health literacy and health actions: a review and a framework from health psychology. Health Educ Behav 2009 Oct;36(5):860-877. [doi: 10.1177/1090198108322819] [Medline: 18728119]
8. Harrington JW, Rosen L, Garnecho A, Patrick PA. Parental perceptions and use of complementary and alternative medicine practices for children with autistic spectrum disorders in private practice. J Dev Behav Pediatr 2006 Apr;27(2 Suppl):S156-S161. [doi: 10.1097/00004703-200604002-00014] [Medline: 16685182]
9. Skelton J, Croyle R. Mental Representation in Health and Illness. New York, US: Springer-Verlag; 1991.
10. Liang P, Dai B. Opinion Mining on Social Media Data. 2013 Presented at: IEEE 14th International Conference on Mobile Data Management; June 03-06, 2013; Milan, Italy. [doi: 10.1109/mdm.2013.73]
11. Antheunis ML, Tates K, Nieboer TE. Patients' and health professionals' use of social media in health care: motives, barriers and expectations. Patient Educ Couns 2013 Sep;92(3):426-431. [doi: 10.1016/j.pec.2013.06.020] [Medline: 23899831]
12. Reavley N, Pilkington P. Use of Twitter to monitor attitudes toward depression and schizophrenia: an exploratory study. PeerJ 2014;2:e647 [FREE Full text] [doi: 10.7717/peerj.647] [Medline: 25374786]
13. Hswen Y, Gopaluni A, Brownstein JS, Hawkins JB. Using Twitter to Detect Psychological Characteristics of Self-Identified Persons With Autism Spectrum Disorder: A Feasibility Study. JMIR Mhealth Uhealth 2019 Feb 12;7(2):e12264 [FREE Full text] [doi: 10.2196/12264] [Medline: 30747718]
14. Beykikhoshk A, Arandjelovi? O, Phung D, Venkatesh S, Caelli T. Data-mining twitter and the autism spectrum disorder: A Pilot study. 2014 Presented at: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM); August 17-20, 2014; Beijing, China. [doi: 10.1109/asonam.2014.6921609]

XSL•FO
RenderX

15. Beykikhoshk A, Arandjelović O, Phung D, Venkatesh S, Caelli T. Using Twitter to learn about the autism community. Soc. Netw. Anal. Min 2015 Jun 2;5(1):11. [doi: 10.1007/s13278-015-0261-5]
16. Shakes P, Cashin A. An Analysis of Twitter Discourse Regarding Identifying Language for People on the Autism Spectrum. Issues Ment Health Nurs 2020 Mar;41(3):221-228. [doi: 10.1080/01612840.2019.1648617] [Medline: 31674850]
17. Arandjelovic O. Overcoming Data Scarcity of Twitter: Using Tweets as Bootstrap with Application to Autism-Related Topic Content Analysis. 2015 Presented at: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining; August 25-28, 2015; Paris, France p. 1354-1361. [doi: 10.1145/2808797.2808908]
18. Burger J, Isvoranu A, Lunansky G, Haslbeck J, Epskamp S, Hoekstra R, et al. Reporting standards for psychological network analyses in cross-sectional data. Psychol Methods 2022 Apr 11:Online ahead of print. [doi: 10.1037/met0000471] [Medline: 35404629]
19. Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: Generalizing degree and shortest paths. Social Networks 2010 Jul;32(3):245-251. [doi: 10.1016/j.socnet.2010.03.006]
20. Dumas G, Serfass DG, Brown NA, Sherman RA. The Evolving Nature of Social Network Research: A Commentary to Gleibs (2014). Analyses of Social Issues and Public Policy 2014 May 06;14(1):374-378. [doi: 10.1111/asap.12055]
21. Hewson C, Buchanan T. Ethics guidelines for internet-mediated research. The British Psychological Society. 2013. URL: https://www.bps.org.uk/sites/www.bps.org.uk/files/Policy/Policy%20-%20Files/Ethics%20Guidelines%20for%20Internet-mediated%20Research.pdf [accessed 2022-06-03]
22. Decoteau CL, Daniel M. Scientific Hegemony and the Field of Autism. Am Sociol Rev 2020 May 22;85(3):451-476. [doi: 10.1177/0003122420922531]
23. O'Reilly M, Lester J, Kiyimba N. Autism in the Twentieth Century: An Evolution of a Controversial Condition. In: Healthy Minds in the Twentieth Century. Cham, Switzerland: Springer Nature; Sep 17, 2019:137-165.
24. Loughman A, Haslam N. Neuroscientific explanations and the stigma of mental disorder: a meta-analytic study. Cogn Res Princ Implic 2018 Nov 14;3(1):43 [FREE Full text] [doi: 10.1186/s41235-018-0136-1] [Medline: 30426319]
25. Walker I, Read J. The differential effectiveness of psychosocial and biogenetic causal explanations in reducing negative attitudes toward "mental illness". Psychiatry 2002;65(4):313-325. [doi: 10.1521/psyc.65.4.313.20238] [Medline: 12530335]
26. Bhandari P, Khanal P. Pride in autistic diversity: against treatment or for inclusion? Lancet 2016 Nov 19;388(10059):2477. [doi: 10.1016/S0140-6736(16)32176-6] [Medline: 27871741]
27. Farina A, Fisher J, Getter H, Fischer E. Some consequences of changing people's views regarding the nature of mental illness. Journal of Abnormal Psychology 1978 Apr;87(2):272-279. [doi: 10.1037/0021-843X.87.2.272]
28. Baldwin T, Cook P, Lui M, MacKinlay A, Wang L. How noisy social media text, how different social media sources? 2013 Presented at: Proceedings of the 6th International Joint Conference on Natural Language Processing (IJCNLP 2013); January 1, 2013; Nagoya, Japan p. 356-364.
29. Petric I, Urbancic T, Cestnik B, Macedoni-Luksic M. Literature mining method RaJoLink for uncovering relations between biomedical concepts. J Biomed Inform 2009 Apr;42(2):219-227 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.004] [Medline: 18771753]
30. Hacking I. 'Style' for historians and philosophers. Studies in History and Philosophy of Science Part A 1992 Mar;23(1):1-20. [doi: 10.1016/0039-3681(92)90024-z]

Abbreviations

ADHD: attention deficit hyperactivity disorder
API: application programming interface
ASD: autism spectrum disorder
MeSH: Medical Subject Headings

Edited by R Kukafka; submitted 14.08.21; peer-reviewed by N Cheak-Zamoran, C Giraud-Carrier; comments to author 03.11.21; revised version received 25.03.22; accepted 04.04.22; published 15.06.22.

Please cite as:
Gauld C, Maquet J, Micoulaud-Franchi JA, Dumas G
Popular and Scientific Discourse on Autism: Representational Cross-Cultural Analysis of Epistemic Communities to Inform Policy and Practice
J Med Internet Res 2022;24(6):e32912
URL: https://www.jmir.org/2022/6/e32912
doi:10.2196/32912
PMID:35704359

XSL•FO
RenderX

©Christophe Gauld, Julien Maquet, Jean-Arthur Micoulaud-Franchi, Guillaume Dumas. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 15.06.2022. 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

Remodeling the Medication Collection Process With Prescription in Locker Box (PILBOX): Prospective Cross-sectional Study

Christina Lim Jit Fan[1*], MSc; Goh Boon Kwang[1*], BSc; Vivian Chee Wing Ling[1*], BSc; Tang Woh Peng[1*], BSc; Bandy Goh Qiuling[1*], MSc

Department of Allied Health (Pharmacy), SingHealth Polyclinics, Singapore, Singapore

[*]all authors contributed equally

Corresponding Author:
Bandy Goh Qiuling, MSc
Department of Allied Health (Pharmacy)
SingHealth Polyclinics
167 Jalan Bukit Merah Connection One (Tower 5)
Singapore, 150167
Singapore
Phone: 65 63507451
Email: bandy.goh.q.l@singhealth.com.sg

Abstract

Background: Traditionally, patients wishing to obtain their prescription medications have had to physically go to pharmacy counters and collect their medications via face-to-face interactions with pharmacy staff. Prescription in Locker Box (PILBOX) is a new innovation allowing patients and their caregivers to collect medication asynchronously, 24/7 at their convenience, from medication lockers instead of from pharmacy staff.

Objective: This study aimed to determine the willingness of patients and caregivers to use this new innovation and factors that affect their willingness.

Methods: This prospective cross-sectional study was conducted over 2 months at 2 public primary health care centers in Singapore. Patients or caregivers aged 21 years and older who came to pharmacies to collect medications were administered a 3-part questionnaire face-to-face by trained study team members after they gave their consent to participate in the study.

Results: A total of 222 participants completed the study. About 40% (89/222, 40.1%) of participants were willing to use PILBOX to collect their medications. Among participants who were keen to use the PILBOX service, slightly more than half (47/89, 53%) were willing to pay for the PILBOX service. Participants felt that ease of use (3.5 [SD 1.2]) of PILBOX was the most important factor affecting their willingness to use the medication pickup service. This was followed by waiting time (3.4 [SD 1.3]), cost of using the medication pickup service (3.0 [SD 1.4]), and 24/7 accessibility (2.6 [SD 1.4]). This study also found that age (P=.01), language literacy (P<.001), education level (P<.001), working status (P=.01), and personal monthly income (P=.01) were factors affecting the willingness of patients or caregivers to use PILBOX.

Conclusions: Patients and caregivers are keen to use PILBOX to collect their medications for its convenience and the opportunity to save time if it is easy to use and not costly.

(J Med Internet Res 2022;24(6):e23266) doi:10.2196/23266

KEYWORDS

PILBOX; medication collection process; dispensing; remodeling; locker box; medication; pharmaceuticals; prescription

Introduction

Many health care systems in the world require patients to physically go to the dispensary or pharmacy to refill or collect their medications. This practice has been ongoing across the globe for a long time [1,2].

Dispensing medications to patients usually encompasses the provision of some basic information about the medication such as the indication, dose, frequency, and duration [3,4]. The face-to-face encounter at the dispensary or pharmacy to collect the medications provides an opportunity for patients to be counseled on the use of their medications and have their concerns pertaining to their medications addressed [5].

XSL•FO
RenderX

This long-standing practice prevails because health care providers and patients instinctively associate the collection of medications by patients as constituting a natural end to their visit to the clinic [6]. A problem may arise if patients are faced with time constraints and cannot afford the extended time waiting for their turn to see the doctor followed by waiting to collect their medications. This problem is especially aggravated for polyclinic patients due to the high volume of patients and resulting long wait times for patients. The inconvenience caused by having to physically collect the prescribed medications and the long wait time the process entails have also contributed to a high medication nonadherence rate worldwide [7].

If one were to adopt a fresh pair of lenses, a visit to the doctor needs not be irreversibly coupled with the medication collection process. It was this out-of-the-box thinking that inspired the study team to envision the development of an innovation that would decouple the medication collection process from the doctor's visit and allow the patient a choice of collecting their prescribed medication at a separate convenient occasion after having visited their doctor for consultation.

Telecommunications and technological advances such as the invention of the phone, broadband, internet, and mobile apps are potentially game-changers in these modern times [8-10]. They enable alternative medication refill or collection modes to be made available to patients [2]. An example of such an alternative is the delivery of medications to patients via courier service or mail-order pharmacy to their preferred delivery locations with a few button clicks or a phone call to the dispensary or pharmacy.

However, not all alternative medication refill or collection modes are feasible to implement, and some may not prove attractive to patients or their caregivers [11-13]. Some solutions are costly to implement and may require dispensaries or pharmacies to pass on some or all of the implementation and operational costs to users (patients or caregivers) [11] . Others may generate new problems or inconveniences to users such as requiring users to block off a long stretch of their valuable and limited time to await the delivery [13]. Improper handling of the delivered medications by the courier personnel may also have an adverse impact on medication safety [14]. Medication safety may present a significant problem to successful implementation of these alternative medication refill or collection modes, particularly in countries with underdeveloped transport networks and infrastructures [15]. In such countries, the ability of courier personnel to safeguard the quality of the medications during transit may not be a given, even with their best efforts.

The current health care system in Singapore, like many countries worldwide, requires patients to collect their medications physically at the hospital or clinic dispensaries or retail pharmacies. Singapore's health care services are provided by both private health care providers and public health care institutions. Although there are established medication delivery services run by private operators, patients may not always be prepared to set aside a few hours to wait for the delivery of their medications as they cope with long work hours and many other competing commitments in fast-paced Singapore, such as running errands and taking children to and from classes.

Thus we have envisioned and developed Prescription in Locker Box (PILBOX) as a new innovation to fill the gap. This new innovation is a first in Singapore. The idea is simple. Rather than requiring users to stay in a given location for a long stretch of time to wait for medications to be delivered to them, why not deposit the medications in a secure location for users to pick up at their convenience? Hence, the PILBOX was conceptualized and developed.

The PILBOX station consists of a few columns of lockers of different sizes with one single attached console that allows users to input a unique access code to retrieve their medication parcels from their allocated locker. The locker station is equipped with a security camera, autolock mechanism, and lock-down system to ensure the medication parcels are safely kept until their collection by their rightful owners. The security camera will deter people with malicious intent from targeting the lockers and the medications contained within. Once a medication parcel has been collected, the locker will be autolocked to ensure the locker is inaccessible to unauthorized personnel. Further, the station lock-down system will prevent power outage situations from triggering the unintended opening of lockers and unauthorized retrieval of medication parcels.

The locker station maintains an ambient temperature (ie, at or below 25 ˚C [77 ˚F]) due to 24/7 air conditioning of the room housing the station. To ensure the medications will be stored at the suitable temperature until retrieval by the user, ambient environment temperature is monitored continuously using temperature loggers. This is important to preserve the quality of the medications until collection by users.

Suitable lighting has also been installed in the locker station to facilitate easy retrieval of medications. Further, the PILBOX is equipped with an information technology (IT) system that allows easy tracking of use of the locker system and booking of the lockers for collection of medications by users. It can also generate an audit trail to ensure accountability of each occasion of access to the lockers.

As PILBOX is a new innovation in Singapore, a study was conducted to determine the willingness of patients and caregivers to use the medication pickup service and factors that may affect their use. The information gleaned will allow us to develop appropriate strategies to promote the use of the PILBOX service to patients and increase adoption. The study also aimed to test the following hypotheses:

- Participants who are willing to use PILBOX are willing to pay for the medication pickup service.
- Participants are more willing to use the PILBOX if their wait time at the pharmacy is long.
- Participants are more willing to use PILBOX if they are not satisfied with their wait time at the pharmacy.
- Participants are more willing to use PILBOX if they refill their prescriptions or collect medications from the pharmacy frequently.
- Participants are more willing to use PILBOX if the traveling time to the PILBOX (located within the polyclinic) is less.

XSL•FO
RenderX

Methods

Study Design

The prospective cross-sectional study was conducted from December 2015 to January 2016. Participants were convenience sampled at 2 polyclinics situated in the eastern part of Singapore. Polyclinics are public primary health care centers that provide medical care for acute and chronic conditions and provide services such as pharmacy, vaccinations, and health screenings. Prior to the development of PILBOX, patients who needed to collect their prescribed medications or who had prescriptions to refill were required to physically go to the polyclinics' respective in-house pharmacies where they would be served by pharmacy staff at counters.

Study Participants and Ethical Approval

The recruited participants were aged 21 years and older. Patients or caregivers visiting the in-house pharmacies at the study sites to fill or refill their prescriptions were invited to participate in the study. Potential participants who could not understand or converse in English or Mandarin were excluded from the study. This study was exempted from ethics review by SingHealth centralized institutional review board (SHP 2015/3035).

Using α=.05, a power of 80%, and an expected correlation coefficient of .30, a calculated sample size of at least 85 participants was required to test the hypotheses on willingness to use PILBOX against their (1) willingness to pay to use the service, (2) wait time at the pharmacy, (3) satisfaction with their wait time at the pharmacy, (4) frequency of prescription refills or medication collections at the pharmacy, and (5) traveling time to the PILBOX.

Study Instrument

A structured questionnaire developed by the study team was administered face-to-face by trained study team members in a standardized manner (Multimedia Appendix 1). Each participant completed the questionnaire once. The questionnaire solicited information on the following:

- Demographics (ie, age, sex, language literacy, education level, working status, housing type, and personal monthly income)
- Relevant factors that might affect participant satisfaction with pharmacy services (ie, estimated traveling time to polyclinic, frequency of prescription refills and medication collections at the pharmacy, average waiting time at the pharmacy, and satisfaction with waiting time at the pharmacy)
- Willingness to use PILBOX and pay for its use

A 5-point Likert scale (1=not keen, 5=very keen) was used to assess willingness of the participants to use PILBOX.

Statistical Methods

Chi-square or Fisher exact tests were used to determine the factors that might affect the willingness of participants to use PILBOX. Pearson correlation tests were used to determine whether participant willingness to use PILBOX correlated with their (1) willingness to pay to use the service, (2) average waiting times at the pharmacy, (3) satisfaction with their waiting times at the pharmacy, and (4) frequency of prescription refills or medication collections at the pharmacy. All analyses were performed using SPSS (version 25.0, IBM Corp) at the 5% significance level.

Results

Participant Characteristics

A total of 222 participants completed the study, with 109 females, 101 males, and 12 unspecified. Among the participants, about half (109/222, 49.1%) were persons aged 65 years and older. Almost 1 in 10 participants (19/222, 8.6%) had no formal education, and 1 in 5 (46/222, 20.7%) had primary education. Almost half (104/222, 46.9%) were not working (Table 1).

XSL•FO
RenderX

Table 1. Demographics of participants (n=222).

Characteristics	Value, n (%)
Age group, years (older adult vs adult)	
≥65	109 (49.1)
<65	99 (44.6)
Unspecified	14 (6.3)
Age (years)	
21-40	12 (5.8)
41-64	87 (41.8)
≥65	109 (52.4)
Sex	
Female	109 (49.1)
Male	101 (45.5)
Unspecified	12 (5.4)
Language literacy (self-reported)	
English	124 (40.0)
Chinese	132 (42.6)
Malay	20 (6.5)
Tamil	3 (1.0)
Other	2 (0.6)
Illiterate	14 (4.5)
Unspecified	15 (4.8)
Education level	
No formal education	19 (8.6)
Primary	46 (20.7)
Secondary	72 (32.4)
Above secondary	67 (30.2)
Unspecified	18 (8.1)
Working status	
Working	100 (45.0)
Not working	104 (46.9)
Unspecified	18 (8.1)
Housing type	
Public housing	165 (74.3)
Private housing	29 (13.1)
Other	8 (3.6)
Unspecified	20 (9.0)
Personal monthly income (S$)	
No income	106 (47.7)
≤3000	63 (28.4)
3001-6000	21 (9.5)
>6000	7 (3.2)
Unspecified	25 (11.3)

XSL•FO
RenderX

Relevant Factors That May Affect Participant Satisfaction With Pharmacy Services

Among the participants who collected their medications in installments, about half (65/150, 43.3%) refilled their prescriptions or collected their medications every 3 months. About 30% (41/150, 27.3%) refilled their prescriptions or collected their medications more than 4 times a year. The median traveling time among the participants was 15 to 30 minutes. Slightly more than half (116/222, 52.3%) of the participants were satisfied with the waiting time at the pharmacy to collect or refill their prescriptions. The median time participants spent at the pharmacies waiting to collect or refill their prescriptions was 31 to 45 minutes (Table 2).

XSL•FO
RenderX

Table 2. Relevant factors that may affect participant satisfaction with pharmacy services.

Factors	Value, n (%)
Estimated traveling time to polyclinic (minutes)	
<15	92 (41.4)
15-30	63 (28.4)
31-45	37 (16.7)
>45	13 (5.8)
Unspecified	17 (7.7)
Collection of medication in installments?	
Yes	150 (67.6)
No	72 (32.4)
Frequency of prescription refills (for patients who collect in installments)	
>1 time per month	2 (1.4)
Once every month	23 (15.3)
Once every 2 months	16 (10.7)
Once every 3 months	65 (43.3)
Once every 4 months and longer	44 (29.3)
Average waiting time at pharmacy (minutes)	
<15	16 (7.2)
15-30	72 (32.9)
31-45	78 (35.1)
>45	55 (24.8)
Satisfaction with waiting time at pharmacy	
Dissatisfied	68 (30.6)
Neutral	37 (16.7)
Satisfied	116 (52.3)
Unspecified	1 (0.4)
Willingness to use PILBOX	
Not keen at all	65 (29.3)
Not keen	25 (11.2)
Neutral	37 (16.7)
Keen	47 (21.2)
Very keen	42 (18.9)
Unspecified	6 (2.7)
Willingness to pay to use PILBOX[a]	
Willing to pay	73 (32.9)
Not willing to pay	123 (55.4)
Unspecified	26 (11.7)
Amount willing to pay to use PILBOX (S$)	
≥2	55 (75.3)
2-5	13 (17.8)
5-10	4 (5.5)
<10	1 (1.4)

XSL•FO
RenderX

[a]PILBOX: Prescription in a Locker Box.

Willingness to Use PILBOX and Willingness to Pay for the Service

About 40% (89/222, 40.1%) of participants were willing to use PILBOX to collect their medications. A similar proportion (90/222, 40.5%) were not willing to, with the rest (37/222, 16.7%) not sure if they would be keen to do so. More than half (123/222, 55.4%) of the participants were not willing to pay for the service, and about a third (73/222, 32.9%) were willing to pay to use PILBOX. Among those who were willing to pay, the majority (55/73, 75%) preferred to pay S$2 (US $1.50) or less to use PILBOX. In Singapore, this amount can pay for a single-direction bus trip or the fee for courier delivery of online purchases, including ordering of food. It also equates to roughly 5% of a typical polyclinic patient's medication bill. Among the participants who were keen to use PILBOX, slightly more than half (47/89, 53%) were willing to pay for PILBOX (Table 2).

Participants who were not keen to use the medication pickup service were hesitant because they were worried that they could not understand the directions to use the locker station. Concerns over not being able to read and understand the directions for use as indicated on the locker station because of literacy limitations and poor eyesight were reported by participants. Some participants felt that they would need assistance to guide them on the use of the locker station but were at the same time concerned that they might hold up the queue. This deterred them from using the service. A small percentage (3/222, 1.4%) did not own a cellphone, and 3.6% (8/222) felt they were not tech savvy enough to be able to use the automated collection system. Some participants were concerned that the payment kiosks at the locker station would not be able to support the use of their medical benefit cards.

A common reason cited by participants was that they did not mind waiting at the pharmacy for their medications as they were already accustomed to waiting for their medications. Some participants did not feel safe collecting their medications from the locker station because they were worried they might get the wrong medications or wrong quantities of the medications from the lockers. Some preferred to collect their medications in person at the pharmacy as they appreciated the availability of pharmacy staff to address any medication queries or concerns they might have. They also wanted someone to explain the use of their medications to them in person.

On the other hand, reasons motivating participants to use PILBOX include no waiting time at the pharmacy (which would save time for busy people) and being able to collect their medications at a time convenient to them. The medication pickup service would presumably appeal more to tech savvy individuals, as suggested by one participant in the survey.

Factors Affecting Willingness to Use PILBOX

The participants felt that the ease of use (3.5 [SD 1.2]) of PILBOX was the most important factor affecting their willingness to use the medication pickup service. This was followed by no wait time (3.4 [SD 1.3]), cost of using the medication pickup service (3.0 [SD 1.4]), and 24/7 accessibility (2.6 [SD1.4]). Participants ranked the location of the locker station (2.6 [SD 1.5]) as the least important among 5 factors that might affect their willingness to use the medication pickup service.

This study found that age (P=.01), literacy (P<.001), education level (P<.001), working status (P=.01), and personal monthly income (P=.01) were factors affecting the willingness of the patients or caregivers to use PILBOX. The same list of factors was also found to be associated with their willingness to pay for use of the medication pickup service (Table 3).

XSL•FO
RenderX

Table 3. Factors affecting willingness to use and willingness to pay to use PILBOX.

Factors	Willingness to use PILBOX[a]		*P* value	Willingness to pay to use PILBOX		*P* value
	Not willing	Willing		Not willing	Willing	
Age (years), n	__[b]	—	.01	—	—	<.001
≥65	54	36	—	76	26	—
<65	31	49	—	44	47	—
Sex, n	—	—	.11	—	—	.72
Female	49	39	—	64	37	—
Male	36	47	—	56	36	—
Language literacy (self-reported), n	—	—	<.001	—	—	—
Non-English	54	18	<.001	61	15	<.001
English	30	68	<.001	59	58	<.001
Illiterate in any language	14	0	<.001	11	0	.01
Literate in at least 1 language	70	86	<.001	109	73	.01
Education level, n	—	—	<.001	—	—	<.001
Have education	65	84	—	103	73	—
No formal education	18	0	—	16	0	—
Primary and below	45	10	—	44	12	—
Secondary and above	38	74	—	75	61	—
Working status, n	—	—	.01	—	—	<.001
Not working	51	35	—	72	27	—
Working	32	49	—	47	46	—
Housing types, n	—	—	.08	—	—	.17
Private housing	7	16	—	14	14	—
Public housing	75	67	—	103	59	—
Personal monthly income, n	—	—	.01	—	—	<.001
Have income	29	46	—	43	43	—
No income	52	36	—	73	28	—
Frequency of medication collection (per year), n	—	—	.79	—	—	.17
≤4	60	25	—	69	32	—
>4	7	17	—	23	18	—
Traveling Time to Polyclinic (minutes), n	—	—	.40	—	—	.94
≤30	38	32	—	91	55	—
>30	16	9	—	29	18	—

[a]PILBOX: Prescription in Locker Box.

[b]Not applicable.

There was weak positive correlation between participant willingness to use PILBOX and their willingness to pay for use of the medication pickup service (159/222, *r*=.23, *P*<.001). Participant willingness to use PILBOX correlated weakly and negatively with waiting time at the pharmacy (180/222, *r*=–.23, *P*<.001). There was very weak negative correlation between participant willingness to use PILBOX and their satisfaction level with waiting time at the pharmacy (180/222, *r*=–.06, *P*=.47). On the other hand, there was very weak positive correlation between willingness to use PILBOX and frequency of prescription refills or collection of medications at the pharmacy (116/222, *r*=.05, *P*=.63) among patients who collected their medications in installments. Participant willingness to use PILBOX correlated weakly and negatively with traveling time to the polyclinics where the medication locker stations were located (168/222, *r*=–.14, *P*=.07).

XSL•FO
RenderX

Discussion

Principal Findings

This study found that about 40.0% of patients and caregivers were willing to use PILBOX to collect their medications. This acceptance rate of a service innovation is in line with the natural diffusion of any new products or services into use among a population of users, comprising innovators (2.5%), early adopters (13.5%), early majority (34.0%), late majority (34.0%), and laggards (16.0%) [16].

Willingness to Use PILBOX

The ease of use of PILBOX was the most important factor affecting the desirability of the new medication collection service to users (patients and caregivers). A possible explanation for this observation was that ease of use could elicit positive emotions associated with the use of PILBOX and lead to service satisfaction [17]. Hence, ease of use of PILBOX would intuitively encourage users to opt for the service [18]. As mentioned in the Results section, participants who were not keen to use PILBOX were hesitant because they were worried they would not be able to understand the directions of use as indicated on the locker stations [19], whether because of literacy limitations or poor eyesight. Some participants also provided feedback that they felt they would need assistance to guide them with use of PILBOX while others were put off because they felt they were not tech savvy enough. Further, as seen in Table 3, willingness to use PILBOX increased with education (P>.99) and literacy levels (P<.001) and decreased with age (P=.01). The working status of participants also influenced whether participants were willing to use PILBOX (P=.01), with working individuals being more willing to use the service presumably due to their greater exposure to and hence greater confidence in interacting with innovative products and services [19]. Participants living in private housing (versus public housing) and presumably deemed more affluent were also more willing to use PILBOX (P=.08), again likely due to their greater exposure to and confidence in interacting with innovative products and services [20]. From the survey results, if greater ease of use such as by providing clear and understandable instructions [21] is designed into PILBOX, there is a higher chance that it will be embraced by more users.

The second most important factor affecting user choice was the zero waiting time advantage offered by the PILBOX option. Previous studies have shown that patient satisfaction levels would reduce with increase in waiting time [22,23]. We would intuitively expect that decreased satisfaction levels on the part of patients would then motivate them to use PILBOX, which could potentially offer them the advantage of zero waiting time, since they could choose when to collect their medications.

Surprisingly, our data showed that user willingness to use PILBOX correlated weakly and negatively with waiting time at the pharmacy (180/222, r=–.23, P<.001), which was not what we would intuitively expect. This could be due to users having become accustomed to longer waiting times at the polyclinics and having realistic expectations given that they were receiving subsidized medical care. In view of these data, it was uncertain whether the zero waiting time advantage would be a factor to induce users to use the PILBOX service. However, in view that many studies have shown negative association between increased waiting time and patient satisfaction [24,25], no wait time could still be a value proposition for the use of PILBOX.

On the other hand, our results showed that patient willingness to use PILBOX correlated negatively and very weakly with patient satisfaction levels over waiting time at the pharmacies (180/222, r=–.06, P=.47). This was to be expected, given that a patient who was already satisfied with the pharmacy waiting times would be less motivated to use an alternative mode of medication collection (ie, PILBOX) [26].

The third most important factor affecting patient choice was the cost of using the medication pickup service. Due to price elasticity of demand [27], it was intuitive to expect that the need to pay for use of PILBOX would deter users from opting to use this innovative service [20]. This was borne out by our data. From Table 3, users who were working or had income were more willing to use and pay for the PILBOX service (159/222, r=.23, P<.001).

The fourth most important factor affecting patient choice was the 24/7 convenience for collection of medications offered to patients by the PILBOX service. This service would be expected to appeal to users as it would allow them to collect their medications at a time convenient to them. Our data showed that there was positive correlation between willingness to use PILBOX and frequency of prescription refills or collection of medications (ie, patients who had to endure more trips to collect their medications would be more willing to explore use of PILBOX to derive greater convenience [116/222, r=.05, P=.63]).

The least important factor affecting patient choice was the location of PILBOX. Previous studies reported mixed findings on the impact of the traveling distance to a pharmacy or dispensary on medication adherence [28,29]. Some studies reported some association between traveling distance with motivation of patients to collect and adhere to administration/dosing of their medications, whereas other studies reported minimal association [28,29]. Our result showed that willingness to use PILBOX correlated negatively and weakly with the traveling time of users for collection of medications (168/222, r=–.14, P=.07). This was consistent with previous studies in that the relation between traveling distance and willingness to invest effort to collect prescribed medications was not strong. The weak negative correlation between traveling distance and willingness to use PILBOX can be attributed to PILBOX being located at the polyclinics. Hence, while it provided greater convenience over the medication collection time, patients would still need to travel to the polyclinics to collect their medications. Given the results, there are opportunities to evolve additional options for collection of medications from locations other than polyclinics themselves. This may entail collaboration with suitable partners operating within community hubs located near the residences of patients.

Remodeling the Medication Collection Process

The rapid development of Singapore into a smart nation and advances in IT capabilities disrupting many industries (eg, transport, food and beverage, and retail industries) have also

XSL•FO
RenderX

changed the way health care providers interact with patients and their caregivers [30].

Specifically, in the area of supply of prescribed medications to patients, the conventional norm was to supply prescribed medications to patients face-to-face at pharmacy counters. In the high-patient volume environment in polyclinics [31], pharmacy staff face many challenges in providing excellent service to patients (ie, high patient volumes and prescription loads and long prescriptions with many items due to an aging population) [32-34], high expectations of a more educated patient population, and the need to fulfill patient waiting time targets [35].

Due to the high patient volume and limited space at polyclinics, one problem faced by pharmacy staff was congestion in the pharmacy waiting areas, particularly during peak hours [24]. Instead of instinctive solutions such as increasing staff numbers, which would push up costs and reduce productivity, or pressurizing staff into rushing through the dispensing process, which could compromise medication safety [36], pharmacy staff tried to think outside of the box in coming up with a solution. Breaking from the paradigm that all patients with prescribed medications had to be served within the polyclinic operating hours, pharmacy staff started to think of how some patients could be served asynchronously outside of polyclinic operating hours [37] (eg, having the medications delivered to them or being given an option to collect their medications after their polyclinic visits). From the latter idea, the staff evolved the idea of a medication pick-up service that was gradually developed and eventually fleshed out into the first-generation PILBOX.

One concern was whether this concept of medication pickup would be embraced by the patient population at large. Hence, we undertook this research study to determine the willingness by patients and caregivers to use this new innovation and find out what factors would affect their willingness to use the innovative service. The research findings provided key insights to develop the first generation PILBOX and would guide the pharmacy team to further streamline the PILBOX innovation to better meet the needs of the users so as to increase uptake of the service and enhance pharmacy overall medication supply service to patients.

Translating Research Findings Into Practice

From the study, the ease of use of PILBOX was a top concern of patients and their caregivers. Hence, when developing the first generation PILBOX service, considerable efforts were devoted to design ease of use into the innovation and assist users with service ambassadors and educational materials [38]. Pictogram decals were pasted on the locker station to provide easy reference to users on use of PILBOX. Educational materials were written at primary 6 level and below and care was taken to ensure that the message content, typography, and visuals of educational materials were designed to promote ease of use [39,40].

To further enhance PILBOX and improve uptake of the innovative service, user inputs will continue to be solicited, either via surveys or focus group discussions [41]. Efforts to educate users on this medication pickup service will continue to be undertaken so that they can confidently use the service [38].

Another finding from the study was that users were more likely to use PILBOX if they did not have to travel too far. Hence, besides giving patients the option of collecting their prescribed medications from PILBOX located within polyclinics, it may be worthwhile for the polyclinics to explore working with suitable partners operating within community hubs located near the residences of patients to make available to users the option of collecting their prescribed medications from locations nearer their homes [19,24].

Future Generations of PILBOX

PILBOX is a new innovation in Singapore that provides a solution to enable patients to pick up their medications at their convenience 24/7. This reduces the need for patients to wait at home to receive their medications via medication delivery service. The waiting time can then be diverted to better uses, translating into better value and experience for the patient. Further, this alternative mode of medication supply enables more efficient use of limited health care resources that every country faces, including Singapore. This modality of supplying medications can still be considered innovative, as the predominant mode of supplying medications to patients is still face-to-face.

Although automated locker stations may be common nowadays, they may not be suitable for medications requiring special considerations (eg, storage, security and compliance with legal requirements). The unique design of PILBOX enables maintenance of ambient temperature at 25 ℃ (77 ℉) and below. This is imperative as medications are required to be stored within the manufacturer's recommended storage temperatures to maintain their safety and integrity. Further, PILBOX's autolock mechanism and lock-down system keep the medication parcel safe and ensure public health safety by preventing unauthorized access without 24/7 human supervision or intervention.

Although the current PILBOX model was well received by patients and their caregivers, future models could include self-cooling features so the locker stations need not be placed only in enclosed locations with 24/7 air conditioning to maintain the required ambient temperatures. The future PILBOX could also have self-refrigerating system to provide appropriate storage condition for cold-chain items such as insulins. This would allow the PILBOX option to be extended to patients prescribed thermolabile medications.

Currently, only pharmacy users are able to book the lockers in PILBOX for use via a designated online portal. Future PILBOX models and IT enablers could provide patients or their caregivers and pharmacy operators with even greater convenience by allowing submission of prescription orders via a mobile app, translation of prescription orders into the pharmacy's system, picking and packing of medications by robots, and automatic placement of picked/packed medications into medication locker stations for collection by the users. The same mobile app could

XSL•FO
RenderX

also provide medication use instructions and advice for reference by users [42,43].

Limitations of Study

Currently, patients or caregivers can visit private dispensaries or pharmacies and hospital pharmacies in addition to polyclinic pharmacies to fill or refill their prescriptions. A limitation of this study is it was confined to patients and caregivers who visited polyclinics to refill their prescriptions or collect their medications. The results would have been more representative of the population if patients or caregivers who visited private dispensaries or pharmacies and hospital pharmacies had been included in the study. However, this limitation is not of great concern as the proportion of patients who visited polyclinics is significant and it would be reasonable and feasible to extrapolate the findings to the entire local population. Another limitation of this study is that only a pictorial PILBOX prototype was shown to participants during the study. The artist impression and features of PILBOX and the medication collection process involving the PILBOX might have been interpreted or imagined differently by different individuals, thereby potentially creating a visual bias.

Future Studies

With more people becoming digitally savvy and better technology providing an even more seamless experience for the user, we would expect a higher adoption rate for the use of PILBOX. Further, with people now spending more time at work or having other competing priorities such as caring for family, the convenience provided by PILBOX may also entice more to use this service and even to pay to use it. However, we are mindful that there will always be a segment of the population who will not be comfortable with this new modality of supply as they may not be digitally savvy or may just be simply averse to use of technology. They tend to be older patients who incidentally are more significant because they tend to have more medical conditions and consume more medications. Hence, while this technology is innovative and bring benefits to patients, it may require a period of time running into years for widespread subscription to its use. This study has provided a baseline that future studies could use and compare with to determine the change in user expectations and their willingness to adopt more advanced technology or services.

Conclusions

A significant proportion of patients and caregivers are keen to use PILBOX as an alternative mode to collect their medications. Addressing patient concerns such as ease of use and language literacy barriers with regard to PILBOX use will help to increase their adoption of this new service.

Acknowledgments

We would to thank Ms Huang Yufang, Mrs Lo Fei Ling, Mr Fang Qinan, and Mr Liaw Zi Yang for their contributions to this study. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. Parts of this study were presented in abstract form and as poster presentations at the Healthcare Information and Management Systems Society–Elsevier Digital Healthcare Conference on Aug 24, 2016, the 10th SingHealth Allied Health Innovation Practice Conference on Feb 24, 2017, and the National Health IT Excellence Awards on May 30, 2017.

Authors' Contributions

CLJF and GBK conceptualized and designed the research study. VCWL implemented the study and collected the data. BGQ analyzed the data. CLJF, BGQ, and TWP wrote the first draft of the manuscript. The whole team collectively reviewed and refined the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire.
[DOCX File , 16 KB - jmir_v24i6e23266_app1.docx]

References

1. Brar Prayaga R, Agrawal R, Nguyen B, Jeong EW, Noble HK, Paster A, et al. Impact of social determinants of health and demographics on refill requests by Medicare patients using a conversational artificial intelligence text messaging solution: cross-sectional study. JMIR Mhealth Uhealth 2019 Nov 18;7(11):e15771 [FREE Full text] [doi: 10.2196/15771] [Medline: 31738170]
2. Schmittdiel JA, Karter AJ, Dyer W, Parker M, Uratsu C, Chan J, et al. The comparative effectiveness of mail order pharmacy use vs. local pharmacy use on LDL-C control in new statin users. J Gen Intern Med 2011 Dec;26(12):1396-1402 [FREE Full text] [doi: 10.1007/s11606-011-1805-7] [Medline: 21773848]
3. Paulino E, Thomas D, Lee S, Cooper J. Dispensing process, medication reconciliation, patient counseling, and medication adherence. In: Thomas D, editor. Clinical Pharmacy Education, Practice and Research: Clinical Pharmacy, Drug Information, Pharmacovigilance, Pharmacoeconomics and Clinical Research. Philadelphia: Elsevier; 2019:109-120.

XSL•FO
RenderX

4. Embry M. Ensuring good dispensing practices. In: Ryan M, editor. MDS-3: Managing Access to Medicines and Health Technologies. Arlington: Management Sciences for Health; 2012:579-595.
5. Liu LST, Goh BQ, Tang WP, Lo FL, Khoo RSY, Lim CJF. Drug information needs and concerns of primary care patients with newly prescribed chronic medications. Proc Singapore Healthc 2018 Jun 13;27(4):294-298. [doi: 10.1177/2010105818779605]
6. Niska R, Ahmad F, Xu J. National Hospital Ambulatory Medical Care Survey: emergency department summary. Natl Health Stat Rep 2010:1-31. [doi: 10.1037/e587172010-001]
7. Brown MT, Bussell JK. Medication adherence: WHO cares? Mayo Clin Proc 2011 Apr;86(4):304-314 [FREE Full text] [doi: 10.4065/mcp.2010.0575] [Medline: 21389250]
8. Ruiz Morilla MD, Sans M, Casasa A, Giménez N. Implementing technology in healthcare: insights from physicians. BMC Med Inform Decis Mak 2017 Jun 27;17(1):92 [FREE Full text] [doi: 10.1186/s12911-017-0489-2] [Medline: 28655299]
9. Meneses F, Moreira A. Technology Enablers for Context-Aware Healthcare Applications. Hershey: IGI Global; 2009.
10. Ranschaert E. The Impact of Information Technology on Radiology Services: An Overview: Article based on PhD dissertation to obtain the degree of doctor in medical sciences, defended at the University of Antwerp on July 14, 2016. J Belg Soc Radiol 2016 Nov 19;100(1):93 [FREE Full text] [doi: 10.5334/jbr-btr.1184] [Medline: 30151487]
11. Palmer J. Drones could be the future of healthcare. Patient Safety Qual Healthc. 2019. URL: https://www.psqh.com/analysis/drones-could-be-the-future-of-healthcare/ [accessed 2020-06-01]
12. Scalea J, Restaino S, Scassero M, Bartlett S, Wereley N. The final frontier? Exploring organ transportation by drone. Am J Transplant 2019 Mar;19(3):962-964 [FREE Full text] [doi: 10.1111/ajt.15113] [Medline: 30203436]
13. Gordon E, Gallant M, Sehgal A, Conti D, Siminoff L. Medication-taking among adult renal transplant recipients: barriers and strategies. Transpl Int 2009 May;22(5):534-545 [FREE Full text] [doi: 10.1111/j.1432-2277.2008.00827.x] [Medline: 19175560]
14. Chawla DS. Poor medicines delivery practices prompt safety review. Pharm J. 2017. URL: https://pharmaceutical-journal.com/article/news/poor-medicines-delivery-practices-prompt-safety-review [accessed 2020-06-01]
15. Taylor J. Recommendations on the control and monitoring of storage and transportation temperatures of medicinal products. Pharm J 2001:267.
16. Dearing JW. Applying diffusion of innovation theory to intervention development. Res Soc Work Pract 2009 Sep 01;19(5):503-518 [FREE Full text] [doi: 10.1177/1049731509335569] [Medline: 20976022]
17. Malik A, Suresh S, Sharma S. Factors influencing consumers' attitude towards adoption and continuous use of mobile applications: a conceptual model. Procedia Comput Sci 2017:106-113. [doi: 10.1016/j.procs.2017.11.348]
18. Liu GS, Tai PT. A study of factors affecting the intention to use mobile payment services in Vietnam. Econ World 2019 Jun 28;4(6):249-273. [doi: 10.17265/2328-7144/2016.06.001]
19. Kassahun A, Gashe F, Mulisa E, Rike W. Nonadherence and factors affecting adherence of diabetic patients to anti-diabetic medication in Assela General Hospital, Oromia Region, Ethiopia. J Pharm Bioallied Sci 2016;8(2):124-129 [FREE Full text] [doi: 10.4103/0975-7406.171696] [Medline: 27134464]
20. Mattila M, Karjaluoto H, Pento T. Internet banking adoption among mature customers: early majority or laggards? J Serv Mark 2003;17:514-528. [doi: 10.1108/08876040310486294]
21. Tang AKY. Mobile app monetization: app business models in the digital era. Int J Innov Manag Technol 2016;7(5):224-227. [doi: 10.18178/ijimt.2016.7.5.677]
22. Camacho F, Anderson R, Safrit A, Jones AS, Hoffmann P. The relationship between patient's perceived waiting time and office-based practice satisfaction. N C Med J 2006 Nov 01;67(6):409-413. [doi: 10.18043/ncm.67.6.409]
23. Leddy KM, Kaldenberg DO, Becker BW. Timeliness in ambulatory care treatment: an examination of patient satisfaction and wait times in medical practices and outpatient test and treatment facilities. J Ambul Care Manage 2003;26(2):138-149. [doi: 10.1097/00004479-200304000-00006] [Medline: 12698928]
24. Huang XM. Patient attitude towards waiting in an outpatient clinic and its applications. Health Serv Manage Res 1994 Feb;7(1):2-8. [doi: 10.1177/095148489400700101] [Medline: 10133292]
25. Schappert S, Rechtsteiner E. Ambulatory medical care utilization estimates for 2006. Natl Health Stat Report 2008 Aug 06(8):1-29 [FREE Full text] [Medline: 18958997]
26. Yang Z, Peterson RT. Customer perceived value, satisfaction, and loyalty: the role of switching costs. Psychol Mark 2004 Oct;21(10):799-822. [doi: 10.1002/mar.20030]
27. Simon H. Dynamics of price elasticity and brand life cycles: an empirical study. J Mark Res 1979 Nov;16(4):439. [doi: 10.2307/3150805]
28. Welty TE, Willis SL, Welty EA. Effect of limited transportation on medication adherence in patients with epilepsy. J Am Pharm Assoc 2010 Nov;50(6):698-703. [doi: 10.1331/japha.2010.09081]
29. Syed ST, Sharp LK, Kim Y, Jentleson A, Lora CM, Touchette DR, et al. Relationship between medication adherence and distance to dispensing pharmacies and prescribers among an urban Medicaid population with diabetes mellitus. Pharmacotherapy 2016 Jun 28;36(6):590-597 [FREE Full text] [doi: 10.1002/phar.1757] [Medline: 27087250]
30. Ortiz E, Clancy CM, AHRQ. Use of information technology to improve the quality of health care in the United States. Health Serv Res 2003 Apr;38(2):xi-xxii [FREE Full text] [doi: 10.1111/1475-6773.00127] [Medline: 12785557]

XSL•FO
RenderX

31. Top 4 conditions of polyclinic attendances. Singapore Ministry of Health. URL: https://www.moh.gov.sg/resources-statistics/singapore-health-facts/top-4-conditions-of-polyclinic-attendances [accessed 2020-06-01]
32. Anderson LA, Goodman RA, Holtzman D, Posner SF, Northridge ME. Aging in the United States: opportunities and challenges for public health. Am J Public Health 2012 Mar;102(3):393-395. [doi: 10.2105/AJPH.2011.300617] [Medline: 22390500]
33. Wong CY, Lee HC. Healthcare in Singapore: challenges and management. Japan Med Assoc J 2008;51(5):343-346.
34. Mafauzy M. The problems and challenges of the aging population of Malaysia. Malaysian J Med Sci 2000;7:1-3.
35. Waiting times for registration and consultation at polyclinics. Singapore Ministry of Health. URL: https://www.moh.gov.sg/resources-statistics/healthcare-institution-statistics/waiting-times-for-registration-and-for-consultation-at-polyclinics [accessed 2020-06-01]
36. Hall LH, Johnson J, Watt I, Tsipa A, O'Connor DB. Healthcare staff wellbeing, burnout, and patient safety: a systematic review. PLoS One 2016;11(7):e0159015 [FREE Full text] [doi: 10.1371/journal.pone.0159015] [Medline: 27391946]
37. Polyclinic locations and operating hours. Polyclinics Singapore. URL: https://polyclinic.singhealth.com.sg/patient-care/our-polyclinics [accessed 2020-06-01]
38. Scott SD, Plotnikoff RC, Karunamuni N, Bize R, Rodgers W. Factors influencing the adoption of an innovation: an examination of the uptake of the Canadian Heart Health Kit (HHK). Implement Sci 2008 Oct 02;3(1):41 [FREE Full text] [doi: 10.1186/1748-5908-3-41] [Medline: 18831766]
39. Wang L, Miller MJ, Schmitt MR, Wen FK. Assessing readability formula differences with written health information materials: application, results, and recommendations. Res Social Adm Pharm 2013 Sep;9(5):503-516. [doi: 10.1016/j.sapharm.2012.05.009] [Medline: 22835706]
40. Kirsch IJ, Jungeblut A, Jenkins L, Kolstad A. Adult literacy in America: A first look at the results of the National Adult Literacy Survey. 1993. URL: https://nces.ed.gov/pubs93/93275.pdf [accessed 2022-06-12]
41. Choudhury D, Bhattacharjee D. Impact of socio economic factors on adoption of e-banking amongst salaried employees. Int J Res Manag Sci Technol 2015;3(3):37-46.
42. Alquran A, Lambert KA, Farouque A, Holland A, Davies J, Lampugnani ER, et al. Smartphone applications for encouraging asthma self-management in adolescents: a systematic review. Int J Environ Res Public Health 2018 Dec 29;15(11):1 [FREE Full text] [doi: 10.3390/ijerph15112403] [Medline: 30380692]
43. Hui CY, Walton R, McKinstry B, Jackson T, Parker R, Pinnock H. The use of mobile applications to support self-management for people with asthma: a systematic review of controlled studies to identify features associated with clinical effectiveness and adherence. J Am Med Inform Assoc 2017 May 01;24(3):619-632. [doi: 10.1093/jamia/ocw143] [Medline: 27694279]

Abbreviations

IT: information technology
PILBOX: Prescription in Locker Box

Edited by T Leung; submitted 07.08.20; peer-reviewed by B Johnson, H De Loof; comments to author 18.11.20; revised version received 12.04.21; accepted 16.03.22; published 27.06.22.

Please cite as:
Lim Jit Fan C, Boon Kwang G, Chee Wing Ling V, Woh Peng T, Goh Qiuling B
Remodeling the Medication Collection Process With Prescription in Locker Box (PILBOX): Prospective Cross-sectional Study
J Med Internet Res 2022;24(6):e23266
URL: https://www.jmir.org/2022/6/e23266
doi:10.2196/23266
PMID:35759321

©Christina Lim Jit Fan, Goh Boon Kwang, Vivian Chee Wing Ling, Tang Woh Peng, Bandy Goh Qiuling. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.06.2022. 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

Patients' Experiences of Web-Based Access to Electronic Health Records in Finland: Cross-sectional Survey

Sari Kujala[1], PhD; Iiris Hörhammer[2], DSC; Akseli Väyrynen[1], MSc; Mari Holmroos[3], MSS; Mirva Nättiaho-Rönnholm[3], MSS; Maria Hägglund[4], PhD; Monika Alise Johansen[5], PhD

[1]Department of Computer Science, Aalto University, Espoo, Finland

[2]Department of Industrial Engineering and Management, Aalto University, Espoo, Finland

[3]Kela, The Social Insurance Institution of Finland, Helsinki, Finland

[4]Healthcare Sciences and e-Health, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden

[5]Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway

Corresponding Author:
Sari Kujala, PhD
Department of Computer Science
Aalto University
Konemiehentie 2
Espoo, FI-00076
Finland
Phone: 358 503862768
Email: sari.kujala@aalto.fi

Abstract

Background: Patient portals that provide access to electronic health records offer a means for patients to better understand and self-manage their health. Yet, patient access to electronic health records raises many concerns among physicians, and little is known about the use practices and experiences of patients who access their electronic health records via a mature patient portal that has been available for citizens for over five years.

Objective: We aimed to identify patients' experiences using a national patient portal to access their electronic health records. In particular, we focused on understanding usability-related perceptions and the benefits and challenges of reading clinical notes written by health care professionals.

Methods: Data were collected from 3135 patient users of the Finnish My Kanta patient portal through a web-based survey in June 2021 (response rate: 0.7%). Patients received an invitation to complete the questionnaire when they logged out of the patient portal. Respondents were asked to rate the usability of the patient portal, and the ratings were used to calculate approximations of the System Usability Scale score. Patients were also asked about the usefulness of features, and whether they had discussed the notes with health professionals. Open-ended questions were used to ask patients about their experiences of the benefits and challenges related to reading health professionals' notes.

Results: Overall, patient evaluations of My Kanta were positive, and its usability was rated as good (System Usability Scale score approximation: mean 72.7, SD 15.9). Patients found the portal to be the most useful for managing prescriptions and viewing the results of examinations and medical notes. Viewing notes was the most frequent reason (978/3135, 31.2%) for visiting the portal. Benefits of reading the notes mentioned by patients included remembering and understanding what was said by health professionals and the instructions given during an appointment, the convenience of receiving information about health and care, the capability to check the accuracy of notes, and using the information to support self-management. However, there were challenges related to difficulty in understanding medical terminology, incorrect or inadequate notes, missing notes, and usability.

Conclusions: Patients actively used medical notes to receive information to follow professionals' instructions to take care of their health, and patient access to electronic health records can support self-management. However, for the benefits to be realized, improvements in the quality and availability of medical professionals' notes are necessary. Providing a standard information structure could help patients find the information they need. Furthermore, linking notes to vocabularies and other information sources could also improve the understandability of medical terminology; patient agency could be supported by allowing them to add comments to their notes, and patient trust of the system could be improved by allowing them to control the visibility of the professionals' notes.

XSL•FO
RenderX

(J Med Internet Res 2022;24(6):e37438) doi:10.2196/37438

KEYWORDS

patient portals; EHR; electronic health record; open notes; patient access; self-management; national survey

Introduction

Patient portals that provide access to electronic health records (EHRs) are becoming increasingly common. Such access to EHRs offers the means for patients to better understand personal health issues, treatment plans, and decisions [1], thus supporting personal health management [2] and informing patients between time- and resource-consuming clinic visits or phone appointments [3].

"Open notes," which are clinical notes that are shared with patients [4], can be considered an essential part of any patient-accessible EHR. In some countries, for example, Sweden [5], Norway [6], and Finland [7], nationwide patient-accessible EHR services, including open notes, are offered to most citizens through national patient portals.

Moreover, the majority of studies in recent reviews [3,8,9] highlighted benefits of patient access to EHRs. Patients were satisfied with the communication and engagement with clinicians, as well as better self-care, achieved as a result of patient access [8]. Improved doctor–patient relationships and patient outcomes were also found to be benefits [3].

Despite these benefits, health care professionals often criticize patient access to EHR [10]; patients, on the other hand, would like more doctors to offer access to their notes [11]. Patient access to EHR changes the physician–patient relationship and power dynamic; physicians have raised concerns [10,12] that such access may worry patients, cause misunderstandings, or cause extra work for physicians [13,14]. Physicians have also been worried that patients who find mistakes or errors would call and ask for corrections to notes which would increase the workload for health care [10].

Many studies [15-18] have also reported lower than anticipated levels of patient uptake of EHR access. Thus, in order to realize the potential of such access to support patient self-management, a better understanding of patient practices, motivations, and challenges is necessary. As de Lusignan et al [15] pointed out, there is still a need to understand how web-based access to EHR might be "redesigned to guide and teach patients in a way that promotes self-management and ultimately improves health."

Patient experiences with access to EHRs have often been explored using surveys, whereby patients were asked to rate usability [19] and attitude [5], usefulness [6], ease of use [20,21], and benefits and risks [22]. In addition, Bell et al [23] used a Likert-scale to study how reading notes affected patient–doctor relationships. Qualitative data have also been collected to understand patient views of access to EHRs. Mishra et al [24] included open-ended questions to identify positive and negative themes related to the usefulness, understandability, and worries caused by patient access; Gerard et al [25] used open-ended questions about the value of reading notes and providing feedback on open notes; Rexhepi et al [26] interviewed patients with cancer and found that patient access helped them prepare for doctor visits and understand their medical issues; and Eriksson-Backa et al [27] conducted focus groups with older adults and identified the uses, enablers, barriers, and behavioral outcomes of the national My Kanta patient portal.

In Finland, My Kanta, a nationwide patient portal, was introduced in 2010 and varied functions were adopted in a step-by-step manner [28]. Since 2015, the My Kanta patient portal has enabled all citizens using public health care services to access their health records and prescriptions, and to renew the latter [28]. The use of My Kanta is very established, with 63% of Finnish adults having accessed the patient portal during the period from 2010 to 2018 [7], and 92% of adults (from 18 to 65 years) used the patient portal in 2021. The most used functions among pharmacy customers were browsing prescription information (97.4%) and health records (96.3%) [20].

The goal of this study was to understand patients' experiences using My Kanta to access their EHRs. While My Kanta has been available for all patients to use for 7 years, little is known about patient use practices and experiences. Thus, we specifically focused on understanding patients' perceptions related to the usability of the patient portal and the benefits and challenges of reading the clinical notes written by health care professionals.

Methods

Design

We conducted a cross-sectional survey to capture patients' experiences using the My Kanta patient portal.

The My Kanta Patient Portal

My Kanta is a web-based patient portal for all residents with a Finnish personal identity number and access to electronic identification. Patients can view their own or their dependents' health data (consisting of records of health care visits, diagnoses, critical risk factors, laboratory tests, x-ray examinations, referrals, health and care plans, and medical certificates, statements [20], and e-prescriptions), request a prescription renewal, and save living wills and organ donation testaments [29].

My Kanta is a part of national Kanta services that integrate and save medical, health, and prescription data for health care providers, citizens, and pharmacies [28]. All public and private health care providers that use electronic patient record systems are obliged by law to send prescription and health data to Kanta services [7]. Health data, test results, and prescriptions can be used by health care units with patient consent [28], which can be given or withdrawn on My Kanta.

According to international benchmarking, My Kanta provided patients and their caregivers with the best access to their health

XSL•FO
RenderX

record data alongside Korea in 2019 [30] and also provided the most functions in 2016 [31]. However, My Kanta does not allow typical patient portal functions, such as appointment booking or communication with health care professionals.

Questionnaire

The web-based questionnaire included 4 open-ended questions and 11 questions with Likert scale or multiple choice response options (Multimedia Appendix 1). The topics of the questions were (1) reasons for logging into the patient portal and whether the visit was successful or not and why; (2) subjective usability of the patient portal; (3) usefulness of the features of the patient portal; (4) the benefits and challenges of reading health care professionals' notes and discussing their notes with them; (5) improvement ideas for the patient portal; (6) guidance on reading the notes; and (7) background information.

To assess perceived usability, a 2-item questionnaire based on the Usability Metric for User Experience [32]—the UMUX-LITE scale [33]—was used. UMUX-LITE scores were transformed, using a corrective regression formula [33], to System Usability Scale scores. The System Usability Scale is the most frequently used questionnaire for measuring the subjective usability of eHealth apps [34]. Borsci et al [35] tested UMUX-LITE with health care professionals and found it to be appropriate for use in the context of health care technology [35].

Open-ended questions about respondents' experiences of the benefits and challenges of reading health care professionals' notes were used in order to collect qualitative data about the most relevant issues from the patients' perspectives. The web-based questionnaire was dynamic; only respondents who reported having read the notes at least once (ie, had actual use experience) were asked the open-ended follow-up questions. If a respondent rated reading the notes as "not useful," they were only asked about challenges (to avoid unnecessarily asking these respondents questions about benefits). The survey was available in both official languages of Finland: Finnish and Swedish.

The questionnaire was reviewed by 2 researchers in the field and 2 experts from the Social Insurance Institution of Finland, which was the organization responsible for developing My Kanta. In addition, we pilot-tested the questionnaire with 3 patients who filled in the questionnaire and simultaneously talked aloud about how they understood the questions. The questionnaire was subsequently revised to clarify wording.

Conducting the Survey

Data were gathered during the period from June 4, 2021 to June 14, 2021 using a web-based questionnaire. Patient users of My Kanta in Finland received an invitation and a link to the questionnaire when they logged out of the patient portal. Thus, all respondents had used the patient portal just before they responded to the questionnaire. Participation was voluntary and anonymous.

Ethics Approval

The study protocol was reviewed and approved by the Ethical Review Board of Aalto University (ethics approval number D/957/03.04/2020 Nordic eHealth for Patients).

Analysis

Descriptive statistics were calculated for quantitative data (respondents' characteristics: age, gender, and portal usage). We performed content analysis (Atlas.ti, version 8.4.5; ATLAS.ti Scientific Software Development GmbH) on the responses to open-ended questions. One researcher first read through the data and used open coding to identify themes in the data without predefined categories. Short sentences were chosen as the analytical unit; themes were defined using in vivo coding, and to ensure that the themes represented the original meaning of the respondents, we used respondents' sentences to label the themes. The number of respondents who mentioned a theme was calculated, and the themes were categorized. A second researcher then reviewed the results. The researchers discussed similarities and differences in themes and combined categories, until a version was agreed upon as the final version.

Results

Respondents

Of 449,922 users who logged in, 3139 users responded to the survey (response rate 0.7%). Most users reported either weekly (889/3112, 28.6%) or monthly use (1120/3112, 36.0%) (Table 1). The frequency of use was comparable to that of My Kanta in May 2019, when users used My Kanta on an average of 2.4 times per month [7]. The proportion of users over the age of 50 years was high (2681/3135, 85.5%). The proportion was 2-fold that in 2021 (44%). Although the frequency of use may vary notably between users, this may suggest overrepresentation of older age groups among respondents.

XSL•FO
RenderX

Table 1. Respondent characteristics (n=3135).

Characteristic	Respondents, n (%)
Gender (n=3118)	
Female	2104 (67.5)
Male	962 (30.8)
Other	52 (1.7)
Age (years) (n=3115)	
<18	5 (0.2)
18-35	93 (3.0)
36-50	336 (10.8)
51-65	1082 (34.7)
66-75	1173 (37.6)
76-85	395 (12.7)
>85	31 (1.0)
Frequency of use (n=3112)	
Daily	194 (6.2)
Weekly	889 (28.6)
Monthly	1120 (36.0)
Less than once per month	878 (28.2)
First time user	31 (1.0)
Success of the visit (n=3125)	
Yes	2247 (71.9)
No	766 (24.5)
Do not know	112 (3.6)
Device used (n=3053)	
Computer	1836 (60.1)
Smartphone	690 (22.6)
Tablet	522 (17.1)
Something else	5 (0.2)
Has discussed the notes with a health care professional (n=3039)	
Yes	1046 (34.4)
No	1993 (65.6)

Experiences With the Patient Portal

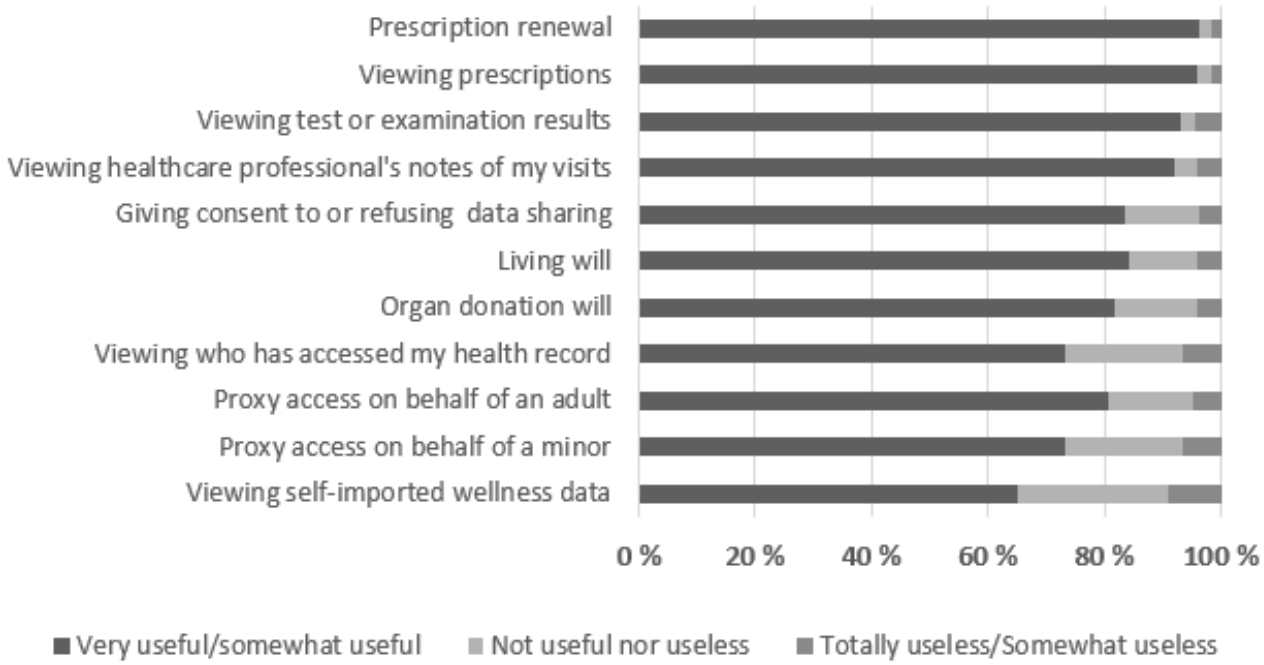
The total mean score for the System Usability Scale approximation was 72.7 (SD 15.9).

The most common reasons for visiting the My Kanta patient portal were viewing medical notes (978/3135, 31.2%), results of examinations (693/3135, 22.1%) or prescriptions (548/3135, 17.5%). Many people also visited the patient portal to renew a prescription (477/3135,15.2%), because there is no other method for renewing prescriptions electronically. At the time of the survey, COVID-19 vaccinations had started in Finland, and many (229/3135, 7.3%) logged into the patient portal to view their vaccination certificates. Other functions were used by only a few respondents (n=6-21). Some users (n=24) tried to use functions that did not exist, such as making appointments or checking their appointments (n=18), contact health care professionals (n=4), or looking for information about the reason that their prescription had not been renewed (n=2).

The most used functions were also deemed to be the most useful (Figure 1); for example, 96.4% (2511/2605) of users considered prescription renewal and 91.9% (2749/2992) of users considered viewing health care professionals' notes to be very useful or somewhat useful; however, the majority of users also considered less rarely used functions useful, with the lowest percentage (759/1165, 65.2%) of users considering self-reported wellness data, and the highest percentage (1483/1759, 84.3%) of users considering living will to be very useful or somewhat useful.

XSL•FO
RenderX

Figure 1. Usefulness of My Kanta patient portal features.



Benefits and Challenges of Reading Notes

Most respondents (2183/3135, 69.6%) answered the open-ended question and mentioned one or more benefits of reading notes (Table 2). Of the perceived benefits, most often respondents stated (560/2178, 25.7%) that notes supported remembering things:

> *One can recall afterwards what happened in the health care visit and what was discussed about.*

Respondents often mentioned that they felt tense or overwhelmed during their appointment, and notes helped in remembering what was said and which instructions were received.

XSL•FO
RenderX

Table 2. Themes of perceived benefits of viewing medical notes.

Benefits	Mentions (n=2178), n (%)
Supports remembering	560 (25.7)
What a doctor or a nurse said	505 (23.2)
The care history	55 (2.5)
Provides information	495 (22.7)
About health and care	223 (15.0)
To check the state of health and remain up-to-date	74 (3.4)
On how I and my disease are perceived	70 (3.2)
On all information concerning myself	43 (2.0)
That is more detailed and was not said during the appointment	34 (1.6)
About what was done during an appointment	30 (1.4)
On diagnoses	21 (1.0)
Convenience of patient portal access	449 (20.6)
Ability to return to view all the saved information	155 (7.1)
Can be checked at leisure	98 (4.5)
No need to call or contact health care	73 (3.4)
Easy of finding information	57 (2.6)
Fast access	56 (2.6)
Clear and reliable information	10 (0.5)
Helps in understanding	339 (15.6)
Own condition or what was said	326 (15.0)
Whether more can be asked if something was unclear	13 (0.6)
Ability to check the notes	234 (10.7)
Identifying potential errors and misunderstandings	142 (6.5)
Asking for error corrections	65 (3.0)
Checking that all essential information was written	21 (1.0)
Increases transparency and reliability	6 (0.3)
Supports self-management	175 (8.0)
Checking the care plan and next steps	63 (2.9)
Following the course of care success	36 (1.7)
Preparing for the next appointment	29 (1.3)
Looking for further information	14 (0.6)
Helps in communicating with health care professionals, learning to express yourself	14 (0.6)
Supports self-care	13 (0.6)
Enables peace of mind	6 (0.3)

Respondents appreciated that notes provided information about their health and care. They were able to follow the course of their care and remain up-to-date. Furthermore, they wanted to identify doctors' perceptions of them and their diseases. Several mentioned that it is important to have all the information concerning themselves:

> *My life and my own information are certainly of primary importance.*

Respondents also noted that the information is provided conveniently in one place, and they can check the information whenever they want. Notes were also perceived as helping them to understand their health conditions and what health care professionals had said during appointments. In addition, many respondents wanted to check the notes to ensure there were no errors or misunderstandings.

Many stated that the reason for accessing the information and remaining up-to-date was to actively self-manage their health.

XSL•FO
RenderX

Respondents wanted to be aware of their care plans and to follow the course of their success. They subsequently prepared themselves for the next appointment and looked for further information related to their condition and care. A few commented that the notes helped in communicating with health care professionals and supported learning to express themselves, and 1046 out of 3135 (33.4%) respondents also discussed the notes with health care professionals.

One-third (1175/3135, 37.5%) of respondents also reported one or more challenges in reading notes (Table 3). The most commonly mentioned challenge was the difficulty in understanding the notes and the medical terminology. For example, one respondent stated:

> *Language that I don't understand. Wikipedia may help in translation work, when you don't understand the crucial words.*

Many mentioned that they used Google to interpret the unfamiliar terms, codes, and abbreviations, and they wanted plain language to be used instead.

Table 3. Perceived challenges of viewing medical notes.

Challenges	Mentions (n=1175), n (%)
Notes are difficult to understand	707 (60.2)
The medical terminology is difficult to understand	523 (44.5)
Abbreviations are difficult	73 (6.2)
Examination and test results are difficult	44 (3.7)
Notes in general are difficult to understand	44 (3.7)
Diagnoses are not understandable	23 (2.0)
Notes are not available	232 (19.7)
Delay in access	121 (10.3)
Missing information	105 (8.9)
Children's information is not visible	6 (0.5)
Notes are incorrect or inadequate	217 (18.5)
Incorrect information or errors	80 (6.8)
Health care professionals' misinterpretations	28 (2.4)
Imprecise notes	27 (2.3)
Very brief notes	16 (1.4)
Negligent writing	15 (1.3)
Irrelevant or too detailed information	12 (1.0)
Repetition	10 (0.9)
Poor language	9 (0.8)
Wrong language (eg, Finnish instead of Swedish)	7 (0.6)
Too personal	5 (0.4)
Inappropriate	4 (0.3)
Follow-up is unclear	4 (0.3)
Problems with usability	167 (17.4)
Information was difficult to find	85 (7.2)
Errors are difficult or impossible to correct	37 (3.1)
Could be easier to use	25 (2.1)
Disorganized	25 (2.1)
No interactivity	8 (0.7)
The search process is cumbersome	5 (0.4)
Worries about privacy	5 (0.4)
Comparing examination results is difficult	5 (0.4)
Reading on mobile devices is difficult	5 (0.4)
The text is small	4 (0.3)

XSL•FO
RenderX

However, the notes were not always available because there were delays in access and some visits were not recorded or visible. It was mentioned that it could take days or weeks before the notes were available, and some information was not available at all.

Many respondents perceived notes to be incorrect or inadequate. Most commonly, they were seen as having errors—some were not significant, such as a wrong date, but some were more severe, such as having a wrong diagnosis or another patient's information. Respondents described,

> *Mainly the challenge is that the communication has been wrongly recorded or it is misunderstood. People should have possibility to say their views on My Kanta*

and

> *Sometimes there have been erroneous information and diagnoses. For example, a cancer that I don't have.*

Many also reported that the notes differed from what they had experienced themselves. Several also wished for more detailed notes. In contrast, some felt that it was unnecessary to include all personal details that they had mentioned during an appointment or the whole message that they had sent. One person also mentioned that they did not want to talk about certain issues, because they would be recorded and seen by all professionals.

Finally, there were challenges related to the usability of the system. Most commonly, it was mentioned that it was difficult to find information. The information was not always in chronological order, and some examination results were not linked to the appropriate appointments. A few respondents also mentioned that there is no interactivity in the system, and they wanted to comment on the notes or request corrections. Furthermore, it was noted that a patient should receive a notification when new information is available.

Discussion

Principal Results

Respondents evaluated the My Kanta patient portal as useful and usable, which is consistent with the findings of earlier studies [20,21]. The total mean score for the System Usability Scale approximation was 72.7 (SD 15.9), which can verbally be described as good usability, according to Bangor et al [36,37]. Prescription renewal and viewing were indicated to be the most useful functions, but viewing medical notes and the results of examinations were the most frequent reasons for visiting the patient portals, which 91.9% (2749/3135) and 92.9 % (2770/3135) of respondents, respectively, considered useful.

Furthermore, respondents explained in their responses to open-ended questions that they appreciated having access to EHRs and information via a patient portal, which supports earlier findings [6,22]. Because My Kanta has been used nationally for several years, respondents were already familiar with the portal and actively used medical notes to prepare for their communications with health care professionals and to take care of their health.

The qualitative responses provided a rich and versatile description of the benefits of patient access to EHRs. Specifically, reading the notes was described as convenient, because they could be accessed easily and quickly, whenever suitable and at leisure. Therefore, easy access via patient portals may help patients to be engaged in self-management of their health. Reading notes were described as supporting remembering and understanding what health professionals said. They were able to check the state of health and care plans, remain up-to-date, look for further information, prepare for the next appointment, and ask further questions if something was unclear. We suggest that these activities support patients in learning about their disease or care, which motivates them to take care of their health.

Furthermore, reading notes can provide information that is not directly addressed during visits with a health care professional. As previously suggested [38], this may improve patient autonomy by reducing dependence on individual health care professionals and providing the opportunity to consult medical literature or other health care professionals to better understand health status and options for care or treatments.

Many respondents stated that it was important to be able to check the notes to identify potential errors and misunderstandings. They were also interested in professionals' perceptions of their situations. Reading the notes was thus seen to help them understand what health care professionals had said and prepare for the next appointment. Thus, patient access to EHR supports patient–provider communication.

Very few patients were concerned about privacy or felt the notes were too personal or inappropriate. Some patients found incorrect information, and a few mentioned serious errors. It was very rarely mentioned, but a few respondents also felt that the notes included irrelevant information or personal information that was too detailed. Although rarely mentioned, the notes sometimes included information about very personal issues that patients were unwilling to share with all health care personnel. In particular, when a patient portal does not allow patients to correct errors or express their views with a comment, we presume that some patients may feel that their self-determination is violated.

It is noteworthy that respondents did not perceive reading the notes to be harmful per se but that challenges, such as understandability of medical terminology, incorrect or inadequate notes, missing information, or difficulties in finding information, interfered with the benefits of reading the notes. Finnish law requires that professionals' notes are sufficiently comprehensive, clear, and understandable and that only commonly known terms or abbreviations are used [39]. Nevertheless, this is clearly not fulfilled according to the survey results.

In order to realize associated benefits, improvements in the quality and availability of medical professionals' notes are needed. In addition to educating health care professionals, the availability of information can also be supported by providing a standard information structure. Because the information structure was confusing to patients, a standard structure would make finding and reading information easier from patients'

XSL•FO
RenderX

perspectives. It is important that the order of the notes is logical from their point of view and that examination results are clearly linked to corresponding appointments. Linking the notes to vocabularies and other information sources could also improve the understandability of medical terminology without increasing the workload of professionals. In addition, patient agency and trust could be supported by enabling them to add comments to their notes, mark some entries as sensitive, and control the visibility of entries.

Limitations

This was a cross-sectional survey study examining patients' self-reported experiences of the national patient portal in Finland, and the results may not be generalizable to other countries or patient portals. The survey was available only to My Kanta users after logging out of the patient portal. Not all users may have actively logged out the portal or noticed the invitation, which may have contributed to the low response rate. Thus, the results do not represent all My Kanta users or the population of Finland. A similar survey study in Sweden [5] also had a low response rate (0.61%).

In addition, the only demographic information available from the survey was age and gender; health and socioeconomic status of the respondents, literacy, and health literacy remained unknown. It is possible that the survey respondents represented users who were most interested in the patient portal and most capable of using it. Our sample did not include persons who had stopped using the portal or were not able to use it. Thus, nonrespondents may differ in their use of the patient portal and may experience barriers (eg, [40,41]) that were not identified in this study.

Moreover, My Kanta includes many functions that were recently added and, thus, not widely used. Therefore, the usefulness of all the functions could not be reliably evaluated. Furthermore, the portal does not have all the potentially useful functions that users could have experienced. As a few respondents complained, My Kanta does not have much interactivity—patients are not allowed to comment on notes or request corrections in the portal. In addition, the lack of notifications on added content frustrated respondents, because they often logged in to look for notes or test results that were not available yet.

By asking open-ended questions on the benefits and challenges, we improved the reliability of the answers as respondents reported their experiences using their own words and were not guided by having to choose from certain options. Because the number of respondents was high, the data that we collected were rich and versatile. However, respondents may have focused on the most significant benefits and challenges they experienced, and they may not have been able to verbalize those that were more abstract and less obvious. Therefore, we believe that our mixed methods survey study complements previous quantitative studies [5,6,19-23].

Comparison With Prior Work

The main benefits experienced by patients were very similar to those identified in a smaller study [25] conducted at a single institution in the United States over a 12-month pilot period, in which participants reported that reading notes helped them to better remember next steps, provided positive emotions, and gave them faster access and results. The participants in the study [25] also valued the opportunity to correct any possible misunderstandings and give feedback to their providers, which are functions that patients also wished were available on My Kanta. In addition, Rexhepi et al [26] and Pyper et al [42] identified similar benefits, in studies in Sweden and the United Kingdom, respectively. Thus, our study provides further details in understanding self-management practices that patient access to EHR can support.

Moreover, several studies [6,20,24,26,27,42-44] have identified that some parts of medical records are difficult to understand. In addition, Johansen et al [14] found that 25.6% of administrative staff and 15.4% of health care professionals had received feedback from patients or their relatives regarding mistakes or missing information in their EHR. In our study, the number of serious mistakes was seldom mentioned, which was not the case in a recent US survey [45] with 29,656 respondents, in which 1 in 5 patients reported finding a mistake, and 40% perceived the mistake to be serious. It is possible that the number of respondents who found errors would have been higher in our study if we had specifically asked about this. This should be explored in future studies.

Pyper et al [42] also found that patients identified errors and omissions and had differences of opinion when they accessed electronic records for the first time. In this context, our study shows that unclear or inadequate notes are very common even when patients are familiar with use of the EHR and the challenges do not disappear when health care professionals gain experience in conveying information to patients. Thus, the findings support the need for applications that provide explanations of medical terms in EHR notes (eg, [46,47]).

Thus, patients' basic needs and self-management processes seem to be similar regardless of the context—benefits and challenges experienced by patients are remarkably similar across countries, different health care systems, and EHRs.

Conclusions

Our findings indicate that patient access to EHR can support self-management—patients actively used medical notes to understand and remember what health care professionals said and to take care of their health. The challenges interfered with the benefits of reading the notes. In order to realize benefits, improvements in the quality and availability of medical professionals' notes are needed, and patients should be encouraged to discuss their concerns with them. In addition, the availability of information can also be supported by using a standard information structure. Specifically, linking the notes to vocabularies and other information sources could also improve the understandability of medical terminology.

XSL•FO
RenderX

Acknowledgments

This work was supported by the Strategic Research Council at the Academy of Finland (grants 327145 and 327147) and NordForsk (project 100477).

Conflicts of Interest

None declared.

Multimedia Appendix 1
The questionnaire.
[DOCX File , 15 KB - jmir_v24i6e37438_app1.docx]

References

1. Mahnke M, Nielsen M. eHealth platforms as user–data communicationxamining patients' struggles with digital health data. Nordicom Review 2021;42:58. [doi: 10.2478/nor-2021-0040]
2. Shah SGS, Fitton R, Hannan A, Fisher B, Young T, Barnett J. Accessing personal medical records online: a means to what ends? Int J Med Inform 2015 Feb;84(2):111-118 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.10.005] [Medline: 25453275]
3. Tapuria A, Porat T, Kalra D, Dsouza G, Xiaohui S, Curcin V. Impact of patient access to their electronic health record: systematic review. Inform Health Soc Care 2021 Jun 02;46(2):192-204. [doi: 10.1080/17538157.2021.1879810] [Medline: 33840342]
4. Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N, et al. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. Ann Intern Med 2012 Oct 2;157(7):461-470 [FREE Full text] [doi: 10.7326/0003-4819-157-7-201210020-00002] [Medline: 23027317]
5. Moll J, Rexhepi H, Cajander Å, Grünloh C, Huvila I, Hägglund M, et al. Patients' experiences of accessing their electronic health records: national patient survey in Sweden. J Med Internet Res 2018 Nov 01;20(11):e278 [FREE Full text] [doi: 10.2196/jmir.9492] [Medline: 30389647]
6. Zanaboni P, Kummervold PE, Sørensen T, Johansen MA. Patient use and experience with online access to electronic health records in Norway: results from an online survey. J Med Internet Res 2020 Feb 07;22(2):e16144 [FREE Full text] [doi: 10.2196/16144] [Medline: 32031538]
7. Jormanainen V, Parhiala K, Niemi A, Erhola M, Keskimäki I, Kaila M. Half of the Finnish population accessed their own data: comprehensive access to personal health information online is a corner-stone of digital revolution in Finnish health and social care. Finish Journal of eHealth and eWelfare 2019 Nov 02;11(4):298-310. [doi: 10.23996/fjhw.83323]
8. Mold F, de LS, Sheikh A, Majeed A, Wyatt JC, Quinn T, et al. Patients' online access to their electronic health records and linked online services: a systematic review in primary care. Br J Gen Pract 2015 Mar;65(632):e141-e151 [FREE Full text] [doi: 10.3399/bjgp15X683941] [Medline: 25733435]
9. Benjamins J, Haveman-Nies A, Gunnink M, Goudkuil A, de Vet E. How the use of a patient-accessible health record contributes to patient-centered care: scoping review. J Med Internet Res 2021 Jan 11;23(1):e17655 [FREE Full text] [doi: 10.2196/17655] [Medline: 33427683]
10. Cajander Å, Grünloh C. Electronic health records are more than a work tool: conflicting needs of direct and indirect stakeholders. : AC; 2019 Presented at: CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland p. 1-13. [doi: 10.1145/3290605.3300865]
11. Esch T, Mejilla R, Anselmo M, Podtschaske B, Delbanco T, Walker J. Engaging patients through open notes: an evaluation using mixed methods. BMJ Open 2016 Jan 29;6(1):e010034 [FREE Full text] [doi: 10.1136/bmjopen-2015-010034] [Medline: 26826154]
12. Grünloh C, Myreteg G, Cajander Å, Rexhepi H. "Why do they need to check me?" patient participation through ehealth and the doctor-patient relationship: qualitative study. J Med Internet Res 2018 Jan 15;20(1):e11 [FREE Full text] [doi: 10.2196/jmir.8444] [Medline: 29335237]
13. Grünloh C, Cajander Å, Myreteg G. "The record is our work tool!"-physicians' framing of a patient portal in Sweden. J Med Internet Res 2016 Jun 27;18(6):e167 [FREE Full text] [doi: 10.2196/jmir.5705] [Medline: 27349531]
14. Johansen MA, Kummervold PE, Sørensen T, Zanaboni P. Health professionals' experience with patients accessing their electronic health records: results from an online survey. Stud Health Technol Inform 2019 Aug 21;264:504-508. [doi: 10.3233/SHTI190273] [Medline: 31437974]
15. de Lusignan S, Mold F, Sheikh A, Majeed A, Wyatt JC, Quinn T, et al. Patients' online access to their electronic health records and linked online services: a systematic interpretative review. BMJ Open 2014 Sep;4(9):e006021 [FREE Full text] [doi: 10.1136/bmjopen-2014-006021] [Medline: 25200561]
16. Niazkhani Z, Toni E, Cheshmekaboodi M, Georgiou A, Pirnejad H. Barriers to patient, provider, and caregiver adoption and use of electronic personal health records in chronic care: a systematic review. BMC Med Inform Decis Mak 2020 Jul 08;20(1):153 [FREE Full text] [doi: 10.1186/s12911-020-01159-1] [Medline: 32641128]

XSL•FO
RenderX

17. Zhao J, Song B, Anand E, Schwartz D, Panesar M, Jackson G, et al. Barriers, facilitators, and solutions to optimal patient portal and personal health record use: a systematic review of the literature. AMIA Annu Symp Proc 2017;2017:1913-1922 [FREE Full text] [Medline: 29854263]
18. Fraccaro P, Vigo M, Balatsoukas P, Buchan I, Peek N. Patient portal adoption rates: a systematic literature review and meta-analysis. 2017 Presented at: MEDINFO: Precision Healthcare through Informatics; August 21-25; Hangzhou, China p. 83. [doi: 10.3233/978-1-61499-830-3-79]
19. Hägglund M, Scandurra I. User evaluation of the Swedish patient accessible electronic health record: System Usability Scale. JMIR Hum Factors 2021 Jul 27;8(3):e24927 [FREE Full text] [doi: 10.2196/24927] [Medline: 34313596]
20. Sääskilahti M, Ahonen R, Timonen J. Pharmacy customers' experiences of use, usability, and satisfaction of a nationwide patient portal: survey study. J Med Internet Res 2021 Jul 16;23(7):e25368 [FREE Full text] [doi: 10.2196/25368] [Medline: 34269687]
21. Lämsä E, Timonen J, Mäntyselkä P, Ahonen R. Pharmacy customers' experiences with the national online service for viewing electronic prescriptions in Finland. Int J Med Inform 2017 Jan;97:221-228. [doi: 10.1016/j.ijmedinf.2016.10.014] [Medline: 27919380]
22. Walker J, Leveille S, Bell S, Chimowitz H, Dong Z, Elmore JG, et al. OpenNotes after 7 years: patient experiences with ongoing access to their clinicians' outpatient visit notes. J Med Internet Res 2019 May 06;21(5):e13876 [FREE Full text] [doi: 10.2196/13876] [Medline: 31066717]
23. Bell SK, Mejilla R, Anselmo M, Darer JD, Elmore JG, Leveille S, et al. When doctors share visit notes with patients: a study of patient and doctor perceptions of documentation errors, safety opportunities and the patient-doctor relationship. BMJ Qual Saf 2016 May 18:262-270. [doi: 10.1136/bmjqs-2015-004697] [Medline: 27193032]
24. Mishra VK, Hoyt RE, Wolver SE, Yoshihashi A, Banas C. Qualitative and quantitative analysis of patients' perceptions of the patient portal experience with OpenNotes. Appl Clin Inform 2019 Jan;10(1):10-18. [doi: 10.1055/s-0038-1676588] [Medline: 30602196]
25. Gerard M, Fossa A, Folcarelli PH, Walker J, Bell SK. What patients value about reading visit notes: a qualitative inquiry of patient experiences with their health information. J Med Internet Res 2017 Jul 14;19(7):e237 [FREE Full text] [doi: 10.2196/jmir.7212] [Medline: 28710055]
26. Rexhepi H, Åhlfeldt RM, Cajander Å, Huvila I. Cancer patients' attitudes and experiences of online access to their electronic medical records: A qualitative study. Health Informatics J 2018 Jun 19;24(2):115-124 [FREE Full text] [doi: 10.1177/1460458216658778] [Medline: 27440056]
27. Eriksson-Backa K, Hirvonen N, Enwald H, Huvila I. Enablers for and barriers to using My Kanta - a focus group study of older adults' perceptions of the National Electronic Health Record in Finland. Inform Health Soc Care 2021 Dec 02;46(4):399-411. [doi: 10.1080/17538157.2021.1902331] [Medline: 33787438]
28. Jormanainen V. Large-scale implementation and adoption of the Finnish national Kanta services in 2010–2017: a prospective, longitudinal, indicator-based study. FinJeHeW 2018 Dec 04;10(4):74511. [doi: 10.23996/fjhw.74511]
29. My Kanta pages. Kanta. URL: https://www.kanta.fi/en/web/guest/my-kanta-pages [accessed 2021-11-25]
30. Ammenwerth E, Duftschmid G, Al-Hamdan Z, Bawadi H, Cheung NT, Cho K, et al. International comparison of six basic ehealth indicators across 14 countries: an ehealth benchmarking study. Methods Inf Med 2020 Dec;59(S 02):e46-e63 [FREE Full text] [doi: 10.1055/s-0040-1715796] [Medline: 33207386]
31. Essén A, Scandurra I, Gerrits R, Humphrey G, Johansen MA, Kierkegaard P, et al. Patient access to electronic health records: differences across ten countries. Health Policy and Technology 2018 Mar;7(1):44-56. [doi: 10.1016/j.hlpt.2017.11.003]
32. Finstad K. The usability metric for user experience. Interact Comput 2010 Sep;22(5):323-327. [doi: 10.1016/j.intcom.2010.04.004]
33. Lewis J, Utesch B, Maher D. UMUX-LITE: when there's no time for the SUS. 2013 Presented at: CHI 2013 Conference; April 27–May 2, 2013; Paris, France p. 2099-2102. [doi: 10.1145/2470654.2481287]
34. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: a scoping review. Int J Med Inform 2019 Jun;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
35. Borsci S, Buckle P, Walne S. Is the LITE version of the usability metric for user experience (UMUX-LITE) a reliable tool to support rapid assessment of new healthcare technology? Appl Ergon 2020 Apr;84:103007. [doi: 10.1016/j.apergo.2019.103007] [Medline: 31785449]
36. Bangor A. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123. [doi: 10.5555/2835587.2835589]
37. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
38. Blease C, Delbanco T, Torous J, Ponten M, DesRoches C, Hagglund M, et al. Sharing clinical notes, and placebo and nocebo effects: can documentation affect patient health? J Health Psychol 2022 Jan;27(1):135-146. [doi: 10.1177/1359105320948588] [Medline: 32772861]
39. Sosiaali- ja terveysministeriön asetus potilasasiakirjoista (kumottu). FinLex. URL: https://www.finlex.fi/fi/laki/ajantasa/2009/20090298 [accessed 2022-02-04]

XSL•FO
RenderX

40. Valeur HS, Lie AK, Moen K. Patient rationales against the use of patient-accessible electronic health records: qualitative study. J Med Internet Res 2021 May 28;23(5):e24090 [FREE Full text] [doi: 10.2196/24090] [Medline: 34047711]
41. Kainiemi E, Vehko T, Kyytsönen M, Hörhammer I, Kujala S, Jormanainen V, et al. The factors associated with nonuse of and dissatisfaction with the national patient portal in Finland in the era of COVID-19: population-based cross-sectional survey. JMIR Med Inform 2022 Apr 22;10(4):e37500 [FREE Full text] [doi: 10.2196/37500] [Medline: 35404831]
42. Pyper C, Amery J, Watson M, Crook C. Patients' experiences when accessing their on-line electronic patient records in primary care. Br J Gen Pract 2004 Jan;54(498):38-43 [FREE Full text] [Medline: 14965405]
43. van Kuppenveld SI, van Os-Medendorp H, Tiemessen NA, van Delden JJ. Real-time access to electronic health record via a patient portal: is it harmful? a retrospective observational study. J Med Internet Res 2020 Feb 06;22(2):e13622 [FREE Full text] [doi: 10.2196/13622] [Medline: 32044753]
44. Lee EH, Patel JP, Fortin AH. Patient-centric medical notes: identifying areas for improvement in the age of open medical records. Patient Educ Couns 2017 Aug;100(8):1608-1611. [doi: 10.1016/j.pec.2017.02.018] [Medline: 28242141]
45. Bell SK, Delbanco T, Elmore JG, Fitzgerald PS, Fossa A, Harcourt K, et al. Frequency and types of patient-reported errors in electronic health record ambulatory care notes. JAMA Netw Open 2020 Jun 01;3(6):e205867 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.5867] [Medline: 32515797]
46. Chen J, Druhl E, Polepalli RB, Houston TK, Brandt CA, Zulman DM, et al. A natural language processing system that links medical terms in electronic health record notes to lay definitions: system development using physician reviews. J Med Internet Res 2018 Jan 22;20(1):e26 [FREE Full text] [doi: 10.2196/jmir.8669] [Medline: 29358159]
47. Lalor JP, Woolf B, Yu H. Improving electronic health record note comprehension with NoteAid: randomized trial of electronic health record note comprehension interventions with crowdsourced workers. J Med Internet Res 2019 Jan 16;21(1):e10793 [FREE Full text] [doi: 10.2196/10793] [Medline: 30664453]

Abbreviations

EHR: electronic health record

Edited by G Eysenbach; submitted 21.02.22; peer-reviewed by C Grünloh, L Petersson; comments to author 23.03.22; revised version received 11.04.22; accepted 06.05.22; published 06.06.22.

Please cite as:
Kujala S, Hörhammer I, Väyrynen A, Holmroos M, Nättiaho-Rönnholm M, Hägglund M, Johansen MA
Patients' Experiences of Web-Based Access to Electronic Health Records in Finland: Cross-sectional Survey
J Med Internet Res 2022;24(6):e37438
URL: https://www.jmir.org/2022/6/e37438
doi:10.2196/37438
PMID:35666563

©Sari Kujala, Iiris Hörhammer, Akseli Väyrynen, Mari Holmroos, Mirva Nättiaho-Rönnholm, Maria Hägglund, Monika Alise Johansen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.06.2022. 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

Electronic Health Record–Based Recruitment and Retention and Mobile Health App Usage: Multisite Cohort Study

Janelle W Coughlin[1,2], PhD; Lindsay M Martin[3], MSc; Di Zhao[2,4], PhD; Attia Goheer[4], PhD; Thomas B Woolf[5], PhD; Katherine Holzhauer[3], BA; Harold P Lehmann[3], MD, PhD; Michelle R Lent[6], PhD; Kathleen M McTigue[7], MPH, MD; Jeanne M Clark[2,3], MD; Wendy L Bennett[2,3], MPH, MD

[1]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, United States

[2]Welch Center for Prevention, Epidemiology and Clinical Research, Johns Hopkins University, Baltimore, MD, United States

[3]Division of General Internal Medicine, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, United States

[4]Department of Health Policy and Management, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States

[5]Department of Physiology, Johns Hopkins University School of Medicine, Baltimore, MD, United States

[6]School of Professional and Applied Psychology, Philadelphia College of Osteopathic Medicine, Philadelphia, PA, United States

[7]Division of General Internal Medicine, University of Pittsburgh, Pittsburgh, PA, United States

Corresponding Author:
Janelle W Coughlin, PhD
Department of Psychiatry and Behavioral Sciences
Johns Hopkins University School of Medicine
5510 Nathan Shock Drive Suite 100
Baltimore, MD, 21224
United States
Phone: 1 410 550 7988
Email: jwilder3@jhmi.edu

Abstract

Background: To address the obesity epidemic, there is a need for novel paradigms, including those that address the timing of eating and sleep in relation to circadian rhythms. Electronic health records (EHRs) are an efficient way to identify potentially eligible participants for health research studies. Mobile health (mHealth) apps offer available and convenient data collection of health behaviors, such as timing of eating and sleep.

Objective: The aim of this descriptive analysis was to report on recruitment, retention, and app use from a 6-month cohort study using a mobile app called Daily24.

Methods: Using an EHR query, adult patients from three health care systems in the PaTH clinical research network were identified as potentially eligible, invited electronically to participate, and instructed to download and use the Daily24 mobile app, which focuses on eating and sleep timing. Online surveys were completed at baseline and 4 months. We described app use and identified predictors of app use, defined as 1 or more days of use, versus nonuse and usage categories (ie, immediate, consistent, and sustained) using multivariate regression analyses.

Results: Of 70,661 patients who were sent research invitations, 1021 (1.44%) completed electronic consent forms and online baseline surveys; 4 withdrew, leaving a total of 1017 participants in the analytic sample. A total of 53.79% (n=547) of the participants were app users and, of those, 75.3% (n=412), 50.1% (n=274), and 25.4% (n=139) were immediate, consistent, and sustained users, respectively. Median app use was 28 (IQR 7-75) days over 6 months. Younger age, White race, higher educational level, higher income, having no children younger than 18 years, and having used 1 to 5 health apps significantly predicted app use (vs nonuse) in adjusted models. Older age and lower BMI predicted early, consistent, and sustained use. About half (532/1017, 52.31%) of the participants completed the 4-month online surveys. A total of 33.5% (183/547), 29.3% (157/536), and 27.1% (143/527) of app users were still using the app for at least 2 days per month during months 4, 5, and 6 of the study, respectively.

Conclusions: EHR recruitment offers an efficient (ie, high reach, low touch, and minimal participant burden) approach to recruiting participants from health care settings into mHealth research. Efforts to recruit and retain less engaged subgroups are needed to collect more generalizable data. Additionally, future app iterations should include more evidence-based features to increase participant use.

XSL•FO
RenderX

(J Med Internet Res 2022;24(6):e34191) doi:10.2196/34191

KEYWORDS

mHealth; mobile apps; recruitment; engagement; retention; timing of eating; timing of sleep; obesity; EHR

Introduction

Obesity and its related medical comorbidities are highly prevalent public health conditions [1-5]. The strongest evidence for preventing and treating obesity targets health behaviors to modify dietary composition, reduce calories, and increase physical activity [6-8]. Although reducing calories and increasing physical activity result in short-term weight loss, there is a need to identify lifelong behavioral patterns that promote longer-term weight loss and maintenance of healthy weight [9-11]. Aligning the timing of eating and sleeping with intrinsic circadian rhythm (ie, a shorter duration of eating, often called time-restricted eating or feeding) has not yet been thoroughly examined in population-based studies, but has the potential to provide a new paradigm to prevent obesity and metabolic conditions [12-15].

Mobile devices and mobile health (mHealth) apps are ubiquitous, readily available approaches to collect real-time data on health behaviors, such as dietary intake, physical activity, and sleep [16-18]. mHealth apps are often designed and marketed to provide behavioral tracking and lifestyle modification support [19-22]. They also provide a convenient and efficient method for collecting information to advance knowledge about the relationship between obesity-related behavioral patterns and health outcomes [23-25].

Although mHealth research has grown exponentially in the last few decades, study attrition is a major problem, and there is a need to identify successful, low-burden, and efficient recruitment and retention strategies [26,27]. The era of COVID-19, in particular, has additionally highlighted the importance of remote research procedures. Electronic health record (EHR)–based recruitment strategies provide potentially efficient (ie, low touch and low participant burden) methods for identifying and recruiting high volumes of patients meeting predetermined medical criteria for population-based research studies [28-30].

This study presents a secondary analysis from a 6-month, multisite, cohort study that used the EHR to identify and recruit participants to use a mobile app (Daily24), designed to assess timing of eating and sleep [31]. The main goal of the parent observational study was to evaluate the longitudinal association between timing of eating and weight changes over time. Because of the growing interest in both EHR-based recruitment strategies and mHealth data collection methods [19,32-34], the goal of this descriptive analysis is to do the following:

1. Describe the EHR-based recruitment and electronic consent (e-consent) methods and response rates for enrolling in the study and downloading the mobile app.
2. Describe engagement strategies, app use, and retention rates during the 6-month study.
3. Evaluate demographic and behavioral predictors of Daily24 app use.

We hypothesized that people who are younger, have greater education, and have higher BMIs would be more likely to use the app. This study has the potential to inform the field of behavioral health in methodology, uptake, and engagement of mHealth approaches for observational research.

Methods

Recruitment

We recruited a cohort of adult patients from three health care systems in the PaTH Clinical Research Network, part of PCORnet (National Patient-Centered Research Network). The three health care systems included the Johns Hopkins Health System, the Geisinger Health System, and the University of Pittsburgh Medical Center [35-37].

Ethics Approval

Institutional Review Board (IRB) approval was obtained from the Johns Hopkins School of Medicine (IRB00174516), which had a reliance agreement with the other institutions' IRBs.

EHR-Based Participant Eligibility Criteria

Potential participants were identified using EHR-based eligibility criteria (ie, "computable phenotype" [38]) to query the EHR. Each site also obtained a list of potentially eligible participants who previously consented to complete PaTH cohort studies at these sites [37]. Eligibility criteria included the following: at least 18 years of age and a minimum of one weight measurement and one height measurement recorded in the EHR between July 2017 and July 2019. Participants were excluded if they were deceased.

Recruitment Messaging Via Email and the Patient Portal

Potentially eligible participants were sent recruitment messages via email or the patient portal (ie, Epic MyChart) from February to July 2019. Each partnering health care system tailored its own strategy to recruit participants from the large pool of potentially eligible patients who were identified using the computable phenotype. One site used patient portal recruitment almost exclusively, focusing on patients who had a health system visit in the last week. The other two sites sent email recruitment letters through their primary care and weight management practices, with messages signed by the clinic medical directors. Multimedia Appendix 1 shows a sample recruitment message, which included a brief study description and link to a web-based e-consent form.

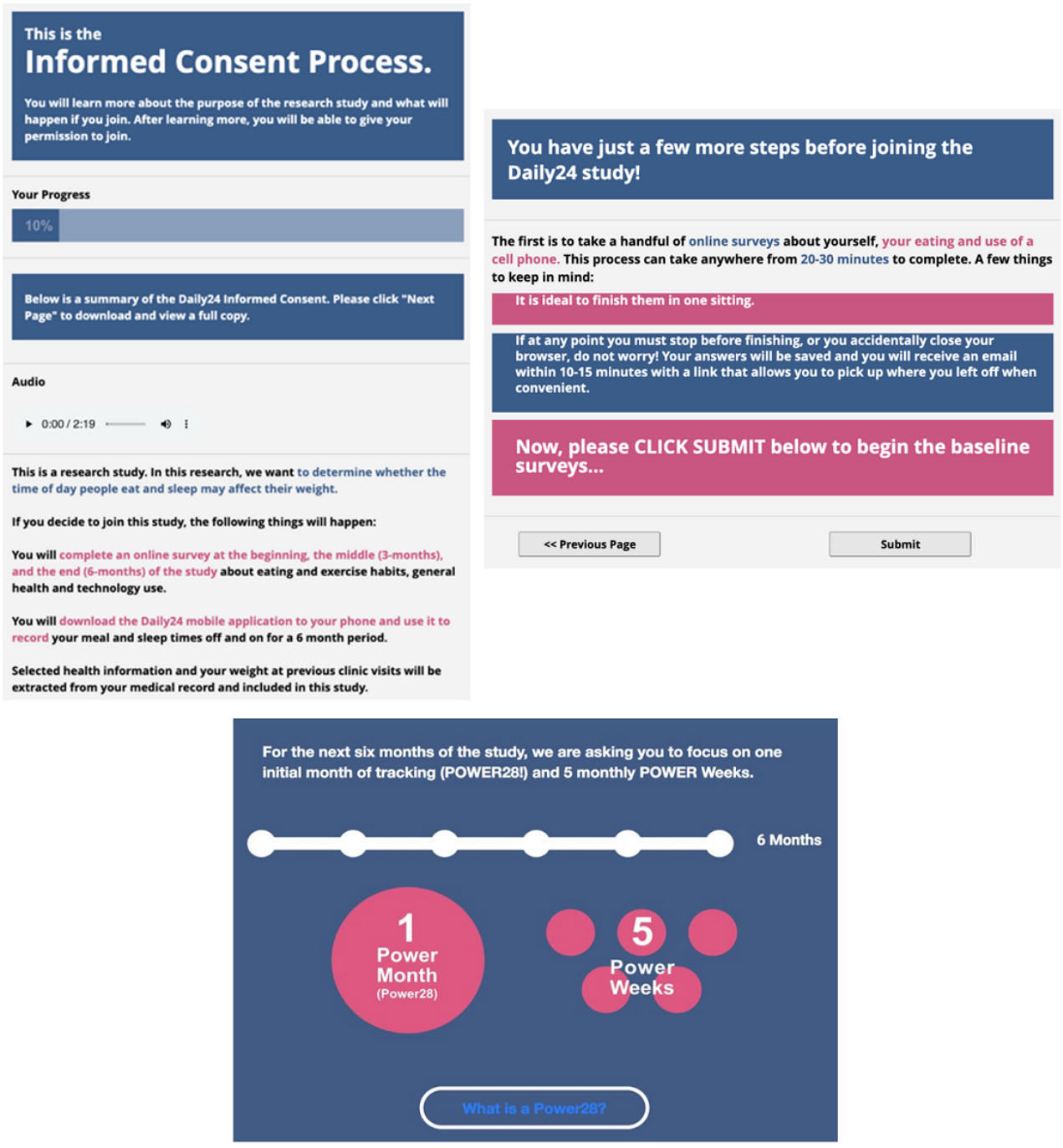
e-Consent and Enrollment Process

We designed a web-based e-consent process in REDCap (Research Electronic Data Capture) beginning with a study description, including participant expectations and duration (see Figure 1). Upon confirming interest, participants proceeded to the e-consent form, which included a supplemental audio clip of the consent form being read aloud, followed by a short quiz

XSL•FO
RenderX

to ensure comprehension. Consenting participants provided identifying information (ie, full name, date of birth, and email), enabling staff to link each participant to the EHR for future analyses (not reported herein). Once consented, participants received a link to complete online baseline surveys (Figure 1). Participants were considered enrolled in the cohort after completing baseline surveys, at which time they received instructions on how to download and use the Daily24 mobile app.

Figure 1. Screenshots of web-based electronic recruitment and onboarding: electronic consent (top left), baseline surveys (top right), and POWER 28 and POWER week information (bottom).



Daily24 Mobile App, Registration, and Download

The Daily24 mobile app was custom designed by our research team to collect information from participants about the timing of eating and sleep, including wake time, sleep time, timing of each eating occasion, and estimate of amount eaten (ie, small, medium, or large meal; small or large snack; or drink, except water, without food) during a 24-hour window (Figure 2). The design of the app is described elsewhere [31]. We benefited from the input of patient and end-user stakeholders in the design of the mobile app, as well as recruitment and retention methods, and we pilot-tested the app [31,39].

Following enrollment in the cohort, participants received a text message on their mobile phones with a unique link to the Daily24 registration form. This unique link contained a token

XSL•FO
RenderX

(ie, 11-character universally unique identifier) that enabled the study team to connect participants' data between the mobile app and their online enrollment information and surveys, while preserving privacy. Registration included an overview of how to use the app, study timeline, and incentives (see next section), followed by selection of a unique "Daily24 name" from a list of randomly generated combined nouns (eg, "FloatHarbor") that, once selected, was the participant's Daily24 username. Participants then received a link to download the Daily24 app via iOS (Apple Store) or Android (Google Play).

Figure 2. Screenshots of the Daily24 app: empty sleep ring (top left), complete sleep ring (top middle), empty food ring (top right), meal size selection (bottom left), complete food ring (bottom middle), and complete day (bottom right).

Engagement Strategies to Promote Use of the Daily24 Mobile App

Although we encouraged participants to enter as much data as possible over the 6-month study, we developed and applied strategies aimed at maximizing app use during their first 4 weeks of participation (ie, 28 days after downloading the app, called "POWER 28") and 1 week per month for the remaining 5 months of the study, called "POWER weeks" (Figure 1). These highly targeted usage days for the study were equivalent to 63 days (POWER 28 + POWER weeks × 5 weeks). Engagement strategies included a leaderboard, badges, raffles, and text reminders. The leaderboard displayed the number of consecutive days tracked on one tab (ie, streak) and total number of all days tracked on the other tab. Earned badges encouraged various types of app use, including one-time badges, streak badges, and POWER week badges (Figure 3). We raffled off US $25 gift cards weekly throughout the study, with those earning more badges having greater odds of winning the raffle. We used emails, SMS text messages, and in-app notifications to encourage usage and to remind participants where they were in their POWER 28 and when a POWER week was coming up. The logic for these messages was triggered both by time (ie, close to a POWER week) as well as by lack of a response (ie, an event missing data). If a participant was on track with logging events, we simply encouraged their continued involvement and did not send additional reminders.

XSL•FO
RenderX

Figure 3. Screenshots of the badges earned to encourage usage in the Daily24 app.

Data Collection

Daily24 app usage data were collected using Amazon Web Services. Self-reported online surveys were administered using REDCap at baseline and at the end of 4 months using standard measures assessing demographics, mHealth use, height, and weight, as well as eating, physical activity, and sleep habits. Technology use and health app use were assessed using the Pew Social Media Update 2016 [40] and a survey measuring characteristics of health app use [16], respectively. Nutritional and eating assessments included the National Health and Nutrition Examination Survey 2009-2010 Dietary Screener Questionnaire [41], which provided estimates of fruit, vegetable, and sugar-sweetened beverage intake over the last 30 days. Physical activity was assessed using the self-administered, short version of the International Physical Activity Questionnaire [42]; physical activity levels were categorized as low, medium, or high over the last 7 days. Sleep measurements included the single sleep quality item from the Pittsburgh Sleep Quality Index [43] and study-created questions about frequency of daytime naps.

To facilitate and encourage baseline survey completion, participants received automated reminders at 15 minutes, 24 hours, and 48 hours after consent and had up until 90 days after receiving the initial survey link to complete the survey. Personalized survey-engagement strategies included a combination of staff emails, text messages, and US $100 raffle gift cards. Participants had up until 60 days to complete the 4-month measures, but this paper only reports baseline survey descriptive results. Data collection was completed in January 2020.

App Usage Categories

App users were defined as using the Daily24 app for at least one day, which was captured by having at least one meal and sleep entry, on at least one day, and pushing "done for the day" on the screen. Nonusers either did not register or download the app or did not push "done for the day" on any day. App use was further categorized into three non–mutually exclusive ways:

1. Immediate use, defined as using the app for 7 days or more during the POWER 28.
2. Consistent use, defined as using the app for 28 days or more during the entire 6-month study, which was based on using the app equal to or more than the median overall days of use for the entire 6-month study.
3. Sustained use, defined as using the app for at least 2 days during the last POWER week (month 6) of the study.

Statistical Analyses

This was a secondary analysis of data from a parent cohort study. We used descriptive statistics (Student t tests or χ^2 tests) for baseline characteristics for all participants and by app use versus nonuse categories. App use was also described in median days of use for the total study, median days used in targeted and nontargeted usage days, and frequency of 2 or more and 7 or more days of use during each month of the study. We selected these two categories based on the following logic:

1. Two or more days: this was selected to represent a low threshold of app use that was not identical to the minimal definition of being an app user.
2. Seven or more days: this was selected because we focused on POWER weeks during months 2 to 6 and wanted to capture those who achieved at least one week of usage.

We evaluated the association between baseline characteristics, with app usage as the dependent variable, using multivariable logistic regression models. Multivariable logistic regression was also used to model the association between baseline characteristics with immediate, consistent, and sustained app use. We used two models with progressive adjustment. Model 1 adjusted for key demographics, including age, sex, race, education, household income, and children younger than 18 years old. Model 2 additionally adjusted for key behavioral factors that could influence engagement, including physical activity, fruit and vegetable servings, sleep quality, and BMI. Covariates were nonmissing, prespecified, and based on a priori hypotheses.

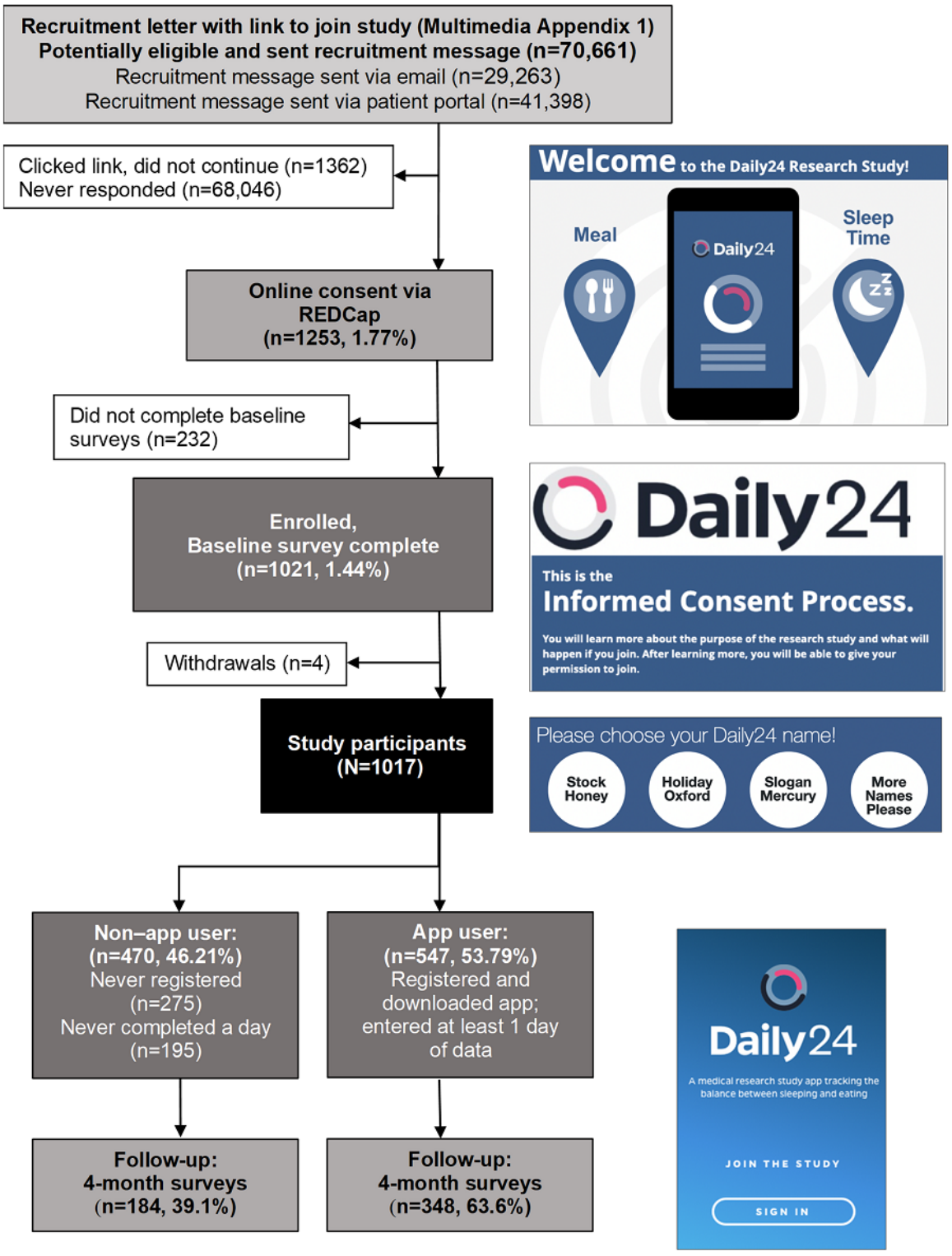
XSL•FO
RenderX

Results

Enrollment and App Use

Figure 4 shows the enrollment flow of eligible participants. Electronic recruitment messages were sent to 70,661 potentially eligible participants, with 1253 participants (1.77%) completing the e-consent process and 1021 (1.44%) enrolling by completing baseline surveys. A total of 4 participants withdrew, leaving 1017 participants included in the analytic sample. Participant characteristics are reported in Table 1. The majority of the 1017 participants were female (n=790, 77.68%), White (n=788, 77.48%), and college graduates (n=749, 73.65%), and the mean age was 51.1 (SD 15.0) years.

Figure 4. Recruitment and retention flow. REDCap: Research Electronic Data Capture.



XSL•FO
RenderX

Table 1. Description of study participants and potential confounders at baseline.

Characteristics and confounders	All participants (N=1017)	Non–app users (n=470)	App users[a] (n=547)	*P* value[b]
Age (years), mean (SD)	51.1 (15.0)	53.2 (14.6)	49.3 (15.0)	<.001
Gender, n (%)				
Male	224 (22.0)	115 (24.5)	109 (19.9)	.07
Female	790 (77.7)	355 (75.5)	435 (79.5)	
Prefer not to answer	3 (0.3)	0 (0)	3 (0.5)	
Race, n (%)				
White	788 (77.5)	351 (74.7)	437 (79.9)	.14
Black	149 (14.7)	82 (17.4)	67 (12.2)	
Asian	29 (2.9)	13 (2.8)	16 (2.9)	
Pacific Islander, American Indian, or others	17 (1.7)	10 (2.1)	7 (1.3)	
Two or more races	34 (3.3)	14 (3.0)	20 (3.7)	
Site, n (%)				
Site A	51 (5.0)	23 (4.9)	28 (5.1)	.004
Site B	282 (27.7)	105 (22.3)	177 (32.4)	
Site C	200 (19.7)	96 (20.4)	104 (19.0)	
Educational level, n (%)				
High school or less	63 (6.2)	40 (8.5)	23 (4.2)	<.001
Some college	205 (20.2)	109 (23.2)	96 (17.6)	
College graduate	749 (73.6)	321 (68.3)	428 (78.2)	
Annual household income (US $), n (%)				
<35,000	120 (11.8)	70 (14.9)	50 (9.1)	.02
35,000 to <50,000	109 (10.7)	53 (11.3)	56 (10.2)	
50,000 to <75,000	148 (14.6)	66 (14.0)	82 (15.0)	
≥75,000	550 (54.1)	234 (49.8)	316 (57.8)	
Don’t know/choose not to answer	90 (8.8)	47 (10.0)	43 (7.9)	
Any child <18 years old, n (%)	248 (24.4)	129 (27.4)	119 (21.8)	.04
Height (cm), mean (SD)	168.9 (50.2)	170.6 (73.3)	167.4 (8.5)	.31
Weight (kg), mean (SD)	85.8 (23.8)	86.8 (25.1)	85.0 (22.5)	.23
BMI[c], mean (SD)	30.5 (7.9)	30.8 (8.2)	30.3 (7.6)	.29
BMI categories, n (%)				
Underweight (<18.5)	14 (1.4)	7 (1.5)	7 (1.3)	.75
Normal (18.5 to <25)	250 (24.6)	111 (23.6)	139 (25.4)	
Overweight (25 to <30)	288 (28.3)	129 (27.4)	159 (29.1)	
Obese (≥30)	465 (45.7)	223 (47.4)	242 (44.2)	
Fruit or vegetable cup equivalent, mean (SD)	2.9 (1.5)	2.8 (1.6)	3.0 (1.4)	.18
Added sugars tsp equivalent from sugar-sweetened beverages, mean (SD)	0.8 (1.3)	1.0 (1.5)	0.7 (1.2)	.004
Physical activity, n (%)				
Low	20 (5.1)	11 (6.3)	9 (4.1)	.53
Medium	221 (55.8)	94 (53.7)	127 (57.5)	
High	155 (39.1)	70 (40.0)	85 (38.5)	

XSL•FO
RenderX

Characteristics and confounders	All participants (N=1017)	Non–app users (n=470)	App users[a] (n=547)	*P* value[b]
Sleep quality, n (%)				
Very good	188 (18.5)	76 (16.2)	112 (20.5)	.02
Fairly good	496 (48.8)	222 (47.2)	274 (50.1)	
Fairly bad	274 (26.9)	136 (28.9)	138 (25.2)	
Very bad	59 (5.8)	36 (7.7)	23 (4.2)	
Nap frequency, n (%)				
<1 per week	581 (57.1)	267 (56.8)	314 (57.4)	.03
1 per week	165 (16.2)	69 (14.7)	96 (17.6)	
2-3 per week	176 (17.3)	79 (16.8)	97 (17.7)	
4-6 per week	57 (5.6)	28 (6.0)	29 (5.3)	
Daily	38 (3.7)	27 (5.7)	11 (2.0)	
Number of health apps used in past month, n (%)				
0	212 (20.8)	127 (27.0)	85 (15.5)	<.001
1-5	705 (69.3)	299 (63.6)	406 (74.2)	
>5	100 (9.8)	44 (9.4)	56 (10.2)	
App use reasons, n (%)				
Track how much exercise I get	665 (65.4)	269 (57.2)	396 (72.4)	<.001
Track what I eat/improve what I eat	531 (52.2)	212 (45.1)	319 (58.3)	<.001
Weight loss	476 (46.8)	206 (43.8)	270 (49.4)	.08
Track a health measure	203 (20.0)	86 (18.3)	117 (21.4)	.22
Track how much sleep I get	346 (34.0)	132 (28.1)	214 (39.1)	<.001

[a]App user is defined as downloading the app and recording at least one entry on at least one day.

[b]The *P* value for a group of variables is reported in the row of the first variable.

[c]BMI is calculated as weight in kilograms divided by height in meters squared.

Out of 1017 participants, 547 (53.79%) were app users (ie, downloaded the app and recorded at least one entry on at least one day). When examining app users by use category, 412 (75.3%), 274 (50.1%), and 139 (25.4%) were categorized as immediate, consistent, and sustained users, respectively. Of the sustained users, 116 (83.5%) used the app at least one day every month of the study, and 133 (95.7%) used the app at least one day for 5 out of the 6 months. In comparison to non–app users (471/1017, 46.31%), app users were younger (mean 49.3 vs 53.3 years; *P*<.001), more likely to be college graduates (78.2% vs 68.3%; *P*<.001), had greater annual income (>US $50,000: 398/547, 72.8% vs 300/470, 63.8%; *P*=.02), and were less likely to have children younger than 18 years old (21.8% vs 27.4%; *P*=.04). There were no differences between app users and nonusers regarding weight, height, mean BMI, and BMI category. App users were less likely to drink sugar-sweetened beverages (mean sugar tsp equivalent: 0.7 vs 1.0; *P*=.004), reported better sleep quality (fairly good or very good: 386/547, 70.6% vs 298/470, 63.4%; *P*=.02), and were less likely to take daily naps (2.0% vs 5.7%; *P*=.03). They were also more likely to use health apps overall (462/547, 84.5% vs 343/470, 73.0%; *P*<.001), and to use them for the purpose of tracking exercise (396/547, 72.4% vs 269/470, 57.2%), eating (319/547, 58.3% vs 212/470, 45.1%), and sleep (214/547, 39.1% vs 132/470, 28.1%; *P*<.001 for all).

The median amount of app use was 28 (IQR 7-75) days over the 6-month study, 20 (IQR 7-35) days during the targeted 63 days of the study, and 6 (IQR 0-41) days during the 117 nontargeted days of the study. Table 2 describes app use by study month. During study month 1, the vast majority of app users (92.3%) used the app for 2 or more days and 76.2% used it for 7 or more days. Usage decreased over time in the cohort study. By month 6, 27.1% of app users used the app for 2 or more days and 20.1% used it 7 or more days.

XSL•FO
RenderX

Table 2. Monthly Daily24 app use during the 6-month cohort study by users who completed at least one day of app use.

Month[a]	Participants who used the app (n=547), n (%)	
	Used ≥2 days	Used ≥7 days
Month 1	505 (92.3)	417 (76.2)
Month 2	269 (49.2)	214 (39.1)
Month 3	213 (38.9)	166 (30.3)
Month 4	183 (33.5)	138 (25.2)
Month 5[b] (n=536)	157 (29.3)	133 (24.8)
Month 6[b] (n=527)	143 (27.1)	106 (20.1)

[a]A study month is defined as 4 weeks (28 days). To enable all study months to begin on a Monday, the time between the end of POWER 28 and start of month 2 ranged from 15 to 21 days.

[b]Due to late registration, some participants were not able to reach months 5 and 6 of the study.

Predictors of Usage of the Daily24 App

Table 3 shows the multivariable regression model for app use versus nonuse. Younger age, White (vs non-White) race, greater education, higher household income, not having children less than 18 years of age, and having used 1 to 5 apps in the past were statistically significantly associated with app use (vs non–app use). Black participants were one-third less likely to use the app than White participants, whereas those with greater than a college education and a higher income (≥US $75,000 vs <US $35,000) were statistically significantly more likely to use the app. Those with children under the age of 18 years were 45% less likely to use the app, and those who had used 1 to 5 apps in the past month were 70% more likely to use the app compared to those who had not used apps in the past month.

Table 4 shows multivariable regression models for immediate, consistent, and sustained use. Older age and lower BMI were statistically significantly associated with increased immediate, consistent, and sustained app use. Having children less than 18 years old was statistically significantly associated with decreased immediate use, and better sleep quality was associated with increased immediate and consistent app use.

XSL•FO
RenderX

Table 3. Multivariable regression models for Daily24 app use versus nonuse.

Risk factors	Model 1[a]		Model 2[b]	
	OR[c] (95% CI)	*P* value	OR (95% CI)	*P* value
Demographic risk factors				
Age, per 10-year increase	0.77 (0.70-0.85)	<.001	0.78 (0.71-0.86)	<.001
Gender				
Male	Ref[d] (1)		Ref (1)	
Female	1.32 (0.96-1.81)	.09	1.22 (0.88-1.69)	.23
Race				
White	Ref (1)		Ref (1)	
Black	0.66 (0.45-0.96)	.03	0.67 (0.46-0.98)	.04
Other	0.80 (0.49-1.31)	.38	0.82 (0.50-1.34)	.43
Educational level				
<College	Ref (1)		Ref (1)	
≥College	1.39 (1.03-1.89)	.03	1.36 (1.00-1.86)	.05
Household income (US $)				
<35,000	Ref (1)		Ref (1)	
35,000 to <50,000	1.58 (0.91-2.72)	.10	1.40 (0.80-2.44)	.24
50,000 to <75,000	2.01 (1.20-3.38)	.01	1.82 (1.07-3.07)	.03
≥75,000	2.30 (1.47-3.61)	<.001	2.00 (1.26-3.17)	.003
Any child <18 years old				
No	Ref (1)		Ref (1)	
Yes	0.53 (0.39-0.73)	<.001	0.55 (0.40-0.75)	<.001
Behavioral risk factors				
Physical activity				
Low or medium	—[e]	—	Ref (1)	
High	—	—	0.93 (0.61-1.43)	.75
Fruit and vegetable cups, per 1-cup increase	—	—	1.04 (0.95-1.14)	.41
Sleep quality				
Very good or fairly good	—	—	Ref (1)	
Very bad or fairly bad	—	—	0.79 (0.59-1.04)	.09
Number of health apps used in past month				
0	—	—	Ref (1)	
1-5	—	—	1.70 (1.22-2.37)	.002
>5	—	—	1.40 (0.84-2.35)	.20
BMI[f], per 1-unit increase	—	—	1.00 (0.98-1.02)	.99

[a]Model 1 was adjusted for age, sex, race, education, household income, and having children younger than 18 years old.

[b]Model 2 included model 1 parameters and was adjusted for physical activity, fruit and vegetable cups, sleep quality, and BMI.

[c]OR: odds ratio.

[d]Ref: reference.

[e]Not calculated since these parameters were not included in model 1.

[f]BMI is calculated as weight in kilograms divided by height in meters squared.

XSL•FO
RenderX

Table 4. Multivariable regression models[a] for immediate, consistent, and sustained Daily24 app use (n=547).

Risk factors	Immediate use: using app for ≥7 days during POWER 28 (n=412)		Consistent use: using app for ≥28 days for 6 months (n=274)		Sustained use: using app for ≥2 days during POWER week 5 (n=139)	
	OR[b] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Demographic risk factors						
Age, per 10-year increase	1.28 (1.09-1.50)	.003	1.40 (1.22-1.61)	<.001	1.54 (1.31-1.82)	<.001
Gender						
Male	Ref[c] (1)		Ref (1)		Ref (1)	
Female	0.69 (0.38-1.26)	.23	0.60 (0.37-0.98)	.04	0.74 (0.44-1.25)	.26
Race						
White	Ref (1)		Ref (1)		Ref (1)	
Black	0.70 (0.38-1.30)	.26	0.86 (0.48-1.52)	.60	0.92 (0.46-1.84)	.82
Other	0.79 (0.38-1.66)	.53	0.67 (0.33-1.36)	.27	1.03 (0.45-2.38)	.94
Educational level						
<College	Ref (1)		Ref (1)		Ref (1)	
≥College	1.00 (0.59-1.71)	.99	0.99 (0.61-1.59)	.96	1.46 (0.82-2.60)	.20
Household income (US $)						
<35,000	Ref (1)		Ref (1)		Ref (1)	
35,000 to <50,000	0.94 (0.38-2.29)	.89	1.05 (0.46-2.42)	.91	0.86 (0.31-2.37)	.77
50,000 to <75,000	0.83 (0.36-1.92)	.66	1.18 (0.54-2.56)	.68	0.87 (0.35-2.16)	.76
≥75,000	1.00 (0.47-2.14)	.99	0.76 (0.38-1.53)	.44	0.62 (0.27-1.41)	.25
Any child <18 years old						
No	Ref (1)		Ref (1)		Ref (1)	
Yes	0.56 (0.34-0.91)	.02	0.68 (0.43-1.07)	.10	0.70 (0.38-1.28)	.24
Behavioral risk factors						
Physical activity						
Low or medium	Ref (1)		Ref (1)		Ref (1)	
High	0.68 (0.35-1.32)	.25	1.01 (0.55-1.83)	.98	1.30 (0.68-2.51)	.43
Fruit and vegetable cups, per 1-cup increase	0.88 (0.75-1.03)	.10	1.03 (0.90-1.19)	.64	0.98 (0.84-1.15)	.82
Sleep quality						
Very good or fairly good	Ref (1)		Ref (1)		Ref (1)	
Very bad of fairly bad	0.59 (0.38-0.93)	.02	0.63 (0.42-0.95)	.03	0.74 (0.45-1.21)	.23
Number of health apps used in past month						
0	Ref (1)		Ref (1)		Ref (1)	
1-5	0.86 (0.45-1.63)	.64	1.12 (0.66-1.92)	.67	1.46 (0.81-2.62)	.20
>5	1.03 (0.42-2.51)	.95	1.04 (0.49-2.24)	.91	0.60 (0.21-1.70)	.34
BMI[d], per 1-unit increase	0.96 (0.94-0.99)	.01	0.95 (0.93-0.98)	.001	0.95 (0.92-0.99)	.01

[a]The model was adjusted for age, sex, race, education, household income, having children younger than 18 years old, physical activity, fruit and vegetable cups, sleep quality, and BMI.

[b]OR: odds ratio.

[c]Ref: reference.

[d]BMI is calculated as weight in kilograms divided by height in meters squared.

XSL•FO
RenderX

Survey Completion and Retention in the EHR-Based Cohort Study

Out of 1017 enrolled participants, 328 (32.25%) completed the 4-month follow-up surveys within 72 hours of receiving the link. Of the remaining 689 participants (67.75%), study staff were able to reach out to 610 participants (88.5%) through personalized emails, text messages, and US $100 raffle invitations delivered up to one week prior to study completion. Of the 610 contacted participants, 113 (18.5%) completed their surveys after one contact, 56 (9.2%) after two contacts, and 35 (5.7%) after three contacts, increasing the overall number of 4-month survey completers to 532 (52.31%).

Discussion

Principal Findings

EHRs and patient portals are readily available through most health care systems, and use of mHealth apps is fairly ubiquitous [16,44]. This study reports on EHR-based recruitment of adults from three health systems to use the Daily24 mobile app to record daily timing of meals, snacks, and sleep for 6 months. We emailed research invitations to over 70,000 potentially eligible participants identified through the EHR using efficient identification (ie, computable phenotype) and messaging methods (ie, emails sent directly through the EHR patient portal or to personal email addresses). A total of 1.4% of participants completed e-consent forms and baseline questionnaires in a period of 6 months, a yield that is slightly lower than reports for other EHR-based recruitment methods [32-34]. In a 2019 single-institution study that included 13 separate EHR-based recruitment strategies using the patient portal recruitment service, the average response rate for patient portal messages was 2.9% [32]. Our lower yield might be explained by the study's expectation to download and actively use an app for 6 months or have no guaranteed compensation be provided for participation [45] (ie, raffles of gift cards). Patients may also be more likely to respond to mHealth research with a behavioral intervention [46] or to disease-related versus wellness-related research [32,45]. In the above study by Miller and colleagues [32], recruitment response rates were higher (3.4%) among condition-specific studies (ie, those with a more inclusive comprehensive phenotype) versus general health studies (1.4%). The latter response rate was identical to this study's recruitment yield, which was also not specific to a health condition. Furthermore, while our app included gaming elements (eg, badges and a leaderboard) [47,48] to increase data entry, we intentionally did not include behavioral techniques (eg, goal setting and personalized behavioral prompts) aimed at behavior change, given the study's primary goal to naturalistically examine the relationship between timing of eating and sleep and weight and medical conditions (findings forthcoming).

Once enrolled, 54% of participants who downloaded the app entered timing of eating or sleep data on at least one day. While the frequency criteria to classify someone as an app user in this study was fairly low (ie, at least one completed day), other studies have used a similarly low frequency to define usage [49]; however, comparisons between studies can be challenging due to disparate study designs and modes of interacting with apps (ie, passive vs active data collection) [50]. For example, in the Asthma Mobile Health Study (AMHS), 85.21% (6470/7593) of enrolled participants (ie, downloaded an asthma health app, e-consented, and verified email) were considered baseline users (ie, at least one in-app survey entry). However, enrollment occurred after the app was already downloaded, and individuals who downloaded the app (N=40,683 in the United States over 6 months) were recruited through a large media blitz versus academic recruitment [49]. Eligibility was also based on having a medical condition (ie, disease related), and the app included behavioral components (eg, goal setting).

Although criteria for defining usage categories differ across studies, our immediate (75%; ≥7 days in the first month) and consistent (50%; ≥28 days over 6 months) rates were higher than the "robust" cohort rates (30%; 5 or more surveys over 6 months) reported in the AMHS; in the case of sustained users (25%; ≥2 days during month 6), our rates were fairly comparable to those in the AMHS [49]. We attribute being able to initially engage three-quarters of our app users, and to retain a quarter of our users, to the food and sleep wheels in Daily24 being fast and easy to use, whereas other apps may include more survey items or require more detailed dietary intake entry [31,51]. Future iterations of the app should employ evidence-based strategies and features for increasing engagement (eg, push notifications with tailored health messages) [52-54].

This study's usage data provides important information about predictors of health app use to guide the design of future observational studies using apps. Our finding that those who were younger, more formally educated, and wealthier were more likely to be app users is consistent with past research [16,49]. This study also found that White participants were more likely to be app users, a finding that is consistent with some research [55]. However, that finding is not consistent with a cross-sectional survey study of 1604 mobile phone users in the United States [16], which found that being Latino or Hispanic (P<.05) or African American (P=.07, trend) were related to a greater likelihood to download a health app. Inconsistencies in findings may be related to different assessment methods (ie, actual app usage vs self-reported use), recruitment methods (ie, national survey vs regional EHR recruitment), and racial and ethnic distribution in recruitment regions [16,56]. Not having children younger than 18 years of age was also associated with app use. While this is perhaps a correlate of being younger, it is also an intuitive finding that those with children may have less time for mHealth app use, supporting the well-documented importance of ease and efficiency of data entry in mHealth apps [16]. While younger age was associated with app use overall, being older was associated with early, consistent, and sustained use. The AMHS study similarly found that among robust users, increasing age was significantly associated with a greater likelihood to use the asthma health app daily [49]. An adherence and retention study of a web-based alcohol intervention also found that being older and not having children predicted a greater likelihood of logging in [57]. We also found that having a lower BMI was associated with early, consistent, and sustained use. Past research has found that having a BMI in the obese range is associated with greater health app use [16], influencing our hypothesis that those with higher BMIs would be more

XSL•FO
RenderX

motivated to download and use a health app. While we did not find a significant association between weight status and app use overall, we did find that those with lower BMIs were more likely to use health apps across time. Given this observational study design, we identified an association between BMI and health app use (ie, those with a lower BMI were more likely to engage in sustained tracking and health monitoring), but we do not know causality or temporality. Future research exploring a causal relationship is needed to determine if apps targeting timing of eating and sleep may have an effect on behaviors that influence weight [25,56].

Limitations

There are several limitations of this study. First, this was an observational cohort study and was not designed with a comparison group to assess differences in app use among participants instructed to log for 6 months without additional guidance on targeted tracking days, compared to our approach of emphasizing tracking on POWER 28 and POWER week days. In designing the study, to optimize longitudinal tracking, we decided to preidentify targeted days to decrease participant burden and, more importantly, to increase the likelihood that we would collect data on some days across each of the 6 months rather than risk the typical pattern of heavier use up front followed by drop-off [18,49]. This approach appeared to be effective. Although we did observe drop-off across each study month, with the biggest decline being from months 1 and 2, about one-quarter of the participants were still using the app during month 6, and those who were using the app during month 6 were using it in the identified POWER week. However, without a two-arm study design, we cannot fully conclude that this was the ideal approach. Second, although we designed badges and a leaderboard to create gaming elements and increase motivation [47,48], we are unable to ascertain if those who earned badges were more motivated individuals in general or were motivated by the badges. Badges were earned based on various categories of usage (eg, first log-in, track 7 days in a row, and 4 days of your POWER week) and were automatically entered into our raffle (ie, participants did not have to enter their badges into the raffle themselves); thus, it is challenging to know whether badges and the resulting raffle were an effective gamification approach. Third, we do not have detailed information on the reasons that a little less than half of the participants did not go on to download the app. With our app being designed by researchers rather than more highly funded industry, we suspect that the onboarding process may have had some cumbersome features. The biggest obstacles may have been problems with the two-factor authentication process, which required participants to receive an SMS code on their device and correctly enter it to verify their identity. Additionally, many people forgot, misplaced, or mis-entered the password they chose when registering for the app and were without an automated password reset option. Although we had research staff available for tech support, it was available only during work hours and via phone or email. Fourth, our sample was largely comprised of White participants, more formally educated participants, and those of middle- to upper-socioeconomic status; thus, the generalizability to other racial, ethnic, and socioeconomic groups is limited. Future research involving EHR-based recruitment independent of technology use might consider partnering with communities from racialized and lower-socioeconomic subgroups to understand how recruitment efforts and health apps can be adapted to improve their impact for marginalized communities. Finally, while our recruitment methods were efficient in terms of participant identification, messaging, and enrollment, we are unable to comment on the cost-effectiveness of EHR enrollment. Each of these health systems have made significant investments into building and maintaining their EHRs and infrastructure to enable these recruitment methods for research purposes. In addition, for this study, we leveraged existing health information technology infrastructure from the PaTH network [30], which enabled efficiency from both a time and resource perspective. However, for this methodology to be used more broadly in a variety of settings, greater institutional and community partnerships and resources are needed.

Conclusions

Health apps aimed at weight loss and related behaviors are among the most highly used mHealth apps [22]. Time-restricted feeding is a novel and promising approach for obesity and related disease management; however, it is largely untested in humans, to a great extent due to the challenges of helping individuals modify their behavior to a shorter window of eating [15,58,59]. This report is a first step in describing efficient EHR recruitment of patients from three large health institutions and the use of an mHealth app to enter information about timing of eating and sleep patterns. Next steps include incorporating behavioral techniques into the app, potentially with health coaching, to assist individuals achieve greater alignment with their circadian rhythms and to determine whether this is a feasible and effective weight loss intervention.

Acknowledgments

We would like to thank Jonathan Martinez for his role in recruitment and retention and Hao Da for her role in data management. We would also like to acknowledge Mengru Liao for her design of the Daily24 web registration flow and contribution to the design of the REDCap e-consent flow. This work was funded by the American Heart Association. The authors acknowledge assistance for clinical data coordination and retrieval from the Johns Hopkins School of Medicine's Core for Clinical Research Data Acquisition, supported, in part, by the Johns Hopkins Institute for Clinical and Translational Research (UL1TR001079).

Conflicts of Interest

TBW founded DaiWare, but has no current commercial projects associated with DaiWare. JMC has served on a Scientific Advisory Board for Novo Nordisk and Boehringer Ingelheim.

XSL•FO
RenderX

Multimedia Appendix 1
Sample recruitment message.
[PDF File (Adobe PDF File), 77 KB - jmir_v24i6e34191_app1.pdf]

References

1. Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: A systematic analysis for the Global Burden of Disease Study 2013. Lancet 2014 Aug 30;384(9945):766-781 [FREE Full text] [doi: 10.1016/S0140-6736(14)60460-8] [Medline: 24880830]
2. Khan SS, Ning H, Wilkins JT, Allen N, Carnethon M, Berry JD, et al. Association of body mass index with lifetime risk of cardiovascular disease and compression of morbidity. JAMA Cardiol 2018 Apr 01;3(4):280-287 [FREE Full text] [doi: 10.1001/jamacardio.2018.0022] [Medline: 29490333]
3. Williams EP, Mesidor M, Winters K, Dubbert PM, Wyatt SB. Overweight and obesity: Prevalence, consequences, and causes of a growing public health problem. Curr Obes Rep 2015 Sep;4(3):363-370. [doi: 10.1007/s13679-015-0169-4] [Medline: 26627494]
4. Arroyo-Johnson C, Mincey KD. Obesity epidemiology worldwide. Gastroenterol Clin North Am 2016 Dec;45(4):571-579 [FREE Full text] [doi: 10.1016/j.gtc.2016.07.012] [Medline: 27837773]
5. Dwivedi AK, Dubey P, Cistola DP, Reddy SY. Association between obesity and cardiovascular outcomes: Updated evidence from meta-analysis studies. Curr Cardiol Rep 2020 Mar 12;22(4):25. [doi: 10.1007/s11886-020-1273-y] [Medline: 32166448]
6. Ramage S, Farmer A, Eccles KA, McCargar L. Healthy strategies for successful weight loss and weight maintenance: A systematic review. Appl Physiol Nutr Metab 2014 Jan;39(1):1-20 [FREE Full text] [doi: 10.1139/apnm-2013-0026] [Medline: 24383502]
7. Cleven L, Krell-Roesch J, Nigg CR, Woll A. The association between physical activity with incident obesity, coronary heart disease, diabetes and hypertension in adults: A systematic review of longitudinal studies published after 2012. BMC Public Health 2020 May 19;20(1):726 [FREE Full text] [doi: 10.1186/s12889-020-08715-4] [Medline: 32429951]
8. Wadden TA, Tronieri JS, Butryn ML. Lifestyle modification approaches for the treatment of obesity in adults. Am Psychol 2020;75(2):235-251 [FREE Full text] [doi: 10.1037/amp0000517] [Medline: 32052997]
9. McHill AW, Wright KP. Role of sleep and circadian disruption on energy expenditure and in metabolic predisposition to human obesity and metabolic disease. Obes Rev 2017 Feb;18 Suppl 1:15-24. [doi: 10.1111/obr.12503] [Medline: 28164449]
10. Nedeltcheva A, Scheer FAJL. Metabolic effects of sleep disruption, links to obesity and diabetes. Curr Opin Endocrinol Diabetes Obes 2014 Aug;21(4):293-298 [FREE Full text] [doi: 10.1097/MED.0000000000000082] [Medline: 24937041]
11. Dong TA, Sandesara PB, Dhindsa DS, Mehta A, Arneson LC, Dollar AL, et al. Intermittent fasting: A heart healthy dietary pattern? Am J Med 2020 Aug;133(8):901-907 [FREE Full text] [doi: 10.1016/j.amjmed.2020.03.030] [Medline: 32330491]
12. LeBlanc ES, Patnode CD, Webber EM, Redmond N, Rushkin M, O'Connor EA. Behavioral and pharmacotherapy weight loss interventions to prevent obesity-related morbidity and mortality in adults: Updated evidence report and systematic review for the US Preventive Services Task Force. JAMA 2018 Sep 18;320(11):1172-1191. [doi: 10.1001/jama.2018.7777] [Medline: 30326501]
13. Hall KD, Kahan S. Maintenance of lost weight and long-term management of obesity. Med Clin North Am 2018 Jan;102(1):183-197 [FREE Full text] [doi: 10.1016/j.mcna.2017.08.012] [Medline: 29156185]
14. MacLean PS, Wing RR, Davidson T, Epstein L, Goodpaster B, Hall KD, et al. NIH working group report: Innovative research to improve maintenance of weight loss. Obesity (Silver Spring) 2015 Jan;23(1):7-15 [FREE Full text] [doi: 10.1002/oby.20967] [Medline: 25469998]
15. Panda S. The arrival of circadian medicine. Nat Rev Endocrinol 2019 Feb;15(2):67-69. [doi: 10.1038/s41574-018-0142-x] [Medline: 30602736]
16. Krebs P, Duncan DT. Health app use among US mobile phone owners: A national survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
17. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile health technology evaluation: The mHealth evidence workshop. Am J Prev Med 2013 Aug;45(2):228-236 [FREE Full text] [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]
18. Dorsey ER, Chan Y, McConnell MV, Shaw SY, Trister AD, Friend SH. The use of smartphones for health research. Acad Med 2017;92(2):157-160. [doi: 10.1097/acm.0000000000001205]
19. Bradway M, Gabarron E, Johansen M, Zanaboni P, Jardim P, Joakimsen R, et al. Methods and measures used to evaluate patient-operated mobile health interventions: Scoping literature review. JMIR Mhealth Uhealth 2020 Apr 30;8(4):e16814 [FREE Full text] [doi: 10.2196/16814] [Medline: 32352394]
20. Feldman DI, Theodore Robison W, Pacor JM, Caddell LC, Feldman EB, Deitz RL, et al. Harnessing mHealth technologies to increase physical activity and prevent cardiovascular disease. Clin Cardiol 2018 Jul;41(7):985-991 [FREE Full text] [doi: 10.1002/clc.22968] [Medline: 29671879]
21. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]

XSL•FO
RenderX

22. Milne-Ives M, Lam C, De Cock C, Van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: Systematic review. JMIR Mhealth Uhealth 2020 Mar 18;8(3):e17046 [FREE Full text] [doi: 10.2196/17046] [Medline: 32186518]
23. Park S, Hwang J, Choi Y. Effect of mobile health on obese adults: A systematic review and meta-analysis. Healthc Inform Res 2019 Jan;25(1):12-26 [FREE Full text] [doi: 10.4258/hir.2019.25.1.12] [Medline: 30788177]
24. Wang Y, Min J, Khuri J, Xue H, Xie B, Kaminsky LA, et al. Effectiveness of mobile health interventions on diabetes and obesity treatment and management: Systematic review of systematic reviews. JMIR Mhealth Uhealth 2020 Apr 28;8(4):e15400 [FREE Full text] [doi: 10.2196/15400] [Medline: 32343253]
25. Ghelani DP, Moran LJ, Johnson C, Mousa A, Naderpoor N. Mobile apps for weight management: A review of the latest evidence to inform practice. Front Endocrinol (Lausanne) 2020;11:412 [FREE Full text] [doi: 10.3389/fendo.2020.00412] [Medline: 32670197]
26. Kearney A, Daykin A, Shaw ARG, Lane AJ, Blazeby JM, Clarke M, et al. Identifying research priorities for effective retention strategies in clinical trials. Trials 2017 Aug 31;18(1):406 [FREE Full text] [doi: 10.1186/s13063-017-2132-z] [Medline: 28859674]
27. Brueton V, Tierney J, Stenning S, Harding S, Meredith S, Nazareth I, et al. Strategies to improve retention in randomised trials. Cochrane Database Syst Rev 2013 Dec 03(12):MR000032 [FREE Full text] [doi: 10.1002/14651858.MR000032.pub2] [Medline: 24297482]
28. Cowie MR, Blomster JI, Curtis LH, Duclaux S, Ford I, Fritz F, et al. Electronic health records to facilitate clinical research. Clin Res Cardiol 2017 Jan;106(1):1-9 [FREE Full text] [doi: 10.1007/s00392-016-1025-6] [Medline: 27557678]
29. Lai YS, Afseth JD. A review of the impact of utilising electronic medical records for clinical research recruitment. Clin Trials 2019 Apr;16(2):194-203. [doi: 10.1177/1740774519829709] [Medline: 30764659]
30. Bennett WL, Bramante CT, Rothenberger SD, Kraschnewski JL, Herring SJ, Lent MR, et al. Patient recruitment into a multicenter clinical cohort linking electronic health records from 5 health systems: Cross-sectional analysis. J Med Internet Res 2021 May 27;23(5):e24003 [FREE Full text] [doi: 10.2196/24003] [Medline: 34042604]
31. Woolf TB, Goheer A, Holzhauer K, Martinez J, Coughlin JW, Martin L, et al. Development of a mobile app for ecological momentary assessment of circadian data: Design considerations and usability testing. JMIR Form Res 2021 Jul 23;5(7):e26297 [FREE Full text] [doi: 10.2196/26297] [Medline: 34296999]
32. Miller H, Gleason K, Juraschek S, Plante T, Lewis-Land C, Woods B, et al. Electronic medical record-based cohort selection and direct-to-patient, targeted recruitment: Early efficacy and lessons learned. J Am Med Inform Assoc 2019 Nov 01;26(11):1209-1217 [FREE Full text] [doi: 10.1093/jamia/ocz168] [Medline: 31553434]
33. Pfaff E, Lee A, Bradford R, Pae J, Potter C, Blue P, et al. Recruiting for a pragmatic trial using the electronic health record and patient portal: Successes and lessons learned. J Am Med Inform Assoc 2019 Jan 01;26(1):44-49 [FREE Full text] [doi: 10.1093/jamia/ocy138] [Medline: 30445631]
34. Gleason KT, Ford DE, Gumas D, Woods B, Appel L, Murray P, et al. Development and preliminary evaluation of a patient portal messaging for research recruitment service. J Clin Transl Sci 2018 Feb 25;2(1):53-56 [FREE Full text] [doi: 10.1017/cts.2018.10] [Medline: 31660218]
35. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. J Am Med Inform Assoc 2014;21(4):578-582 [FREE Full text] [doi: 10.1136/amiajnl-2014-002747] [Medline: 24821743]
36. Forrest CB, McTigue KM, Hernandez AF, Cohen LW, Cruz H, Haynes K, et al. PCORnet® 2020: Current state, accomplishments, and future directions. J Clin Epidemiol 2021 Jan;129:60-67 [FREE Full text] [doi: 10.1016/j.jclinepi.2020.09.036] [Medline: 33002635]
37. Amin W, Tsui F, Borromeo C, Chuang CH, Espino JU, Ford D, PaTH network team. PaTH: Towards a learning health system in the Mid-Atlantic region. J Am Med Inform Assoc 2014;21(4):633-636 [FREE Full text] [doi: 10.1136/amiajnl-2014-002759] [Medline: 24821745]
38. Pathak J, Kho AN, Denny JC. Electronic health records-driven phenotyping: Challenges, recent advances, and perspectives. J Am Med Inform Assoc 2013 Dec;20(e2):e206-e211 [FREE Full text] [doi: 10.1136/amiajnl-2013-002428] [Medline: 24302669]
39. Goheer A, Holzhauer K, Martinez J, Woolf T, Coughlin JW, Martin L, et al. What influences the "when" of eating and sleeping? A qualitative interview study. Appetite 2021 Jan 01;156:104980. [doi: 10.1016/j.appet.2020.104980] [Medline: 32980457]
40. Greenwood S, Perrin A, Duggan M. Social Media Update 2016. Washington, DC: Pew Research Center; 2016 Nov 11. URL: https://www.pewresearch.org/internet/2016/11/11/social-media-update-2016/ [accessed 2022-05-19]
41. Thompson FE, Midthune D, Kahle L, Dodd KW. Development and evaluation of the National Cancer Institute's Dietary Screener Questionnaire scoring algorithms. J Nutr 2017 Jun;147(6):1226-1233 [FREE Full text] [doi: 10.3945/jn.116.246058] [Medline: 28490673]
42. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]

XSL•FO
RenderX

43. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4]
44. Office of the National Coordinator for Health Information Technology. Percent of hospitals, by type, that possess certified health IT. Health IT quick-stat #52. HealthIT.gov. 2018 Sep. URL: https://www.healthit.gov/data/quickstats/percent-hospitals-type-possess-certified-health-it [accessed 2022-05-19]
45. Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, et al. Indicators of retention in remote digital health studies: A cross-study evaluation of 100,000 participants. NPJ Digit Med 2020;3:21 [FREE Full text] [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
46. Lee J, Choi M, Lee SA, Jiang N. Effective behavioral intervention strategies using mobile health applications for chronic disease management: A systematic review. BMC Med Inform Decis Mak 2018 Feb 20;18(1):12 [FREE Full text] [doi: 10.1186/s12911-018-0591-0] [Medline: 29458358]
47. Edwards EA, Lumsden J, Rivas C, Steed L, Edwards LA, Thiyagarajan A, et al. Gamification for health promotion: Systematic review of behaviour change techniques in smartphone apps. BMJ Open 2016 Oct 04;6(10):e012447 [FREE Full text] [doi: 10.1136/bmjopen-2016-012447] [Medline: 27707829]
48. King D, Greaves F, Exeter C, Darzi A. 'Gamification': Influencing health behaviours with games. J R Soc Med 2013 Mar;106(3):76-78 [FREE Full text] [doi: 10.1177/0141076813480996] [Medline: 23481424]
49. Chan YY, Wang P, Rogers L, Tignor N, Zweig M, Hershman SG, et al. The Asthma Mobile Health Study, a large-scale clinical observational study using ResearchKit. Nat Biotechnol 2017 Apr;35(4):354-362 [FREE Full text] [doi: 10.1038/nbt.3826] [Medline: 28288104]
50. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med 2019 Nov 14;381(20):1909-1917. [doi: 10.1056/nejmoa1901183]
51. Chen J, Berkman W, Bardouh M, Ng CYK, Allman-Farinelli M. The use of a food logging app in the naturalistic setting fails to provide accurate measurements of nutrients and poses usability challenges. Nutrition 2019 Jan;57:208-216. [doi: 10.1016/j.nut.2018.05.003] [Medline: 30184514]
52. Bidargaddi N, Almirall D, Murphy S, Nahum-Shani I, Kovalcik M, Pituch T, et al. To prompt or not to prompt? A microrandomized trial of time-varying push notifications to increase proximal engagement with a mobile health app. JMIR Mhealth Uhealth 2018 Nov 29;6(11):e10123 [FREE Full text] [doi: 10.2196/10123] [Medline: 30497999]
53. 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]
54. Wei Y, Zheng P, Deng H, Wang X, Li X, Fu H. Design features for improving mobile health intervention user engagement: Systematic review and thematic analysis. J Med Internet Res 2020 Dec 09;22(12):e21687 [FREE Full text] [doi: 10.2196/21687] [Medline: 33295292]
55. Bender MS, Choi J, Arai S, Paul SM, Gonzalez P, Fukuoka Y. Digital technology ownership, usage, and factors predicting downloading health apps among Caucasian, Filipino, Korean, and Latino Americans: The digital link to health survey. JMIR Mhealth Uhealth 2014 Oct 22;2(4):e43 [FREE Full text] [doi: 10.2196/mhealth.3710] [Medline: 25339246]
56. Robbins R, Krebs P, Jagannathan R, Jean-Louis G, Duncan DT. Health app use among us mobile phone users: Analysis of trends by chronic disease status. JMIR Mhealth Uhealth 2017 Dec 19;5(12):e197 [FREE Full text] [doi: 10.2196/mhealth.7832] [Medline: 29258981]
57. Murray E, White IR, Varagunam M, Godfrey C, Khadjesari Z, McCambridge J. Attrition revisited: Adherence and retention in a web-based alcohol trial. J Med Internet Res 2013 Aug 30;15(8):e162 [FREE Full text] [doi: 10.2196/jmir.2336] [Medline: 23996958]
58. Jamshed H, Beyl R, Della Manna D, Yang E, Ravussin E, Peterson C. Early time-restricted feeding improves 24-hour glucose levels and affects markers of the circadian clock, aging, and autophagy in humans. Nutrients 2019 May 30;11(6):1234 [FREE Full text] [doi: 10.3390/nu11061234] [Medline: 31151228]
59. Arble D, Bass J, Behn C, Butler M, Challet E, Czeisler C, et al. Impact of sleep and circadian disruption on energy balance and diabetes: A summary of workshop discussions. Sleep 2015 Dec 01;38(12):1849-1860 [FREE Full text] [doi: 10.5665/sleep.5226] [Medline: 26564131]

Abbreviations

AMHS: Asthma Mobile Health Study
e-consent: electronic consent
EHR: electronic health record
IRB: Institutional Review Board
mHealth: mobile health
PCORnet: National Patient-Centered Research Network
REDCap: Research Electronic Data Capture

XSL•FO
RenderX

Edited by R Kukafka; submitted 15.10.21; peer-reviewed by B Schreiweis, Y Chu; comments to author 18.12.21; revised version received 01.03.22; accepted 21.03.22; published 10.06.22.

Please cite as:
Coughlin JW, Martin LM, Zhao D, Goheer A, Woolf TB, Holzhauer K, Lehmann HP, Lent MR, McTigue KM, Clark JM, Bennett WL
Electronic Health Record–Based Recruitment and Retention and Mobile Health App Usage: Multisite Cohort Study
J Med Internet Res 2022;24(6):e34191
URL: https://www.jmir.org/2022/6/e34191
doi:10.2196/34191
PMID:35687400

©Janelle W Coughlin, Lindsay M Martin, Di Zhao, Attia Goheer, Thomas B Woolf, Katherine Holzhauer, Harold P Lehmann, Michelle R Lent, Kathleen M McTigue, Jeanne M Clark, Wendy L Bennett. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.06.2022. 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 Remote Digital Monitoring Platform to Assess Cognitive and Motor Symptoms in Huntington Disease: Cross-sectional Validation Study

Florian Lipsmeier[1], PhD; Cedric Simillion[1], PhD; Atieh Bamdadian[1], PhD; Rosanna Tortelli[2], MD; Lauren M Byrne[2], PhD; Yan-Ping Zhang[1], PhD; Detlef Wolf[1], MSc; Anne V Smith[3], PhD; Christian Czech[1,4], PhD; Christian Gossens[1], PhD; Patrick Weydt[5,6], MD; Scott A Schobel[7], MD; Filipe B Rodrigues[2], MD; Edward J Wild[2], PhD; Michael Lindemann[1], PhD

[1]Roche Pharma Research and Early Development, pRED Informatics, Pharmaceutical Sciences, Clinical Pharmacology, and Neuroscience, Ophthalmology, and Rare Diseases Discovery and Translational Area, Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd, Basel, Switzerland

[2]Huntington's Disease Centre, UCL Queen Square Institute of Neurology, University College London, London, United Kingdom

[3]Ionis Pharmaceuticals Inc, Carlsbad, CA, United States

[4]Rare Disease Research Unit, Pfizer, Nice, France

[5]Department of Neurology, University of Ulm Medical Center, Ulm, Germany

[6]Department of Neurodegenerative Disease and Gerontopsychiatry/Neurology, University of Bonn Medical Center, Bonn, Germany

[7]F. Hoffmann-La Roche Ltd, Basel, Switzerland

Corresponding Author:
Florian Lipsmeier, PhD
Roche Pharma Research and Early Development, pRED Informatics, Pharmaceutical Sciences, Clinical Pharmacology, and Neuroscience, Ophthalmology, and Rare Diseases Discovery and Translational Area
Roche Innovation Center Basel
F. Hoffmann-La Roche Ltd
Grenzacherstrasse 124
Basel, CH 4070
Switzerland
Phone: 41 61 6877909
Email: florian.lipsmeier@roche.com

Abstract

Background: Remote monitoring of Huntington disease (HD) signs and symptoms using digital technologies may enhance early clinical diagnosis and tracking of disease progression, guide treatment decisions, and monitor response to disease-modifying agents. Several recent studies in neurodegenerative diseases have demonstrated the feasibility of digital symptom monitoring.

Objective: The aim of this study was to evaluate a novel smartwatch- and smartphone-based digital monitoring platform to remotely monitor signs and symptoms of HD.

Methods: This analysis aimed to determine the feasibility and reliability of the Roche HD Digital Monitoring Platform over a 4-week period and cross-sectional validity over a 2-week interval. Key criteria assessed were feasibility, evaluated by adherence and quality control failure rates; test-retest reliability; known-groups validity; and convergent validity of sensor-based measures with existing clinical measures. Data from 3 studies were used: the predrug screening phase of an open-label extension study evaluating tominersen (NCT03342053) and 2 untreated cohorts—the HD Natural History Study (NCT03664804) and the Digital-HD study. Across these studies, controls (n=20) and individuals with premanifest (n=20) or manifest (n=179) HD completed 6 motor and 2 cognitive tests at home and in the clinic.

Results: Participants in the open-label extension study, the HD Natural History Study, and the Digital-HD study completed 89.95% (1164/1294), 72.01% (2025/2812), and 68.98% (1454/2108) of the active tests, respectively. All sensor-based features showed good to excellent test-retest reliability (intraclass correlation coefficient 0.89-0.98) and generally low quality control failure rates. Good overall convergent validity of sensor-derived features to Unified HD Rating Scale outcomes and good overall known-groups validity among controls, premanifest, and manifest participants were observed. Among participants with manifest

XSL•FO
RenderX

HD, the digital cognitive tests demonstrated the strongest correlations with analogous in-clinic tests (Pearson correlation coefficient 0.79-0.90).

Conclusions: These results show the potential of the HD Digital Monitoring Platform to provide reliable, valid, continuous remote monitoring of HD symptoms, facilitating the evaluation of novel treatments and enhanced clinical monitoring and care for individuals with HD.

(J Med Internet Res 2022;24(6):e32997) doi:10.2196/32997

KEYWORDS

Huntington disease; digital monitoring; digital biomarkers; remote monitoring; smartphone; smartwatch; cognition; motor; clinical trials; mobile phone

Introduction

Background

Huntington disease (HD) is a genetic, neurodegenerative, and ultimately fatal disease characterized by a triad of cognitive, behavioral, and motor symptoms leading to functional decline and progressive loss of independence [1,2]. The clinical assessment of HD primarily relies on periodic in-person clinical assessments and may include administration of clinician-rated outcomes (which are dependent on rater experience and expertise) or patient-reported outcomes [3,4]. The infrequency of these assessments can result in subtle changes in cognition, behavior or motor abilities being unnoticed, and fluctuations in signs and symptoms being undetected [3,5]. Moreover, in-clinic assessments of disease symptoms that affect patients' daily experiences are removed from the daily context in which patients experience these symptoms [3]. Taken together, there is a need for improvement in the monitoring of HD signs, symptoms, and functional impacts to enhance accurate characterization of the clinical course and detection of treatment effects.

Recent studies have demonstrated the feasibility and initial validation of wearable sensors as objective measures of HD motor symptoms in the home setting [6-10]. Acquired sensor data on motor function differentiated individuals with HD from control participants, as well as individuals with HD grouped by motor impairment as measured by the Unified HD Rating Scale-Total Motor Score (UHDRS-TMS) [6]. Additionally, sensor data revealed disease features not observed during in-clinic assessments, such as an increased proportion of time spent lying down among participants with HD who were ambulatory compared with control participants [7,9]. A pilot study of a smartphone app in 23 participants showed a significant difference in chorea score and tap rate between individuals with and without manifest HD [11]. Furthermore, the digital measure of tap rate strongly correlated with the UHDRS finger tapping score [11]. Finally, the feasibility and validity of a smartphone app for remote assessment of HD cognitive measures were evaluated in a study of 42 participants. The study found that the digital cognitive tests had robust test-retest reliability for participants with manifest HD (intraclass correlation coefficient [ICC] 0.71-0.96) [12]. Correlations between the digital cognitive tasks and selected Enroll-HD cognitive tasks varied in strength (r=0.36-0.68) [12]. Despite these promising exploratory findings in HD and favorable results from digital platforms for Parkinson disease [13-21], to date, there have been no formal validation efforts of an at-home digital-based monitoring system that includes assessments for both motor and nonmotor symptoms of HD.

Although most studies on digital measures of neurodegenerative diseases have focused on motor symptom assessments, a more comprehensive assessment of function that includes both motor and nonmotor outcomes is needed to provide a more holistic disease characterization. Furthermore, digital platforms that include a combination of motor and nonmotor active tests, passive monitoring of daily activities, and patient-reported outcomes can generate data that can be interpreted as being meaningful to patients. Indeed, previous studies have demonstrated that gait and balance impairments increase fall risk and greatly influence the quality of life of people with HD [22]. Furthermore, the cognitive and neuropsychiatric characteristics of HD contribute greatly to the loss of functional independence and quality of life, and hence require evaluation [23-26].

This Study

In this study, a smartwatch- and smartphone-based remote digital monitoring platform was developed to assess motor, cognitive, behavioral, and functional domains in HD using frequent active and continuous passive monitoring [27,28]. This platform was applied to individuals with premanifest HD (individuals genetically confirmed to have HD but not having diagnostic motor symptoms of HD), manifest HD (individuals with diagnostic motor symptoms of HD), and control participants to determine its feasibility, reliability, and cross-sectional validity for monitoring motor and cognitive features, which are key domains that change with clinical progression across the continuum of adult HD [1]. Digital-based outcomes were compared at baseline with analogous in-clinic tests during the screening period from 3 independent studies (a recently completed open-label extension [OLE; NCT03342053] of a tominersen phase I/IIa study and 2 untreated natural history cohorts: the HD Natural History Study [NHS; NCT03664804] and the University College London Digital-HD study) to cross-sectionally validate the Roche HD Digital Monitoring Platform, with results reported herein.

Methods

Study Design and Setting

The analysis sample included pretreatment data from the OLE of a tominersen phase I/IIa study and 2 untreated natural history cohorts: HD NHS and the University College London

XSL•FO
RenderX

Digital-HD study. The OLE study was designed to assess the safety, tolerability, pharmacokinetics, and pharmacodynamics of tominersen in patients with manifest HD as well as its effects on digital and standard clinical measures. The OLE study included a 4-week screening period before the start of the treatment period. The HD NHS was designed to evaluate the relationship between changes in cerebrospinal fluid mutant huntingtin protein levels and clinical outcomes in untreated patients with manifest HD. A 4-week screening period was included in the HD NHS. Digital-HD was an observational study that evaluated the tolerability and feasibility of conducting smartphone- and smartwatch-based remote patient monitoring in HD over 18 months.

In this study, we report an analysis of sensor-based outcomes from the 3 studies to assess the cognitive and motor domains. Data from the first 4 weeks after issuing the remote monitoring devices to the participants were considered for the analysis; for the OLE study and the HD NHS, this included only data collected during the 4-week screening periods, up to the baseline assessment. Adherence metrics were collected upon deployment of the digital devices to participants. As part of the study setup for the digital monitoring platform solution, adherence monitoring and processes for follow-up were implemented in case of drops in participant adherence. However, there were no incentives, financial or otherwise, for high adherence, nor did poor adherence lead to exclusion of the participant from the study. The longitudinal effects of tominersen on digital outcomes acquired from the OLE study are not the focus of this study and hence not reported here.

Participants

All participants enrolled in each respective study were eligible for this analysis. Written, informed consent was obtained from all participants. To be eligible for the OLE study, patients must have completed the treatment period of the phase I/IIa study. Patients in the phase I/IIa study had early manifest HD, Shoulson-Fahn stage I disease (UHDRS-Total Functional Capacity [TFC] score 11-13). Participants in the HD NHS had early manifest HD, Shoulson-Fahn stage I/II disease (UHDRS-TFC score 7-13). Participants from the phase I/IIa study (N=46) and the HD NHS (N=94) were aged 25 to 65 years. The Digital-HD study (N=79) enrolled adults (aged 18 to 75 years) with manifest HD (diagnostic confidence level=4, stage I-III, UHDRS-TFC 4-13, cytosine adenine guanine [CAG] expansion≥36), premanifest HD (diagnostic confidence level<4, CAG expansion≥40), and healthy control participants (no known family history or CAG expansion<36).

In-Clinic Assessments

All in-clinic assessments were performed at the screening visit. Clinical signs and symptoms were assessed by trained raters using the UHDRS [29]. The scale assesses 4 domains associated with HD: motor function, cognitive function, behavioral abnormalities, and functional capacity. To assess motor performance, the UHDRS Maximal Chorea item, UHDRS Finger Taps item, and UHDRS-TMS were used. In addition, an in-clinic balance score was generated by summing the UHDRS-TMS Retropulsion Pull test item and UHDRS-TMS Tandem Walking test item scores.

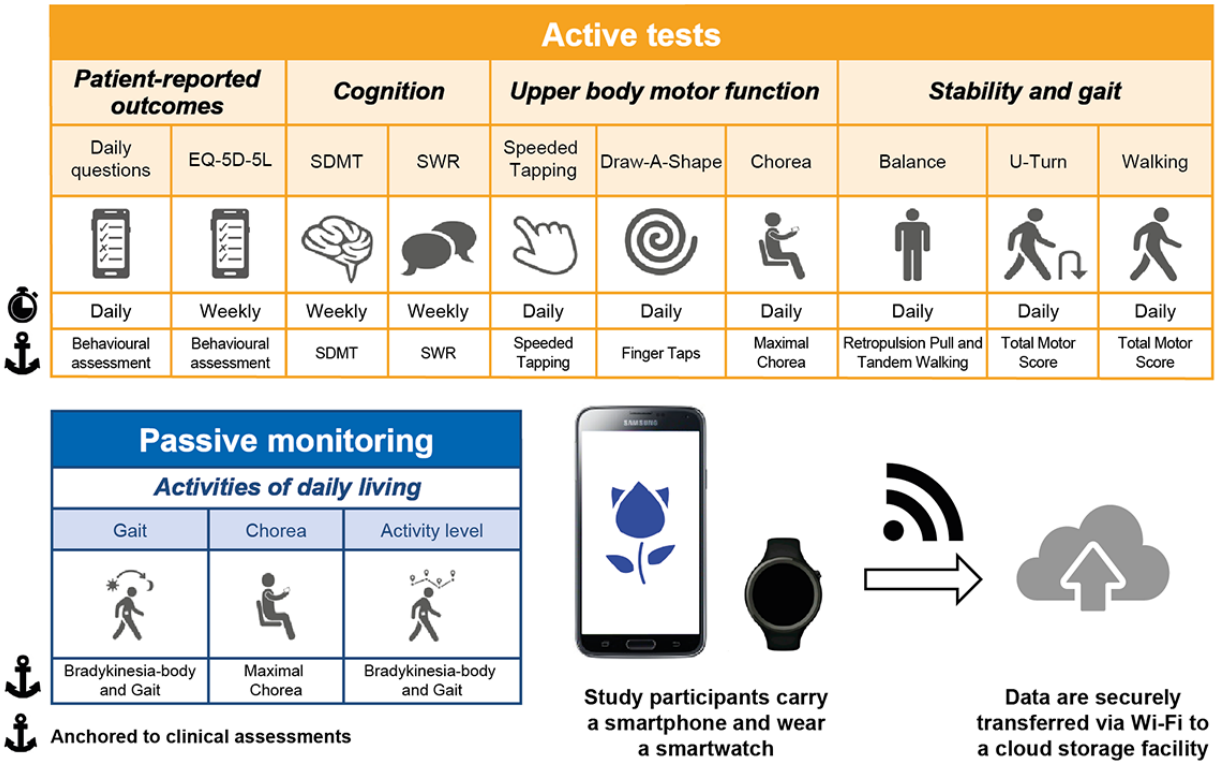
To assess cognitive performance, the Stroop Word Reading (SWR) [30] and the Symbol Digit Modalities Test (SDMT) [31] were used. The SWR is a measure of attention and psychomotor speed and relies on verbal motor output and ability to articulate words. The SDMT was used to assess attention, visuoperceptual processing, working memory, and psychomotor speed. The Speeded Tapping test [32] was applied to measure bradykinesia and motor timing. In this computerized test, participants were instructed to tap on the mouse key as fast as possible for 30 seconds using their index finger. The mouse was fixed on the mouse platform and placed on the table.

Digital Monitoring Hardware

Participants were provided with a wrist-worn smartwatch (Moto G 360 2nd Gen Sport; Motorola), a smartphone (Galaxy J7; Samsung), and a belt containing a pouch to carry the smartphone. Participants received training on their use at the screening visit, at which time the devices were deployed for remote continuous monitoring (Figure 1). The devices were locked and configured to only collect assessments in this study and contained no additional functionality.

XSL•FO
RenderX

Figure 1. Overview of the Roche HD monitoring app and workflow for the daily assessments. The smartphone (Galaxy J7; Samsung) and smartwatch (Moto G 360 2nd Gen Sport; Motorola) were provided with a preinstalled custom app (Roche HD monitoring app version 1; Roche). Participants were instructed to carry the smartphone in their trouser pocket, or a belt containing a pouch around the waist and wear the smartwatch. The app requested the completion of active tests daily and subsequently recorded sensor data during daily living (passive monitoring). EQ-5D-5L: EuroQol 5-dimension 5-level; SDMT: Symbol Digit Modalities Test; SWR: Stroop Word Reading.



Active Digital Tests

A novel Android app (Roche HD monitoring app version 1) was designed and installed on both the smartphone and smartwatch to measure HD clinical features. Participants were asked to complete specific tests using the devices (*active tests*) and then to carry the devices with them as they conducted their daily routine, during which sensor data were recorded continuously for *passive monitoring* of gait, chorea, and activity level. Results from patient-reported outcomes and passive monitoring are not reported here. Table 1 describes the active tests included in the app and a video depicting the tests can be found in Multimedia Appendix 1. Participants were prompted by the smartphone to complete active tests daily except SWR, SDMT, and EuroQol 5-dimension 5-level, which were completed weekly. Each test was preceded by an instruction screen that named and explained the task (Multimedia Appendix 2).

The digital SDMT assessed the participant's attention, visuoperceptual processing, working memory, and psychomotor speed by recording the participant's performance in tapping a number corresponding to a symbol shown on the smartphone screen. A total of 3 sets of symbol-digital mappings were used, and the symbol sequence was fixed for all tests. For a given symbol, participants were required to match it with a 1- to 9-digit using the keypad. The number of correct answers was defined as the number of matching events.

The digital SWR test assessed the participant's attention, psychomotor speed, and ability to articulate words by recording the participant's performance in reading color names out loud, row by row, as fast as possible. Names of colors were displayed in black on the screen in a randomly generated sequence (4 words per row and a total of 60 words). Spoken words were automatically recognized by a custom-written word recognizer. The custom-written word recognizer was validated on >30 annotated Stroop tests per language spanning the whole range of severity as defined by the in-clinic SWR test. The number of correctly read words was defined as the number of matching words.

The Speeded Tapping test assessed fine motor impairment by recording the participant's performance in tapping one button on the screen as fast and as regularly as possible.

The Draw-A-Shape test assessed visuomotor coordination and fine motor impairment by recording the participant's performance in tracing a series of increasingly complex shapes on the smartphone screen.

The Chorea test captured the degree of chorea by recording upper body physical movements as the seated participant held the smartphone as still as possible in their outstretched arm and hand while wearing the smartwatch on the preferred wrist. As a dual task, the participant counted backward aloud during the test; these data are not included in these analyses as the methodologies needed to analyze these data are still under development.

The Balance test assessed the participant's static balance function by recording movements as the participant stood as still as possible while wearing the smartphone and smartwatch.

XSL•FO
RenderX

The U-Turn test was designed to assess gait and lower body bradykinesia. The participant walked and turned at least five times between 2 points that were at least four steps apart while wearing the smartphone and smartwatch.

The Walk test captured elements of gait, body bradykinesia, and tandem walking abnormalities. The participant walked as fast as was safely possible for 200 meters or 2 minutes.

The inertial measurement unit (accelerometer, gyroscope, or magnetometer) captured continuous measurements from the smartphone and smartwatch during active tests. The digital SWR and Chorea tests were captured using the microphone in addition to inertial measurement unit recordings. For the digital SDMT, Speeded Tapping, and Draw-A-Shape tests, touchscreen events were recorded. For the digital SDMT, actual answers with timestamps were recorded. Participants were instructed to carry the phone in the provided pouch for the Balance, U-Turn, and Walking tests and in the provided pouch or their trouser pocket for passive monitoring.

Table 1. Descriptions of the active tests included in the Roche HD monitoring app version 1.

Domain and test	Description
Cognition	
SDMT[a]	Digital version of the pen-and-paper SDMT; tap the number corresponding to a symbol shown on the screen as fast as possible for 90 seconds
SWR[b]	Modified digital version of the pen-and-paper SWR; read the color names out loud, row by row, as fast as possible for 45 seconds
Upper body motor function	
Speeded Tapping[c]	Repeatedly tap a virtual button as fast as possible for 30 seconds with the phone flat on a surface
Draw-A-Shape[c]	Trace a series of reference shapes (diagonal lines, square, circle, figure of eight, or spiral) on the screen with the index finger as quickly and accurately as possible with the phone flat on a surface
Chorea[c]	Hold phone in the palm of the hand and keep arm outstretched for 30 seconds while keeping eyes closed and counting backward aloud in sevens from a random number shown on the smartphone screen
Stability and gait	
Balance	Stand upright as still as possible for 30 seconds with arms hanging loosely by the sides and phone in waist pouch
U-Turn	Walk between two points, at least four steps apart, and turn 180 degrees at least five times with the phone in waist pouch during a 60-second period
Walking	Walk as fast as safely possible for 200 meters or 2 minutes, with phone in waist pouch. Ideally, the test was performed in a straight path with no obstacles

[a]SDMT: Symbol Digit Modalities Test.

[b]SWR: Stroop Word Reading.

[c]Tests are repeated for each hand.

Data Transfer and Processing

Participants received instructions on how to connect the smartphone to the internet at home. For participants with no wireless internet connection (Wi-Fi) at home, data were uploaded during clinic visits. All data were encrypted and uploaded to secure servers each time the smartphone was connected to Wi-Fi.

Digital Test Outcomes

The raw data for each test were converted into a single predefined readout, hereafter referred to as a feature. The values reported here are the medians for each feature over 2-week intervals. If, for a participant, less than n observations that passed the quality criteria (*Statistical Analysis*) for a given test in an interval were available, the data for that participant and interval were considered as missing. The value of n was 1 for the SDMT and SWR test and 3 for all other tests. The interval length of 2 weeks was found to be the optimal trade-off between period length and robustness against missing values. The following active test features were prospectively selected based on their face validity as tests of relevant cognitive and motor function in HD and digital surrogates of existing in-clinic tests: number of correct answers (SDMT), number of correctly read words (SWR), and mean intertap interval (Speeded Tapping). The rationale behind the intertap interval is to assess the time that the finger is in the air (not on the glass), as we hypothesized that uncontrolled movements would influence this time span. For all other tests, features were prospectively chosen based on previous literature and their relevance to HD: sway path (Chorea and Balance) [33,34], spiral drawing speed variability (Draw-A-Shape), median turn speed (U-Turn) and step frequency variance (Walking) [35]. The sway path feature offers a straightforward way to measure the amount of movement occurring when a study participant is trying to hold the body or hand as still as possible; this feature has been successfully used in other disease areas in the same context [36]. The median turn speed feature has shown good performance in Parkinson disease and multiple sclerosis [14,37]. This feature is influenced by gait and postural instability problems, which are both prevalent symptoms in HD. Therefore, it was hypothesized that turn speed would measure relevant HD signals. A meta-analysis showed that variability in gait parameters was increased among patients with HD compared with healthy controls, even in comparison

XSL•FO
RenderX

with patients with other neurological disorders [38]. These results and expert input led to the decision to select step frequency variance as a feature to measure gait variability while being algorithmically as robust as possible. Variability measures, in general, seem to be sensitive in detecting disease-relevant signals in the upper limb domain also, as has been shown for a tapping test [39]. Although the Draw-A-Shape test offers a plethora of different features to select from, a feature that measures variability, drawing speed variability (as measured by the coefficient of variation of drawing speed, for the shape where we expected to see the biggest challenges in maintaining drawing speed, ie, the spiral) was chosen.

Statistical Analysis

As a quality control (QC) measure, active tests were excluded via quality criteria assessing the correct test execution (Multimedia Appendix 3). Overall adherence is reported as the proportion of tests completed during the 4-week study period.

Test-retest reliability of active test feature data from participants was calculated using the ICC [40] between the median values of the first 2 weeks and those of the second 2 weeks, and occurred predrug exposure in the OLE study.

To investigate convergent validity (ie, the degree to which 2 measures of the same construct are related) of sensor-derived features, Spearman correlation was calculated between clinical scores acquired at baseline visit and sensor-derived features that were median-aggregated over the first 2 weeks of data collection. Pearson correlation was used when both variables were normally distributed and a linear relationship between them was expected to exist. All analyses were conducted with Python (version 3.6; Python Software Foundation) scripts using the pandas and SciPy libraries.

Known-groups validity was assessed by comparing the feature values between the 3 cohorts from the Digital-HD study (premanifest HD, manifest HD, and controls) and the 2 manifest HD cohorts from the OLE study and HD NHS. Comparison was done by first fitting for each value a mixed linear effect model with a fixed effect for age and a random intercept for study. The residuals of this model were then compared using Kruskal-Wallis test and pairwise Mann-Whitney *U* tests between these 5 groups.

Ethics Approval

All studies were approved by the respective local ethics committees and conducted in accordance with the Declaration of Helsinki and the International Conference on Harmonisation Guidelines for Good Clinical Practice.

The OLE study protocol was approved by the following ethics committees: National Research Ethics Service Committee London-West London and GTAC, London, United Kingdom; Ethik-Kommission der Medizinischen Fakultät der Universität Ulm, Germany; and University of British Columbia Clinical Ethics Review Board, Canada. The HD NHS protocol was approved by the following ethics committees: The University of British Columbia Clinical Research Ethics Office, Vancouver, British Columbia, Canada; Advarra, Aurora, Ontario, Canada; Universität Ulm – Ethik - Kommission, Germany; Ethik - Kommission der Med. Fakultät der Ruhr Universität Bochum, Germany; London - Camden & Kings Cross Research Ethics Committee, London, United Kingdom; Advarra, Columbia, Maryland, United States; University of California Davis Institutional Review Board (IRB) Administration, Sacramento, California, United States; Institutional Review Board, Columbia University Medical Center, New York, New York, United States; Johns Hopkins Medicine Office of Human Subjects Research-IRB East Baltimore Campus, Baltimore, Maryland, United States; HealthOne IRB, Denver, Colorado, United States; Georgetown University Institutional Review Board, Washington DC, United States; and NHS Health Research Authority/Research Ethics Service, London, United Kingdom. The Digital-HD study protocol was approved by London-Central Research Ethics Committee, London, United Kingdom.

Code Availability

To collect at-home and in-clinic digital data (active tests of cognitive and motor performance, passive monitoring of daily life, and electronic patient-reported outcomes), we relied on custom Android apps built specifically for these studies deployed on smartphones and smartwatches. All other data for the studies (including demographics and clinical scores) were collected manually without the use of software code.

Raw data signal processing (feature extraction) and data analysis (to compute descriptive statistics for demographic variables, implement test-retest reliability, and for group comparisons and correlations) were carried out with Python. The code used to complete the analysis can be made available upon request.

Results

Participants

In the OLE study and HD NHS, no participants were lost to follow-up or withdrew during the 4-week period after receiving the digital monitoring equipment, allowing all to be included in the planned test-retest evaluation. In the Digital-HD study, one participant withdrew from the study. Participants' key baseline demographic and clinical characteristics are provided in Table 2.

XSL•FO
RenderX

Table 2. Baseline characteristics of participants in the open-label extension (OLE) study, the Huntington disease Natural History Study (HD NHS), and the Digital-HD study.

Characteristics	OLE study (N=46)	HD NHS (N=94)	Digital-HD study		
			Healthy control (N=20)	Premanifest HD (N=20)	Manifest HD (N=39)
Age (years), mean (SD)	48.6 (10.2)	48.2 (9.9)	48.0 (13.8)	44.9 (10.0)	56.3 (11.0)
Gender, n (%)					
Male	28 (61)	58 (62)	13 (65)	10 (50)	21 (54)
Female	18 (39)	36 (38)	7 (35)	10 (50)	18 (46)
Number of CAG[a] repeats, mean (SD)	44.3 (3.0)	44.2 (3.1)	N/A[b]	41.6 (2.0)	42.7 (3.3)[c]
Right hand dominance (laterality), n (%)	43 (93)	81 (86)	17 (85)	20 (100)	29 (74)
TMS[d], mean (SD)	23.6 (12.5)	22.1 (10.9)	1.4 (2.4)	4.9 (3.9)	32.9 (16.6)
TFC[e], mean (SD)	11.2 (1.6)	11.0 (1.5)	13.0 (0.0)	12.9 (0.3)	10.6 (2.2)
SWR[f], mean (SD)	74.0 (21.9)	72.2 (20.0)	100.1 (19.5)	102.3 (19.0)	67.5 (19.4)
SDMT[g], mean (SD)	33.7 (12.1)	32.3 (11.6)	62.1 (8.4)[h]	56.4 (12.2)[h]	29.0 (12.2)[h]

[a]CAG: cytosine adenine guanine.

[b]N/A: not applicable.

[c]Number of CAG repeats for 3 participants with manifest HD in Digital-HD study not available.

[d]TMS: Total Motor Score.

[e]TFC: Total Functional Capacity.

[f]SWR: Stroop Word Reading.

[g]SDMT: Symbol Digit Modalities Test.

[h]Reported data are an average of 68 participants, as data from 11 participants were discarded owing to tests conducted in 45 seconds rather than 90 seconds.

QC of Digital Active Test Execution

For the SDMT, Chorea, Speeded Tapping, and SWR tests, a low proportion (ranging between 0% and 6.7% [Digital-HD, healthy controls: 16/238 (6.7%) Chorea—nondominant hand—tests were excluded]) of digital active tests across the 3 studies were excluded from the analysis due to QC criteria not being met, indicating improper test execution (Multimedia Appendix 4). For the other tests, the QC fail rates were higher. The percentage of improperly executed Draw-A-Shape tests was higher among participants with manifest HD across the 3 studies in comparison with participants with premanifest HD and control participants in the Digital-HD study. Across the 3 studies, 8.60% (HD NHS: 94/1093) to 10% (OLE study: 33/331) and 13.05% (HD NHS: 140/1073) to 20.1% (OLE study: 66/329) of Draw-A-Shape tests performed by participants with manifest HD with the dominant (D) and nondominant (ND) hands, respectively, failed to pass the QC criteria. Moreover, an analysis of the per-subject QC pass rate for the Draw-A-Shape tests showed that this rate is negatively correlated with UHDRS-TMS and Maximal Chorea upper limb scores (Multimedia Appendix 5). Study participants were instructed to carry out the Walking, U-Turn, and Balance tests with the phone placed in the provided pouch around the waist. For the Walking test, it was found that not adhering to this instruction and carrying the device in the pocket instead resulted in skewed step frequency variance values. As a result, these test instances were discarded, which amounted to 17.5% (Digital-HD, premanifest HD: 40/228) to 30.9% (OLE study: 84/272) of all Walking tests. To be consistent, the same criterion was also applied to the Balance and the U-Turn tests, resulting in similar percentages of tests being discarded. It should be noted though that despite the relatively high number of discarded tests, only 11% (19/172), 12.8% (23/179), and 13.8% (25/181) of all participants were lost for subsequent analysis for the Walking, U-Turn, and Balance tests, respectively.

Adherence

The active tests, excluding the Walking test, required on average (median) 8 to 9 minutes (OLE: 8:49 minutes, HD NHS: 8:59 minutes, and Digital-HD: 8:18 minutes) for the days without the SWR, the SDMT, and the EuroQol 5-dimension 5-level, and 14 to 15 minutes (OLE: 14:52 minutes and HD NHS: 14:18 minutes) for days with these nondaily tests. For Digital-HD, the nondaily tests were split over multiple days, leading to an average test time of 9 to 11 minutes for these days. In the OLE study, participants completed 1164 out of 1294 active tests (89.95%). Participants in the HD NHS and Digital-HD study performed a total of 2025 out of 2812 (72.01%) and 1454 out of 2108 (68.98%) tests, respectively.

Test-Retest Reliability

Test-retest reliability was good to excellent for the active tests across the 3 studies and varied from 0.89 (95% CI 0.83-0.93) to 0.98 (95% CI 0.97-0.99; Table 3).

XSL•FO
RenderX

Table 3. Test-retest reliability intraclass correlation coefficients (ICCs) for the tests and selected face valid features.

Test	Clinical score	Digital test feature		Test-retest ICC of digital test		
			D/ND[a] hand	OLE[b] study	HD[c] NHS[d]	Digital-HD
SDMT[e]	Number of correct answers for in-clinic SDMT	Number of correct answers (95% CI)	N/A[f]	0.94 (0.77-0.98)	0.93 (0.65-0.97)	0.98 (0.89-0.99)
SWR[g]	Number of correctly read words for in-clinic SWR	Number of correctly read words (95% CI)	N/A	0.92 (0.85-0.95)	0.93 (0.90-0.96)	0.96 (0.94-0.98)
Speeded Tapping	Mean intertap interval for in-clinic Speeded Tapping	Mean intertap interval (ms; 95% CI)	D[h]	0.94 (0.86-0.98)	0.97 (0.95-0.98)	0.98 (0.97-0.99)
Speeded Tapping	Mean intertap interval for in-clinic Speeded Tapping	Mean intertap interval (ms; 95% CI)	ND[i]	0.96 (0.88-0.98)	0.97 (0.95-0.98)	0.97 (0.96-0.98)
Draw-A-Shape	UHDRS[j] Finger Taps	Spiral drawing speed variability (mm/s; 95% CI)	D	0.93 (0.85-0.97)	0.93 (0.87-0.96)	0.91 (0.85-0.95)
Draw-A-Shape	UHDRS Finger Taps	Spiral drawing speed variability (mm/s; 95% CI)	ND	0.93 (0.84-0.97)	0.92 (0.87-0.95)	0.97 (0.94-0.98)
Chorea	UHDRS Maximal Chorea upper limb	Sway path (m/s^2; 95% CI)	D	0.96 (0.92-0.98)	0.96 (0.94-0.97)	0.98 (0.96-0.99)
Chorea	UHDRS Maximal Chorea upper limb	Sway path (m/s^2; 95% CI)	ND	0.97 (0.94-0.99)	0.94 (0.90-0.96)	0.98 (0.97-0.99)
Balance	Balance score	Sway path (m/s^2; 95% CI)	N/A	0.91 (0.73-0.97)	0.89 (0.83-0.93)	0.94 (0.89-0.97)
U-Turn	TMS[k]	Median turn speed (rad/sec; 95% CI)	N/A	0.95 (0.89-0.98)	0.95 (0.92-0.97)	0.94 (0.91-0.97)
Walking	TMS	Step frequency variance (Hz^2; 95% CI)	N/A	0.95 (0.88-0.98)	0.93 (0.89-0.96)	0.95 (0.82-0.97)

[a]D/ND: dominant/nondominant.
[b]OLE: open-label extension.
[c]HD: Huntington disease.
[d]NHS: Natural History Study.
[e]SDMT: Symbol Digit Modalities Test.
[f]N/A: not applicable.
[g]SWR: Stroop Word Reading.
[h]D: dominant.
[i]ND: nondominant.
[j]UHDRS: Unified HD Rating Scale.
[k]TMS: Total Motor Score.

Clinical Cross-sectional Validity

Convergent Validity of Sensor-Based Measures With Standard Clinical Outcome Measures

Across the 3 studies for participants with manifest HD, the digital SDMT and SWR tests were strongly associated with the in-clinic SDMT (OLE: r=0.85, HD NHS: r=0.79, Digital-HD [manifest HD cohort]: r=0.80; P<.001 for all) and SWR (OLE: r=0.84, HD NHS: r=0.87, Digital-HD [manifest HD cohort]: 0.90; P<.001 for all tests), respectively (Figures 2A and 2B, Table 4, and Multimedia Appendix 6).

For remote monitoring of upper body motor function, 3 active tests were used: Chorea test, Speeded Tapping test, and Draw-A-Shape test. In the OLE study, the digital Speeded Tapping test was strongly associated with the in-clinic Speeded Tapping test (D: r=0.70, ND: r=0.75; P<.001 for both; Figure 2C, Table 4, and Multimedia Appendix 6). The in-clinic Speeded Tapping test was not conducted in the HD NHS and the Digital-HD study. The spiral drawing speed variability showed moderate association with the UHDRS Finger Taps item across the 3 studies for participants with manifest HD when using the ND hand (OLE: ρ=0.47, P=.001; HD NHS: ρ=0.47, P<.001; Digital-HD [manifest HD cohort]: ρ=0.57, P<.001; Figure 3A, Table 4, and Multimedia Appendix 6). When using the D hand, the spiral drawing speed variability showed moderate association with the UHDRS Finger Taps item in the HD NHS (ρ=0.41; P<.001) and Digital-HD study (manifest HD cohort: ρ=0.55; P=.002). Sway path during Chorea tests showed moderate-to-strong associations across the studies, bilaterally with the UHDRS Maximal Chorea upper limb item (OLE: D/ND, ρ=0.50/0.58, HD NHS: D/ND, ρ=0.46/0.45, Digital-HD [manifest HD cohort]: D/ND, ρ=0.47/0.65; P<.001 for all except

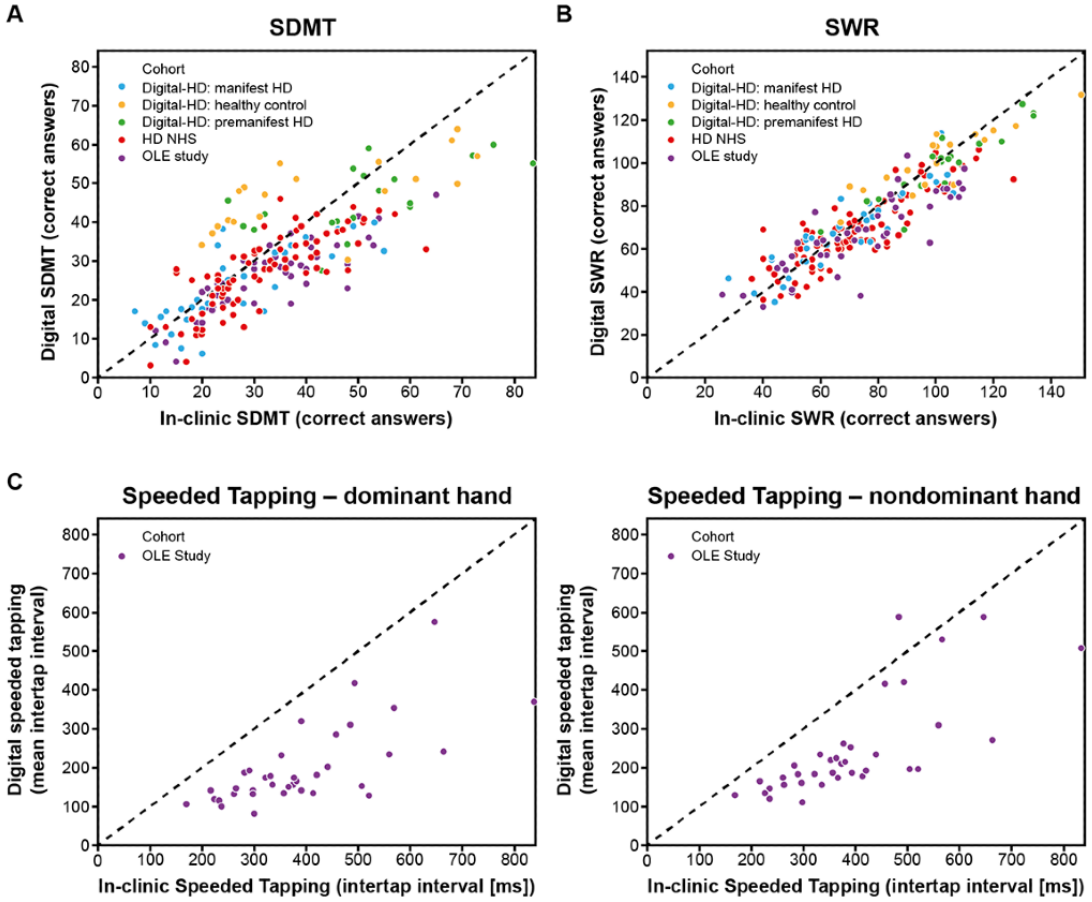
XSL•FO
RenderX

P=.006 for Digital-HD: D; Figure 3B, Table 4, and Multimedia Appendix 6).

For whole body and lower limb tests, significant weak-to-moderate associations were found with respective UHDRS items across the 3 studies for participants with manifest HD with the exception of median turn speed during the U-Turn test in the HD NHS and the Digital-HD study, and sway path during the Balance test in the Digital-HD study (Figure 4, Table 4, and Multimedia Appendix 6): sway path during the Balance test was associated with the in-clinic balance score (OLE: ρ=0.51, HD NHS: ρ=0.28, Digital-HD [manifest HD cohort]: ρ=0.24; *P*=.23), median turn speed when doing U-turns with UHDRS-TMS (OLE: ρ=–0.51, HD NHS: ρ=–0.16; *P*=.18, Digital-HD [manifest HD cohort]: ρ=–0.20; *P*=.32), and step frequency variance while walking for 2 minutes with UHDRS-TMS (OLE: ρ=0.71, HD NHS: ρ=0.26, and Digital-HD [manifest HD cohort]: ρ=0.47).

Using data from participants with premanifest HD in the Digital-HD study, the digital SDMT and SWR tests were strongly associated with the in-clinic SDMT (*r*=0.64; *P*=.002) and SWR (*r*=0.91; *P*<.001) tests, and sway path during the Chorea test showed moderate association with the UHDRS Chorea item when using the D hand (*r*=0.58; *P*=.01).

Figure 2. Clinical validity of digital cognitive tests and digital Speeded Tapping test. (A) Correlation of in-clinic Symbol Digit Modalities Test (SDMT) with digital SDMT. (B) Correlation of in-clinic Stroop Word Reading (SWR) test with digital SWR test. (C) Correlation of in-clinic Speeded Tapping test with digital Speeded Tapping test. HD: Huntington disease; NHS: Natural History Study; OLE: open-label extension.



XSL•FO
RenderX

Table 4. Correlation coefficients between clinical scores and digital tests.

Test	Clinical score	Digital test feature		Correlation coefficient between clinical score and digital test				
				OLE[a] study	HD[b] NHS[c]	Digital-HD study		
			D/N[d] hand			Healthy controls	Premanifest HD	Manifest HD
SDMT[e]	Number of correct answers for in-clinic SDMT	Number of correct answers (95% CI)	N/A[f]	0.85 (0.73 to 0.91)[g,h]	0.79 (0.69 to 0.86)[g,h]	0.68 (0.35 to 0.86)[g,h]	0.64 (0.28 to 0.84)[h,i]	0.80 (0.65 to 0.89)[g,h]
SWR[j]	Number of correctly read words for in-clinic SWR	Number of correctly read words (95% CI)	N/A	0.84 (0.72 to 0.91)[g,h]	0.87 (0.80 to 0.91) [g,h]	0.87 (0.69 to 0.95) [g,h]	0.91 (0.79 to 0.97)[g,h]	0.90 (0.82 to 0.95)[g,h]
Speeded Tapping	Mean intertap interval for in-clinic Speeded Tapping	Mean intertap interval (ms; 95% CI)	D[k]	0.70 (0.49 to 0.84)[g,h]	—[l]	—[l]	—[l]	—[l]
Speeded Tapping	Mean intertap interval for in-clinic Speeded Tapping	Mean intertap interval (ms; 95% CI)	ND[m]	0.75 (0.56 to 0.86)[g,h]	—[l]	—[l]	—[l]	—[l]
Draw-A-Shape	UHDRS[n] Finger Taps	Spiral drawing speed variability (mm/s; 95% CI)	D	0.19 (−0.12 to 0.47)	0.41 (0.21 to 0.58)[g]	0.13 (−0.35 to 0.55)	0.02 (−0.44 to 0.47)	0.55 (0.23 to 0.76)[i]
Draw-A-Shape	UHDRS Finger Taps	Spiral drawing speed variability (mm/s; 95% CI)	ND	0.47 (0.20 to 0.68)[i]	0.47 (0.27 to 0.62)[g]	0.21 (−0.28 to 0.62)	−0.17 (−0.58 to 0.31)	0.57 (0.29 to 0.77)[g]
Chorea	UHDRS Maximal Chorea upper limb	Sway path (m/s^2; 95% CI)	D	0.50 (0.23 to 0.70)[g]	0.46 (0.27 to 0.62)[g]	−0.06 (−0.50 to 0.41)	0.58 (0.17 to 0.82)[i]	0.47 (0.15 to 0.70)[i]
Chorea	UHDRS Maximal Chorea upper limb	Sway path (m/s^2; 95% CI)	ND	0.58 (0.34 to 0.75)[g]	0.45 (0.25 to 0.61)[g]	−0.26 (−0.64 to 0.22)	0.27 (−0.22 to 0.66)	0.65 (0.40 to 0.81)[g]
Balance	Balance score	Sway path (m/s^2; 95% CI)		0.51 (0.05 to 0.79)[o]	0.28 (0.05 to 0.48)[o]	−0.20 (−0.62 to 0.30)	—[p]	0.24 (−0.16 to 0.56)
U-Turn	TMS[q]	Median turn speed (rad/sec; 95% CI)		−0.51 (−0.77 to −0.09)[o]	−0.16 (−0.38 to −0.07)	−0.19 (−0.61 to 0.32)	−0.22 (−0.64 to 0.32)	−0.20 (−0.55 to 0.20)
Walking	TMS	Step frequency variance (Hz^2; 95% CI)		0.71 (0.42 to 0.87)[g]	0.26 (0.02 to 0.47)[o]	0.32 (−0.21 to 0.70)	0.05 (−0.44 to 0.52)	0.47 (0.06 to 0.72)[o]

[a]OLE: open-label extension.

[b]HD: Huntington disease.

[c]NHS: Natural History Study.

[d]D/ND: dominant/nondominant.

[e]SDMT: Symbol Digit Modalities Test.

[f]N/A: not applicable.

[g]$P<.001$.

[h]Indicates Pearson correlation coefficient; Spearman correlation coefficients are used otherwise.

[i]$P<.01$.

[j]SWR: Stroop Word Reading.

[k]D: dominant.

[l]The in-clinic Speeded Tapping test was not conducted in the HD NHS and Digital-HD study.

[m]ND: nondominant.

[n]UHDRS: Unified HD Rating Scale.

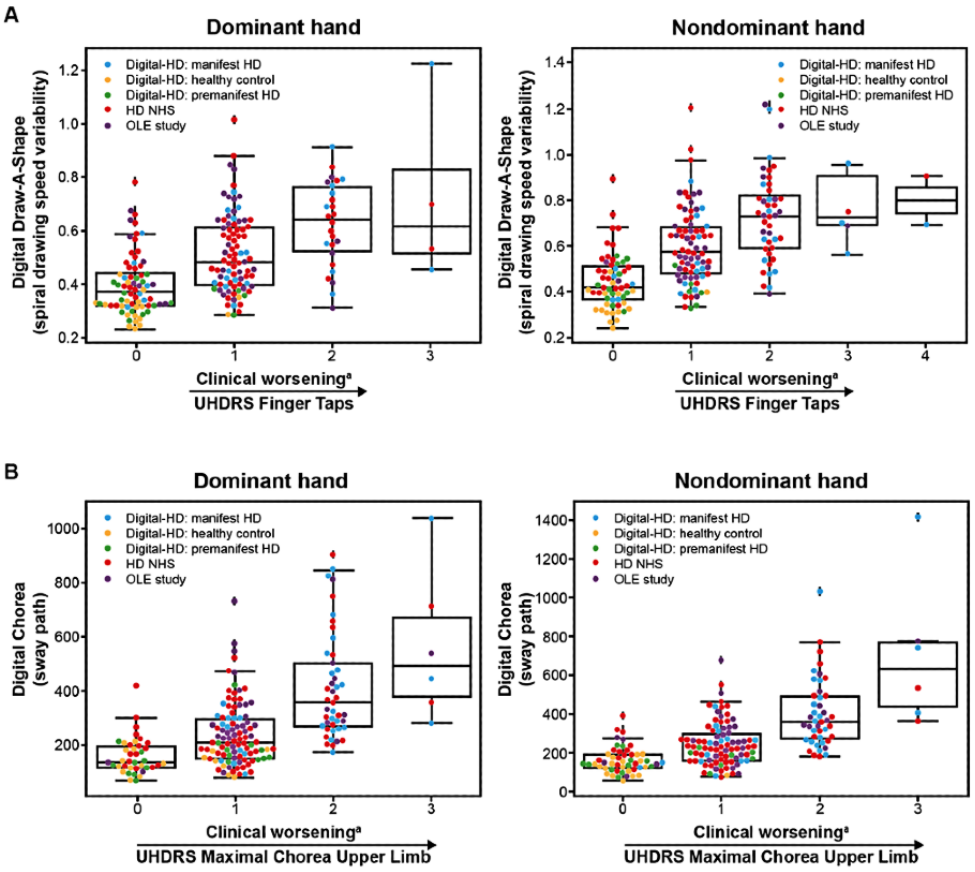
[o]$P<.05$.

[p]Data not available.

[q]TMS: Total Motor Score.

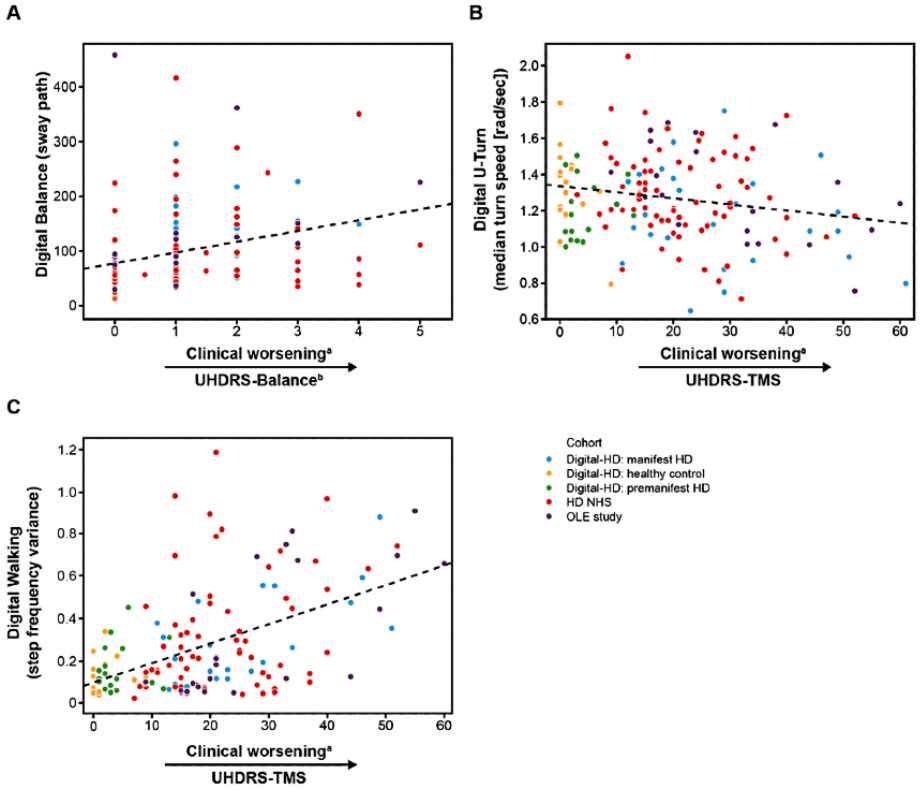
XSL•FO
RenderX

Figure 3. Clinical validity of Draw-A-Shape and digital Chorea tests. (A) Association of Unified HD Rating Scale (UHDRS) Finger Taps with spiral drawing speed variability during the Draw-A-Shape test. (B) Association of the UHDRS Maximal Chorea upper limb item with the log sway path measured during the digital Chorea test. [a]Higher scores represent increased clinical worsening. HD: Huntington disease; NHS: Natural History Study; OLE: open-label extension.



XSL•FO
RenderX

Figure 4. Clinical validity of digital whole body and lower limb tests. (A) Comparison of the balance score with the logarithm of the sway path based on smartphone signal while standing still. (B) Comparison of Unified HD Rating Scale-Total Motor Score (UHDRS-TMS) with the median turn speed during the U-Turn test. (C) Comparison of the UHDRS-TMS with the step frequency variance during the Walking test. The dotted line in (B) and (C) shows the regression line. [a]Higher scores represent increased clinical worsening. [b]Balance is the sum of UHDRS-TMS Retropulsion Pull test item and UHDRS-TMS Tandem Walking test item scores. HD: Huntington disease; NHS: Natural History Study; OLE: open-label extension.

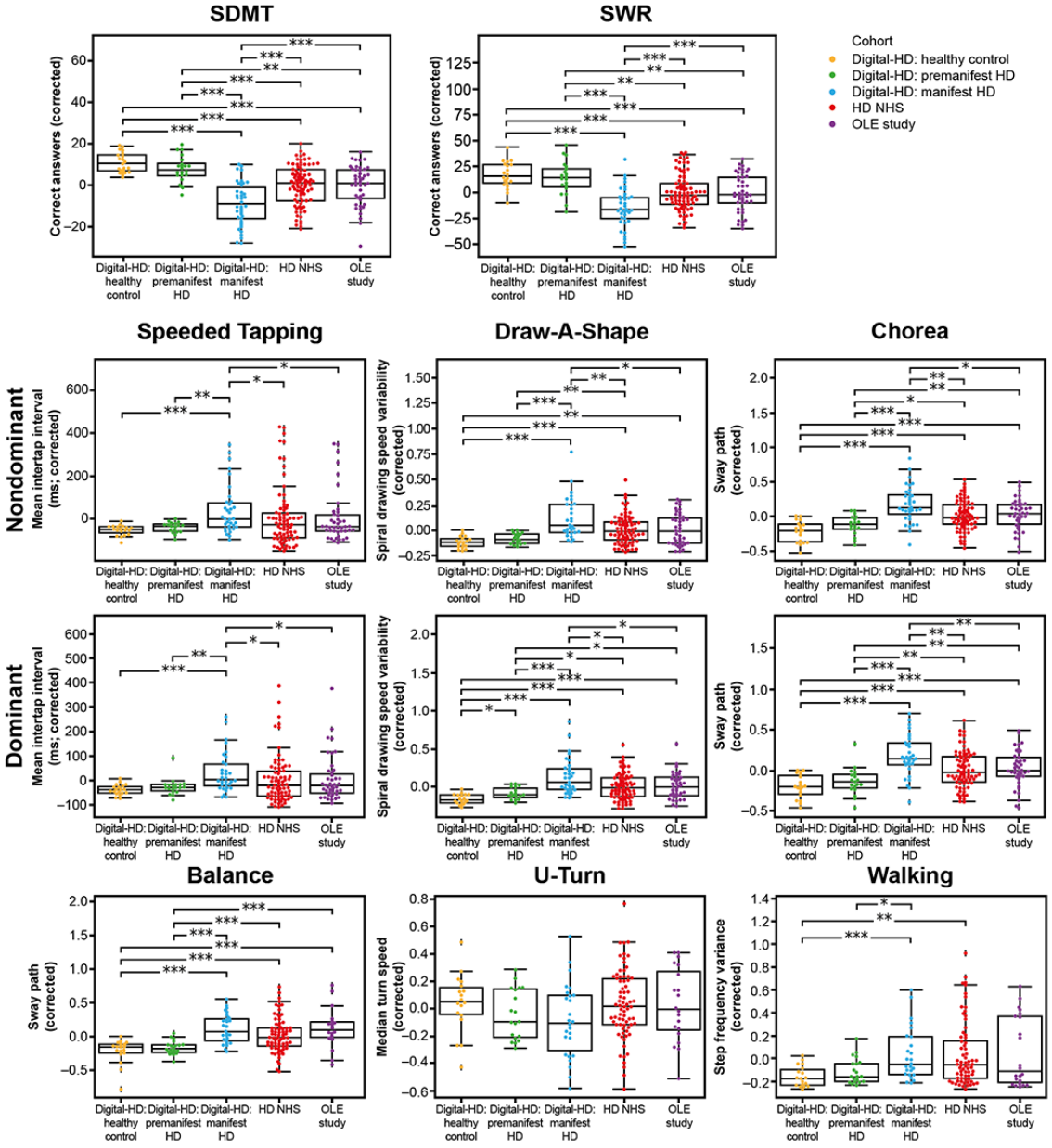


Known-Groups Validity of Sensor-Based Measures

All 3 studies were used to determine the cross-sectional association of each of the 6 motor features and 2 cognitive features to disease status (premanifest HD vs manifest HD vs controls). The results demonstrated that, across features, there was an increasing pattern of abnormality as a function of disease stage (controls vs manifest HD and premanifest HD vs manifest HD), indicating that the digital motor and cognitive features are disease-status associated (Figure 5 and Multimedia Appendix 7). The results within disease stage from the Digital-HD study (manifest HD cohort), OLE study, and HD NHS showed a consistent level of abnormality across the features assessed in the Balance, U-Turn, and Walking tests.

XSL•FO
RenderX

Figure 5. Known-groups validity of digital active tests. All 3 studies were used to compare sensor-derived feature values between the control, premanifest HD, and manifest HD groups to determine an association between digital feature value and disease status or stage. *P<.05; **P<.01; ***P<.001. HD: Huntington disease; NHS: Natural History Study; OLE: open-label extension; SDMT: Symbol Digit Modalities Test; SWR: Stroop Word Reading.



Discussion

Principal Findings

Continuous remote digital assessment of motor and cognitive features of HD appears feasible, reliable, and valid in cross-section across 3 independent HD cohorts. The selected features are robustly associated with disease stage, as well as clinical severity, as measured by standard in-clinic assessments. Overall adherence to the active tests was good to excellent across the 3 studies, demonstrating that the length and number of daily tests were acceptable to most participants.

Interpretation and Comparison With Prior Work

Both cognitive and motor features, except for the turn speed during the U-Turn test in the HD NHS and Digital-HD study, sway path during the Balance test in the Digital-HD study, and spiral drawing speed variability during the Draw-A-Shape test in the OLE study, were significantly correlated with corresponding in-clinic assessments. Good overall convergent validity suggests that the tests measure the same domains as the in-clinic counterparts. Of all digital tests, the cognitive tests demonstrated the strongest correlations with analogous in-clinic tests. Importantly, both digital and standard cognitive outcomes studied here are based upon the same underlying pseudocontinuous scale, which may explain the high degree of association observed. Indeed, a prior and much smaller study of 4 smartphone-based cognitive assessments in participants with HD showed varying degrees of association between the digital cognitive measures and Enroll-HD cognitive tasks, which included SWR and SDMT (Pearson correlation coefficients 0.36-0.68) [12]. Overall, 1 of the 4 digital cognitive assessments showed no significant correlations with Enroll-HD cognitive tasks [12], indicating that this digital assessment may not be

XSL•FO
RenderX

measuring the same constructs that the Enroll-HD cognitive tasks are measuring.

In this study, digital motor tests, though analogous to the in-clinic motor assessment in the UHDRS, are quantitative measures that may offer advantages for improved objectivity and sensitivity, whereas the UHDRS scores are based on Likert-type clinician-rated scales. Notably, sway path during digital Chorea tests showed moderate-to-strong associations across the studies, bilaterally with the UHDRS Maximal Chorea upper limb item (Spearman correlation coefficients 0.45-0.65). These results are consistent with a pilot study of a smartphone app for HD that showed a positive correlation between a digital assessment of chorea and the clinical UHDRS Maximal Chorea score (r=0.53 for the left hand and r=0.54 for the right hand), although the effect was not found to be statistically significant in this small study of 8 participants with manifest HD [11].

Both cognitive features and 2 of the 6 motor features, Speeded Tapping and Chorea, had a low proportion of improperly executed tests (eg, ≤8%). The relatively high amount of improperly executed, and therefore nonevaluable, tests for the Draw-A-Shape task could be owing to how the test was implemented. The attempt to draw a shape is considered as completed as soon as the participant lifts a finger from the screen. This implementation detail could explain why participants with manifest HD had a higher proportion of improperly executed tests. Further research may indicate if changing the implementation to be more tolerant to lifting the finger from the screen will result in a higher proportion of correctly executed tests. The high proportion of failed tests observed with the Balance, U-Turn, and Walking tests was due to the exclusion of data that were collected from participants who performed the tests with the smartphone in their trouser pocket instead of the provided running belt; these data were excluded to account for any possible influence of sensor placement differences on the digital measures.

All sensor-based features had excellent test-retest reliability (ICCs≥0.8). Generally low QC failure, high reliability, and good adherence indicate that these measures possess the properties required to be used as outcome measures in clinical trials. The difference in adherence overall between studies is likely due to the difference in study design, where participants in an interventional trial, in this case the OLE study, are more likely to be motivated compared with participants in observational trials, as with the HD NHS and Digital-HD study. Furthermore, although adherence to the active tests was acceptable across the 3 studies over the 4-week study period, the digital monitoring platform should be evaluated over a longer period to further assess feasibility.

Of note, the digital Speeded Tapping test showed a shift toward shorter mean intertap intervals relative to the in-clinic analog, a shift that most likely reflects a systematic difference between the different devices and platforms used. In some cases, the selected digital features (eg, speed variability of the Draw-A-Shape test) could not map directly to the in-clinic analog (eg, UHDRS Finger Taps item), which may in part explain the lower degree of association between the measures. However, novel assessment of fine motor skills has the potential to detect small changes in motor function that may not otherwise be detected by traditional in-clinic assessments, as supported by the ability of the Draw-A-Shape test to differentiate participants with premanifest HD from controls.

Strengths and Limitations

In summary, data from remote patient-driven digital monitoring systems have the potential to advance insights into HD disease features and progression that may enable improved clinical trial design and disease management. As demonstrated in this study, the Roche HD Digital Monitoring Platform appears to fulfill the criteria of cross-sectional validation required for a novel platform to be useful in this context. An important limitation of this study is the cross-sectional nature of the data and the lack of a comprehensive evaluation of the platform's clinical validity. Accordingly, the next goal for the platform is to demonstrate sensitivity to clinical change over time and ability to measure drug effect. Such additional longitudinal data are critical to judge the true value of the digital approach versus the standard approach, and these data are presently being generated across the Roche tominersen clinical development program in interventional and observational contexts. Another limitation is the limited understanding of how these prespecified features are linked to what matters for patients in daily life. As explained above, the feature preselection was driven by literature and expert input and as such is mainly signal identification driven. Following recommendations previously outlined for the development of meaningful digital measures [41,42], a qualitative research study to investigate what matters most for patients in their daily life in relation to the HD Digital Monitoring Platform is ongoing. One core strength of digital testing is that it entails the high-frequency collection of data. This enables the development of a broader feature space that has the potential to show even stronger signals, such as features that can differentiate between healthy controls and those with premanifest HD (as demonstrated in this study) and show greater relevance to what matters to patients.

Conclusions

Taken together, the analyses presented support the use of wearable devices and mobile apps to provide further insight into HD disease features and clinical progression previously not possible with standard clinical assessments, enabling improved clinical trial design and, potentially, disease management.

Acknowledgments

The authors thank all the patients who participated in our studies and their families. The authors also acknowledge Jens Schjodt-Eriksen, Luis Matos, Hans Van Wesenbeeck, and Sarper Guerel for their contributions in implementing data capture and

XSL•FO
RenderX

processing solutions; Timothy Kilchenmann and Eduardo Arteaga Bracho for study management; Wei-Yi Cheng, David Slater, and Gregoire Pointeau for contributing to feature development; and Alexis Dorier for conducting data analyses.

The Huntington disease Natural History Study is funded by F. Hoffmann-La Roche Ltd. The open-label extension study was initially sponsored by Ionis Pharmaceuticals and transferred to F. Hoffmann-La Roche Ltd. The Digital-HD study is sponsored by University College London, supported by a research grant to University College London from F. Hoffmann-La Roche Ltd.

The authors thank Kristina Rodriguez, PhD, of MediTech Media for medical writing assistance, which was funded by F. Hoffmann-La Roche Ltd.

CC was affiliated with the Roche Pharma Research and Early Development, pRED Informatics, Pharmaceutical Sciences, Clinical Pharmacology, and Neuroscience, Ophthalmology, and Rare Diseases Discovery and Translational Area at Roche Innovation Center Basel at the time of this study and is currently affiliated with the Rare Disease Research Unit at Pfizer. PW was affiliated with the Department of Neurology at University of Ulm Medical Center at the time of this study and is currently affiliated with the Department of Neurodegenerative Disease and Gerontopsychiatry/Neurology at University of Bonn Medical Center.

Data Availability

For eligible studies, qualified researchers may request access to individual patient-level clinical data through a data request platform. At the time of writing, this request platform is Vivli [43]. Up-to-date details on Roche's global policy on the sharing of clinical information and how to request access to related clinical study documents are available on the Roche website [44]. Anonymised records for individual patients across more than one data source external to Roche can not, and should not, be linked due to a potential increase in risk of patient reidentification.

Authors' Contributions

FL, CC, CG, PW, SAS, EJW, and ML have made substantial contributions to the conception and design of the work. All authors have contributed to the acquisition, analysis, or interpretation of data for this work. All authors have contributed to drafts of the article, revised the manuscript critically for important intellectual content, approved the final version to be published, and agreed to be accountable for all aspects of the article. FL had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Conflicts of Interest

FL is an employee of F. Hoffmann-La Roche Ltd; in addition, FL has a patent WO2020157083A1 pending to F. Hoffmann-La Roche AG. CS is an employee of F. Hoffmann-La Roche Ltd. AB is a consultant to F. Hoffmann-La Roche AG via Inovigate. RT reports grants from F. Hoffmann-La Roche Ltd during the conduct of the study. RT is an employee of F. Hoffmann-La Roche Ltd, though was employed at University College London during the conduct of the study. LMB reports grants from the Huntington Disease Society of America, grants from Hereditary Disease Foundation, grants from Medical Research Council, personal fees from F. Hoffmann-La Roche AG, personal fees from Genentech, and personal fees from the Journal of Huntington Disease outside the submitted work. YPZ is an employee of F. Hoffmann-La Roche Ltd. DW is an employee of F. Hoffmann-La Roche Ltd. AVS is an employee of Ionis Pharmaceuticals. CC was an employee of F. Hoffmann-La Roche Ltd during the conduct of the study. In addition, CC received personal fees from Therachon AG and Pfizer outside the submitted work and is a shareholder of F. Hoffmann-La Roche Ltd. CG is an employee of F. Hoffmann-La Roche Ltd. PW reports personal fees from F. Hoffmann-La Roche Ltd and grants from the European Huntington Disease Network during the conduct of the study. SAS is an employee of F. Hoffmann-La Roche Ltd. FBR reports grants from CHDI Foundation and grants from F. Hoffmann-La Roche Ltd during the conduct of the study and personal fees from Cochrane Collaboration and from Gerson Lehrman Group outside the submitted work. EJW reports grants and personal fees from F. Hoffmann-La Roche Ltd, grants from Medical Research Council UK, and grants from CHDI Foundation during the conduct of the study; and personal fees from Takeda Pharmaceuticals, PTC Therapeutics, Triplet Therapeutics, Mitoconix, and Loqus23 outside the submitted work. ML is a consultant to F. Hoffmann-La Roche AG via Inovigate; in addition, ML has a patent US20190200915A1 pending to Hoffmann-La Roche Inc, and patents WO2020157083A1, EP3701542A2, and WO2019215230A1 pending to F. Hoffmann-La Roche AG.

Multimedia Appendix 1
Animated video describing the Roche Huntington disease digital monitoring platform.
[MP4 File (MP4 Video), 96450 KB - jmir_v24i6e32997_app1.mp4]

Multimedia Appendix 2
Screenshots of active tests on the Roche HD Monitoring app.
[DOCX File , 449 KB - jmir_v24i6e32997_app2.docx]

XSL•FO
RenderX

Multimedia Appendix 3
Quality control pass criteria for digital active tests.
[DOCX File , 33 KB - jmir_v24i6e32997_app3.docx]

Multimedia Appendix 4
Quality control summary statistics.
[DOCX File , 38 KB - jmir_v24i6e32997_app4.docx]

Multimedia Appendix 5
Correlation statistics for the QC pass rate for the Draw-A-Shape test with UHDRS-TMS and Maximal Chorea upper limb item.
[DOCX File , 35 KB - jmir_v24i6e32997_app5.docx]

Multimedia Appendix 6
Correlation coefficients and *P* values between clinical score and digital test.
[DOCX File , 40 KB - jmir_v24i6e32997_app6.docx]

Multimedia Appendix 7
P values for known-groups validity of digital active tests.
[DOCX File , 50 KB - jmir_v24i6e32997_app7.docx]

References

1. Bates GP, Dorsey R, Gusella JF, Hayden MR, Kay C, Leavitt CR, et al. Huntington disease. Nat Rev Dis Primers 2015 Apr 23;1:15005. [doi: 10.1038/nrdp.2015.5] [Medline: 27188817]
2. Roos RA. Huntington's disease: a clinical review. Orphanet J Rare Dis 2010 Dec 20;5:40 [FREE Full text] [doi: 10.1186/1750-1172-5-40] [Medline: 21171977]
3. Dorsey ER, Papapetropoulos S, Xiong M, Kieburtz K. The first frontier: digital biomarkers for neurodegenerative disorders. Digit Biomark 2017 Jul 4;1(1):6-13 [FREE Full text] [doi: 10.1159/000477383] [Medline: 32095743]
4. Post B, Merkus MP, de Bie RM, de Haan RJ, Speelman JD. Unified Parkinson's disease rating scale motor examination: are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable? Mov Disord 2005 Dec;20(12):1577-1584. [doi: 10.1002/mds.20640] [Medline: 16116612]
5. Teipel S, König A, Hoey J, Kaye J, Krüger F, Robillard JM, et al. Use of nonintrusive sensor-based information and communication technology for real-world evidence for clinical trials in dementia. Alzheimers Dement 2018 Sep;14(9):1216-1231 [FREE Full text] [doi: 10.1016/j.jalz.2018.05.003] [Medline: 29936147]
6. Andrzejewski KL, Dowling AV, Stamler D, Felong TJ, Harris DA, Wong C, et al. Wearable sensors in Huntington disease: a pilot study. J Huntingtons Dis 2016 Jun 18;5(2):199-206. [doi: 10.3233/JHD-160197] [Medline: 27341134]
7. Adams JL, Dinesh K, Xiong M, Tarolli CG, Sharma S, Sheth N, et al. Multiple wearable sensors in Parkinson and Huntington disease individuals: a pilot study in clinic and at home. Digit Biomark 2017 Aug 17;1(1):52-63 [FREE Full text] [doi: 10.1159/000479018] [Medline: 32095745]
8. Cohen S, Waks Z, Elm JJ, Gordon MF, Grachev ID, Navon-Perry L, et al. Characterizing patient compliance over six months in remote digital trials of Parkinson's and Huntington disease. BMC Med Inform Decis Mak 2018 Dec 20;18(1):138 [FREE Full text] [doi: 10.1186/s12911-018-0714-7] [Medline: 30572891]
9. Dinesh K, Snyder CW, Xiong M, Tarolli CG, Sharma S, Dorsey ER, et al. A longitudinal wearable sensor study in Huntington's disease. J Huntingtons Dis 2020;9(1):69-81. [doi: 10.3233/JHD-190375] [Medline: 31868675]
10. Tortelli R, Rodrigues FB, Wild EJ. The use of wearable/portable digital sensors in Huntington's disease: a systematic review. Parkinsonism Relat Disord 2021 Feb;83:93-104 [FREE Full text] [doi: 10.1016/j.parkreldis.2021.01.006] [Medline: 33493786]
11. Waddell EM, Dinesh K, Spear KL, Elson MJ, Wagner E, Curtis MJ, et al. GEORGE®: a pilot study of a smartphone application for Huntington's disease. J Huntingtons Dis 2021;10(2):293-301. [doi: 10.3233/JHD-200452] [Medline: 33814455]
12. McLaren B, Andrews SC, Glikmann-Johnston Y, Mercieca EC, Murray NW, Loy C, et al. Feasibility and initial validation of 'HD-Mobile', a smartphone application for remote self-administration of performance-based cognitive measures in Huntington's disease. J Neurol 2021 Feb;268(2):590-601. [doi: 10.1007/s00415-020-10169-y] [Medline: 32880724]
13. Kassavetis P, Saifee TA, Roussos G, Drougkas L, Kojovic M, Rothwell JC, et al. Developing a tool for remote digital assessment of Parkinson's disease. Mov Disord Clin Pract 2015 Oct 20;3(1):59-64 [FREE Full text] [doi: 10.1002/mdc3.12239] [Medline: 30363542]
14. 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]

XSL•FO
RenderX

15. Arora S, Venkataraman V, Zhan A, Donohue S, Biglan KM, Dorsey ER, et al. Detecting and monitoring the symptoms of Parkinson's disease using smartphones: a pilot study. Parkinsonism Relat Disord 2015 Jun;21(6):650-653. [doi: 10.1016/j.parkreldis.2015.02.026] [Medline: 25819808]
16. Lee CY, Kang SJ, Hong SK, Ma HI, Lee U, Kim YJ. A validation study of a smartphone-based finger tapping application for quantitative assessment of bradykinesia in Parkinson's disease. PLoS One 2016 Jul 28;11(7):e0158852 [FREE Full text] [doi: 10.1371/journal.pone.0158852] [Medline: 27467066]
17. Prince J, Arora S, de Vos M. Big data in Parkinson's disease: using smartphones to remotely detect longitudinal disease phenotypes. Physiol Meas 2018 Apr 26;39(4):044005. [doi: 10.1088/1361-6579/aab512] [Medline: 29516871]
18. Elm JJ, Daeschler M, Bataille L, Schneider R, Amara A, Espay AJ, et al. Feasibility and utility of a clinician dashboard from wearable and mobile application Parkinson's disease data. NPJ Digit Med 2019 Sep 25;2:95 [FREE Full text] [doi: 10.1038/s41746-019-0169-y] [Medline: 31583283]
19. Printy BP, Renken LM, Herrmann JP, Lee I, Johnson B, Knight E, et al. Smartphone application for classification of motor impairment severity in Parkinson's disease. Annu Int Conf IEEE Eng Med Biol Soc 2014;2014:2686-2689. [doi: 10.1109/EMBC.2014.6944176] [Medline: 25570544]
20. Zhan A, Mohan S, Tarolli C, Schneider RB, Adams JL, Sharma S, et al. Using smartphones and machine learning to quantify Parkinson disease severity: the mobile Parkinson disease score. JAMA Neurol 2018 Jul 01;75(7):876-880 [FREE Full text] [doi: 10.1001/jamaneurol.2018.0809] [Medline: 29582075]
21. Isaacson SH, Boroojerdi B, Waln O, McGraw M, Kreitzman DL, Klos K, et al. Effect of using a wearable device on clinical decision-making and motor symptoms in patients with Parkinson's disease starting transdermal rotigotine patch: a pilot study. Parkinsonism Relat Disord 2019 Jul;64:132-137. [doi: 10.1016/j.parkreldis.2019.01.025] [Medline: 30948242]
22. Vuong K, Canning CG, Menant JC, Loy CT. Gait, balance, and falls in Huntington disease. Handb Clin Neurol 2018;159:251-260. [doi: 10.1016/B978-0-444-63916-5.00016-1] [Medline: 30482318]
23. Hamilton JM, Salmon DP, Corey-Bloom J, Gamst A, Paulsen JS, Jerkins S, et al. Behavioural abnormalities contribute to functional decline in Huntington's disease. J Neurol Neurosurg Psychiatry 2003 Jan;74(1):120-122 [FREE Full text] [doi: 10.1136/jnnp.74.1.120] [Medline: 12486282]
24. Nehl C, Paulsen JS, Huntington Study Group. Cognitive and psychiatric aspects of Huntington disease contribute to functional capacity. J Nerv Ment Dis 2004 Jan;192(1):72-74. [doi: 10.1097/01.nmd.0000106004.67587.57] [Medline: 14718780]
25. Ready RE, Mathews M, Leserman A, Paulsen JS. Patient and caregiver quality of life in Huntington's disease. Mov Disord 2008 Apr 15;23(5):721-726 [FREE Full text] [doi: 10.1002/mds.21920] [Medline: 18175350]
26. Simpson JA, Lovecky D, Kogan J, Vetter LA, Yohrling GJ. Survey of the Huntington's disease patient and caregiver community reveals most impactful symptoms and treatment needs. J Huntingtons Dis 2016 Dec 15;5(4):395-403. [doi: 10.3233/JHD-160228] [Medline: 27983566]
27. Lipsmeier F, Cheng W, Wolf D, Zhang YP, Kilchenmann T, Bamdadian A, et al. Digital, high-frequency, long-term monitoring of motor and non-motor symptoms in huntington’s disease (hd) patients. J Neurol Neurosurg Psychiatry 2018 Sep 01;89(Suppl 1):A61. [doi: 10.1136/jnnp-2018-EHDN.162]
28. Lipsmeier F, Simillion C, Bamdadian A, Smith A, Schobel S, Czech C, et al. Preliminary reliability and validity of a novel digital biomarker smartphone application to assess cognitive and motor symptoms in Huntington’s disease (HD) (P1.8-042). Neurology 2019 Apr 09;92(15 Supplement):P1.8-P042 [FREE Full text]
29. Kremer HPH, Hungtington Study Group. Unified Huntington's Disease Rating Scale: reliability and consistency. Huntington Study Group. Mov Disord 1996 Mar;11(2):136-142. [doi: 10.1002/mds.870110204] [Medline: 8684382]
30. Stroop JR. Studies of interference in serial verbal reactions. J Exp Psychol 1935;18(6):643-662. [doi: 10.1037/h0054651]
31. Smith A. Symbol Digit Modalities Test. Los Angeles, CA: Western Psychological Services; 1973.
32. Michell AW, Goodman AO, Silva AH, Lazic SE, Morton AJ, Barker RA. Hand tapping: a simple, reproducible, objective marker of motor dysfunction in Huntington's disease. J Neurol 2008 Aug;255(8):1145-1152. [doi: 10.1007/s00415-008-0859-x] [Medline: 18465109]
33. Salomonczyk D, Panzera R, Pirogovosky E, Goldstein J, Corey-Bloom J, Simmons R, et al. Impaired postural stability as a marker of premanifest Huntington's disease. Mov Disord 2010 Oct 30;25(14):2428-2433. [doi: 10.1002/mds.23309] [Medline: 20818666]
34. Tian JR, Herdman SJ, Zee DS, Folstein SE. Postural control in Huntington's disease (HD). Acta Otolaryngol Suppl 1991;481:333-336. [doi: 10.3109/00016489109131415] [Medline: 1833947]
35. Hausdorff JM, Cudkowicz ME, Firtion R, Wei JY, Goldberger AL. Gait variability and basal ganglia disorders: stride-to-stride variations of gait cycle timing in Parkinson's disease and Huntington's disease. Mov Disord 1998 May;13(3):428-437. [doi: 10.1002/mds.870130310] [Medline: 9613733]
36. Mancini M, Salarian A, Carlson-Kuhta P, Zampieri C, King L, Chiari L, et al. ISway: a sensitive, valid and reliable measure of postural control. J Neuroeng Rehabil 2012 Aug 22;9:59 [FREE Full text] [doi: 10.1186/1743-0003-9-59] [Medline: 22913719]

XSL•FO
RenderX

37. Cheng WY, Bourke AK, Lipsmeier F, Bernasconi C, Belachew S, Gossens C, et al. U-turn speed is a valid and reliable smartphone-based measure of multiple sclerosis-related gait and balance impairment. Gait Posture 2021 Feb;84:120-126 [FREE Full text] [doi: 10.1016/j.gaitpost.2020.11.025] [Medline: 33310432]
38. Moon Y, Sung J, An R, Hernandez ME, Sosnoff JJ. Gait variability in people with neurological disorders: a systematic review and meta-analysis. Hum Mov Sci 2016 Jun;47:197-208. [doi: 10.1016/j.humov.2016.03.010] [Medline: 27023045]
39. Bechtel N, Scahill RI, Rosas HD, Acharya T, van den Bogaard SJ, Jauffret C, et al. Tapping linked to function and structure in premanifest and symptomatic Huntington disease. Neurology 2010 Dec 14;75(24):2150-2160 [FREE Full text] [doi: 10.1212/WNL.0b013e3182020123] [Medline: 21068430]
40. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-428. [doi: 10.1037//0033-2909.86.2.420] [Medline: 18839484]
41. Taylor KI, Staunton H, Lipsmeier F, Nobbs D, Lindemann M. Outcome measures based on digital health technology sensor data: data- and patient-centric approaches. NPJ Digit Med 2020 Jul 23;3:97 [FREE Full text] [doi: 10.1038/s41746-020-0305-8] [Medline: 32715091]
42. 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 Sep 15;4(3):69-77 [FREE Full text] [doi: 10.1159/000509725] [Medline: 33083687]
43. Our members. Vivli. URL: https://vivli.org/ourmember/roche/ [accessed 2022-05-20]
44. Our commitment to data sharing. Roche. URL: https://www.roche.com/innovation/process/clinical-trials/data-sharing/ [accessed 2022-05-20]

Abbreviations

CAG: cytosine adenine guanine
D: dominant
HD: Huntington disease
ICC: intraclass correlation coefficient
IRB: Institutional Review Board
ND: nondominant
NHS: Natural History Study
OLE: open-label extension
QC: quality control
SDMT: Symbol Digit Modalities Test
SWR: Stroop Word Reading
TFC: Total Functional Capacity
TMS: Total Motor Score
UHDRS: Unified HD Rating Scale

Edited by T Leung; submitted 31.08.21; peer-reviewed by I Clay, T Woelfle, E Kuosmanen; comments to author 22.09.21; revised version received 17.02.22; accepted 04.05.22; published 28.06.22.

Please cite as:
Lipsmeier F, Simillion C, Bamdadian A, Tortelli R, Byrne LM, Zhang YP, Wolf D, Smith AV, Czech C, Gossens C, Weydt P, Schobel SA, Rodrigues FB, Wild EJ, Lindemann M
A Remote Digital Monitoring Platform to Assess Cognitive and Motor Symptoms in Huntington Disease: Cross-sectional Validation Study
J Med Internet Res 2022;24(6):e32997
URL: https://www.jmir.org/2022/6/e32997
doi:10.2196/32997
PMID:35763342

©Florian Lipsmeier, Cedric Simillion, Atieh Bamdadian, Rosanna Tortelli, Lauren M Byrne, Yan-Ping Zhang, Detlef Wolf, Anne V Smith, Christian Czech, Christian Gossens, Patrick Weydt, Scott A Schobel, Filipe B Rodrigues, Edward J Wild, Michael Lindemann. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.06.2022. 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

XSL•FO
RenderX

information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Disease Identification Algorithm for Medical Crowdfunding Campaigns: Validation Study

Steven S Doerstling[1,2], BSPH; Dennis Akrobetu[1,2], BS; Matthew M Engelhard[3], MD, PhD; Felicia Chen[4], BS; Peter A Ubel[5,6], MD

[1]Duke University School of Medicine, Duke University, Durham, NC, United States
[2]Margolis Center for Health Policy, Duke University, Durham, NC, United States
[3]Department of Biostatistics & Bioinformatics, Duke University, Durham, NC, United States
[4]Apple, Inc, Cupertino, CA, United States
[5]Fuqua School of Business, Duke University, Durham, NC, United States
[6]Sanford School of Public Policy, Duke University, Durham, NC, United States

Corresponding Author:
Steven S Doerstling, BSPH
Duke University School of Medicine
Duke University
DUMC 3710
Durham, NC, 27710
United States
Phone: 1 704 497 7007
Email: sd363@duke.edu

Abstract

Background: Web-based crowdfunding has become a popular method to raise money for medical expenses, and there is growing research interest in this topic. However, crowdfunding data are largely composed of unstructured text, thereby posing many challenges for researchers hoping to answer questions about specific medical conditions. Previous studies have used methods that either failed to address major challenges or were poorly scalable to large sample sizes. To enable further research on this emerging funding mechanism in health care, better methods are needed.

Objective: We sought to validate an algorithm for identifying 11 disease categories in web-based medical crowdfunding campaigns. We hypothesized that a disease identification algorithm combining a named entity recognition (NER) model and word search approach could identify disease categories with high precision and accuracy. Such an algorithm would facilitate further research using these data.

Methods: Web scraping was used to collect data on medical crowdfunding campaigns from GoFundMe (GoFundMe Inc). Using pretrained NER and entity resolution models from Spark NLP for Healthcare in combination with targeted keyword searches, we constructed an algorithm to identify conditions in the campaign descriptions, translate conditions to *International Classification of Diseases, 10th Revision, Clinical Modification* (ICD-10-CM) codes, and predict the presence or absence of 11 disease categories in the campaigns. The classification performance of the algorithm was evaluated against 400 manually labeled campaigns.

Results: We collected data on 89,645 crowdfunding campaigns through web scraping. The interrater reliability for detecting the presence of broad disease categories in the campaign descriptions was high (Cohen κ: range 0.69-0.96). The NER and entity resolution models identified 6594 unique (276,020 total) ICD-10-CM codes among all of the crowdfunding campaigns in our sample. Through our word search, we identified 3261 additional campaigns for which a medical condition was not otherwise detected with the NER model. When averaged across all disease categories and weighted by the number of campaigns that mentioned each disease category, the algorithm demonstrated an overall precision of 0.83 (range 0.48-0.97), a recall of 0.77 (range 0.42-0.98), an F_1 score of 0.78 (range 0.56-0.96), and an accuracy of 95% (range 90%-98%).

Conclusions: A disease identification algorithm combining pretrained natural language processing models and ICD-10-CM code–based disease categorization was able to detect 11 disease categories in medical crowdfunding campaigns with high precision and accuracy.

(J Med Internet Res 2022;24(6):e32867) doi:10.2196/32867

XSL•FO
RenderX

KEYWORDS

crowdfunding; natural language processing; named entity recognition; health care costs; GoFundMe

Introduction

Many patients share details about their health care experiences on the internet. Various platforms, ranging from social media to discussion forums, provide an outlet to convey aspects of the patient experience that may not be captured by feedback surveys or academic studies. For example, researchers have analyzed web-based hospital reviews to understand which hospital quality metrics are the most important to patients [1]. Other work has analyzed Twitter posts to detect adverse drug reactions alongside traditional adverse event reporting systems [2,3]. Considering the public availability and high volume of patient-authored, web-based content, these posts constitute an important data source for gleaning insights about the real-world impact of health care from the patient perspective.

One such data source that has gained recent attention is web-based crowdfunding, which has become a popular method that many in the United States use to raise money for medical expenses. As of April 2019, more than US $3 billion was raised for personal medical expenses on GoFundMe (GoFundMe Inc)—the largest web-based crowdfunding platform. To understand how different patient populations are impacted by medical expenses, recent studies have used data from GoFundMe to identify campaigns associated with specific, narrowly defined medical conditions, focusing on, for example, cancer [4,5], injuries [6], or neurologic diseases [7]. However, because GoFundMe campaigns do not contain any structured data on medical conditions, these details must be inferred from the free text of each campaign description. To address this challenge, a variety of methods have been explored to identify campaigns associated with specific medical conditions, including manual reviews [8]; rule-based approaches based on keywords and regular expressions [9]; and, more recently, biomedical word embeddings for establishing similarities to reference words for broad disease categories [6].

Each of these approaches has important shortcomings. Rule-based approaches might systematically overlook misspelled diagnoses or the conversational phrasing of medical terms. Manual reviews are time intensive and thus scale poorly to larger sample sizes. Strategies based on biomedical word embeddings are promising but are highly context dependent and may perform unpredictably with crowdfunding campaigns because of frequent misspellings and vague medical terminology. Additionally, most medical crowdfunding studies have focused on a single or small number of disease categories, and disease categories are often treated as mutually exclusive at the campaign level [8,10,11], even though many people seek money to pay for the cost of multiple illnesses.

Considering these challenges, better methods are needed to answer important questions about the scale and impact of medical crowdfunding. To facilitate this work, we sought to construct an algorithm to more accurately and comprehensively identify medical diagnoses in medical crowdfunding campaigns. We used a named entity recognition (NER) model, which can be trained to predict phrases that represent medical conditions and have been successfully applied to medical corpora for disease identification [12]. Medical conditions identified by the NER model were then converted to *International Classification of Diseases, 10th Revision, Clinical Modification* (ICD-10-CM) codes to group conditions into disease categories. In this paper, we present data on the precision and reliability of a new algorithm that was designed to detect the presence or absence of 11 mutually inclusive disease categories in medical crowdfunding campaigns.

Methods

Data Collection

We wrote a web scraping program to collect data from medical crowdfunding campaigns that are hosted by GoFundMe. The program accessed a random sample of the GoFundMe sitemap [13], which contains links to GoFundMe crowdfunding campaigns that are made available to search engines. Web scraping was completed in August 2020. Data were collected from campaigns that were self-categorized as *Medical, Illness & Healing* and located in the United States.

Ethics Approval

This study was approved by the Duke University Institutional Review Board (IRB number 2020-0435). All data collected from GoFundMe were publicly available and aggregated for research purposes in accordance with fair use. The source code is available on GitHub [14].

Disease Identification and Resolution to the ICD-10-CM

In order to identify medical diagnoses in the descriptions of crowdfunding campaigns, we used an NER model developed by Spark NLP for Healthcare [15]. The NER model identifies segments of text that are predicted to represent medical diagnoses. Each text segment that was identified as a medical diagnosis was subsequently entered into an entity resolution model, which was also developed by Spark NLP for Healthcare [16]. The entity resolution model selects the ICD-10-CM codes that most closely match the input text according to the distance between embedding vectors. Together, this pipeline generates a list of medical diagnoses and their corresponding ICD-10-CM codes for each campaign description.

Categorizing ICD-10-CM Codes

Our goal was to sort ICD-10-CM codes into clinically coherent disease categories. We used the 2021 Clinical Classifications Software Refined (CCSR; Healthcare Cost and Utilization Project) for ICD-10-CM diagnoses [17], which groups ICD-10-CM codes at the following two levels of specificity: a narrow CCSR clinical category (eg, *Heart failure*) and a broad diagnosis chapter (eg, *Diseases of the Circulatory System*). ICD-10-CM codes from certain CCSR clinical categories were reassigned to a different diagnosis chapter to consolidate the number of disease categories and prioritize a system-based

XSL•FO
RenderX

classification of diseases (Multimedia Appendix 1). For example, congenital abnormalities were reassigned to categories related to the impacted organ systems. Afterward, we selected 11 diagnosis chapters for disease categories that we sought to identify in crowdfunding campaigns. These categories were chosen because they represented common medical conditions in the United States and were suitable for principal diagnoses that are made according to ICD-10-CM documentation. The diagnosis chapters included for our analysis were renamed to differentiate them from those in the official ICD-10-CM and CCSR documentation (Multimedia Appendix 2). The final assignment of ICD-10-CM codes to disease categories is provided in Multimedia Appendix 3.

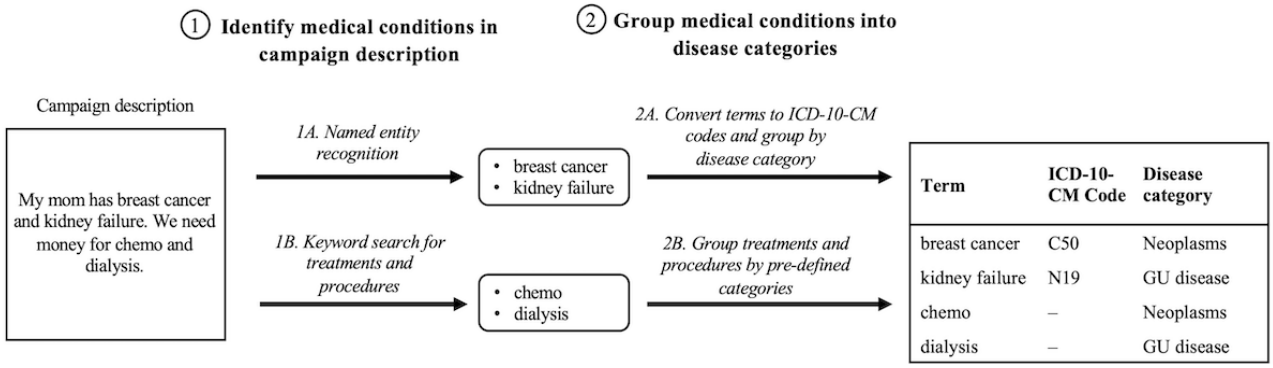
Identification of Disease Categories by Using a Word Search

Our research team, which was comprised of a senior physician, 2 medical students, and research assistants with undergraduate and master's degrees, conducted several rounds of exploratory reading of crowdfunding campaigns to understand how medical details were conveyed. We observed that crowdfunding campaigns sometimes did not explicitly state a medical diagnosis but instead referenced a procedure or treatment that implied the presence of a diagnosis. For example, mentioning *chemotherapy* suggests the presence of a neoplasm. Considering that the NER model used in our study was trained to identify medical diagnoses and not procedures or treatments, campaigns that failed to mention a diagnosis would be missed. Other pretrained NER models exist for the detection of treatments and procedures, but the use of these models was outside the scope of this project. Instead, we compiled a list of treatments and procedures that appeared during our team's review and assigned each term to a disease category. If a term was present in a campaign description, we indicated that the term's corresponding disease category was present in the campaign.

Recoding ICD-10-CM Codes

Certain ICD-10-CM codes that were identified by the entity resolution model did not have an exact match in the CCSR data that were used to group codes into disease categories. To align these codes from the entity resolution model with the CCSR data, we removed the last character of the unmatched code, thereby creating a trimmed code, and checked if any code in the CCSR data began with the resulting trimmed code. This process was repeated until a match was found. If multiple CCSR codes were found to begin with the trimmed code, then the unmatched code was assigned to the disease category that was the most common among the matched CCSR codes. The final set of recoded ICD-10-CM codes was then merged with disease categories that were derived from CCSR, thereby aligning each campaign's identified medical conditions with their corresponding disease categories. ICD-10-CM codes that mapped to the *Other* category were then removed. Each remaining disease category was summarized as "present" or "absent" for all campaigns. A schematic of the algorithm is shown in Figure 1.

Figure 1. A schematic diagram of the disease identification algorithm. This figure shows how this study's algorithm determines which disease categories are present in a hypothetical example that is representative of web-based medical crowdfunding text. Medical conditions are identified in the text by using a named entity recognition model to identify diagnoses and keyword searches to identify treatments and procedures. Diagnoses identified by the named entity recognition model are assigned to best-matching ICD-10-CM codes by using an entity resolution model and grouped according to the disease category definitions outlined in the *Methods* section. Treatments and procedures were used to indicate the presence of corresponding disease categories (defined in Table 1). GU: genitourinary; ICD-10-CM: *International Classification of Diseases, 10th Revision, Clinical Modification*.



Evaluation of Classification Performance

We created a manually labeled reference set to evaluate the ability of our algorithm to detect medical diagnoses in crowdfunding campaign text. A subset of campaigns (n=400) was independently reviewed by 2 medical students, which we considered as the ground truth. The reviewers identified medical diagnoses in the campaign descriptions and identified the best corresponding disease category for each term according to the groups of ICD-10-CM codes defined in the *Categorizing ICD-10-CM Codes* section. Each disease category was indicated as "present" or "absent" in the campaign. Interrater reliability was evaluated by using the Cohen κ. Discrepancies in labeling were reconciled in a group meeting among the students, and remaining disagreements were resolved by a senior physician. The presence or absence of each disease category was similarly determined by the algorithm, constituting a test set. Classification performance metrics for each disease category were then calculated in comparison with our expert consensus reference set. The reference set is provided in Multimedia Appendix 4. All analyses were done using Python version 3.8.8.

Results

After applying the modifications described in the *Methods* section to CCSR, each ICD-10-CM code mapped to a single

XSL•FO
RenderX

disease category. The NER and entity resolution models identified 6594 unique (276,020 total) ICD-10-CM codes among the 89,645 crowdfunding campaigns in our sample. Of the 6594 unique ICD-10-CM codes identified by the entity resolution model, 2884 (43.7%) did not have an exact match in the ICD-10-CM codes constituting our disease categories. Of these 2884 unmatched codes, 2544 (88.2%) were matched to a code with an identical stem and additional alphanumeric characters, indicating a more precise diagnosis. For example, the code "C5091" was identified by the NER model, and it represents cancer of the breast at an unspecified site. More precise codes, such as "C50911," which indicates cancer of the right female breast, are included in the official CCSR documentation. Therefore, the iterative trimming process would match these codes and allow the unmatched code to inherit the proper disease category assignment.

Our manual review of the NER and entity resolution model outputs demonstrated the algorithm's ability to appropriately identify and categorize misspelled diagnoses. For example, the phrase *brain aneruism* (correct spelling: *aneurysm*) was appropriately identified as a cerebral aneurysm and mapped to the *cardiovascular diseases* category. Another campaign contained the phrase *myeloid lukemia* (correct spelling: *leukemia*), but this was nonetheless appropriately categorized as a neoplasm.

Search terms for additional indicators of a disease category are shown in Table 1. Through our word search, we identified 3261 additional campaigns for which a medical condition was not otherwise detected with the NER model. Search terms for injuries and external causes allowed us to identify the most additional campaigns (n=1586), followed by search terms for cardiovascular diseases (n=598), neoplasms (n=486), genitourinary diseases (n=428), gastrointestinal diseases (n=135), and respiratory diseases (n=74). Furthermore, the word search often identified additional disease categories outside of those identified by the NER model. Among these campaigns, search terms identifying a new instance of neoplasms were most common (campaigns: n=19,079), followed by search terms for injuries and external causes (campaigns: n=9238), genitourinary diseases (campaigns: n=2086), cardiovascular diseases (campaigns: n=1919), gastrointestinal diseases (campaigns: n=648), and respiratory diseases (campaigns: n=440). The contribution of each individual search term is shown in Multimedia Appendix 5.

The relative contribution of the word search to identifying disease categories that were not otherwise found by the NER model was small (Figure 2). Instances of disease categories that were detected exclusively via the word search ranged from 2.6% (993/38,221) for neoplasms to 25.2% (1185/4698) for genitourinary diseases. The word search more often identified disease categories that were also identified by the NER model. However, the exclusive contributions of the word search varied by disease category. For example, 94.9% (18,572/19,565) of the word search–identified campaigns mentioning neoplasms were identified by the NER model. Further, only 52.3% (269/514) of the word search–identified campaigns mentioning respiratory disease were identified by the NER model.

To understand the extent of overlap between disease categories that were identified by the word search and those that were identified by the NER model, we calculated how often the disease categories that were identified by each method co-occurred (Figure 3). The rates of co-occurrence also varied by disease category. For example, 53.5% (8371/15,634) of the NER model–identified campaigns mentioning injuries and external causes were also identified by the word search; the overlap was slightly lower for campaigns mentioning neoplasms (18,572/37,228, 49.9%) and genitourinary diseases (1329/3513, 37.8%). Co-occurrence rates were modest for the remaining disease categories that were common among those identified by the NER model and word search.

When preparing the reference set, the interrater reliability for detecting the presence of broad disease categories in the campaign descriptions was high (Cohen κ: range 0.69-0.96). The Cohen κ values for each disease category are shown in Multimedia Appendix 6. Discrepancies in coder annotation often occurred due to imprecise or vague descriptions of medical conditions. For example, one campaign described complications of a feeding tube, but it was unclear if the text sufficiently described a medical condition that was related to the gastrointestinal system. Other campaigns described a "sternum issue" or the patient getting "badly hurt" in an accident. After resolving these discrepancies, a reference set of disease categories in each campaign was established. The presence or absence of each disease category was then determined by the algorithm, and these outputs were compared to those in the reference set.

Classification performance metrics are detailed in Table 2 (additional values are included in Multimedia Appendix 7). The number of campaigns in our reference set that mentioned each disease category ranged from 18 (gastrointestinal diseases) to 162 (neoplasms). Classification performance also varied by disease category. When averaged across all disease categories and weighted by the number of campaigns that mentioned each disease category, the algorithm demonstrated an overall precision of 0.83 (range 0.48-0.97), a recall of 0.77 (range 0.42-0.98), an F_1 score of 0.78 (range 0.56-0.96), and an accuracy of 95% (range 90%-98%). Representative examples of false positives and false negatives are provided in Multimedia Appendix 8.

XSL•FO
RenderX

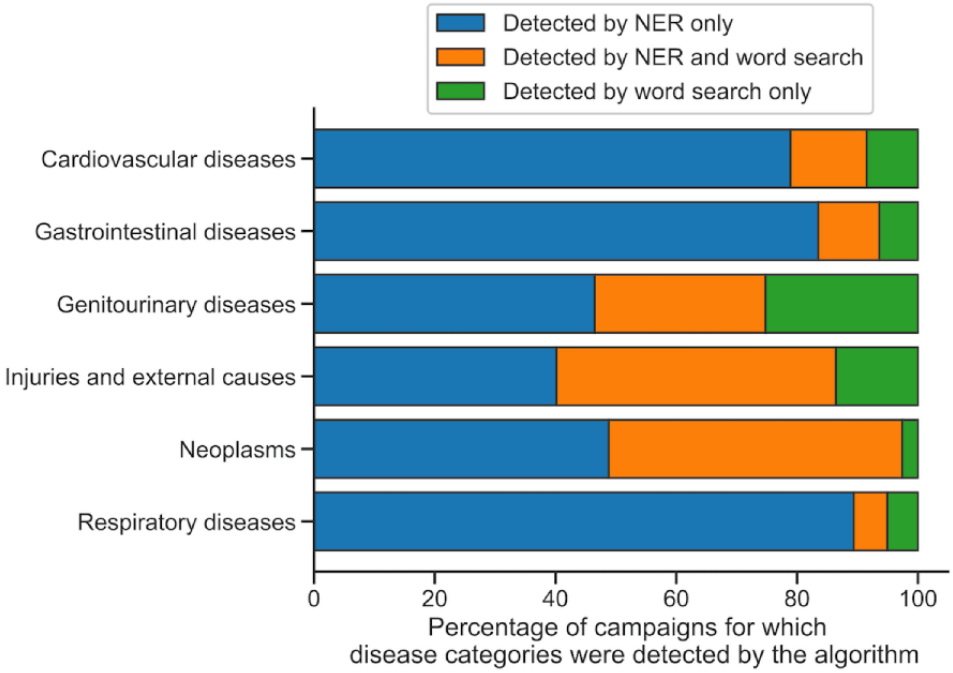
Table 1. The keywords used to identify additional disease categories in campaign descriptions.

Disease category[a]	Keywords searched in campaign descriptions[b]	Representative examples from campaign descriptions
Injuries and external causes	*accident*, *injury/injuries/injured*, *crash*, *collision*, and *burn/burns/burned*	"[He] got into a serious accident in October. All four extremities were injured but the most severe were his legs."
Cardiovascular diseases	*heart transplant* and *heart surgery*	"His cardiologist has informed him that a heart transplant is [his] only hope for survival."
Neoplasms	*chemo/chemotherapy*, *radiation/radiotherapy*, and *bone marrow transplant*	"The chemotherapy did not stabilize the lymphoma so we were unable to move forward with the transplant."
Genitourinary diseases	*dialysis* and *kidney/renal transplant*	"This disease resulted in my kidneys failing and having to start dialysis."
Gastrointestinal diseases	*liver transplant*	"...the cirrhosis is incurable without a complete liver transplant."
Respiratory diseases	*lung transplant*	"Her desire to live life...will only be possible with the double lung transplant."

[a]Each disease category was indicated as present in a campaign if any of the corresponding terms were included in the campaign description.

[b]Keywords were selected during the exploratory reading of crowdfunding campaigns as indicators of a disease category that did not specify a diagnosis.

Figure 2. The relative contributions of the NER model and word search to detecting disease categories. All campaigns for which the disease categories on the y-axis were detected by the disease identification algorithm are presented. The colored bars represent the percentage of those campaigns for which the disease categories were detected by the NER model only (blue), the NER model and word search (orange), or the word search only (green). NER: named entity recognition.



XSL•FO
RenderX

Figure 3. The co-occurrence of disease categories identified by the NER model and word search. The heat map values represent the percentage of campaigns containing the disease category in each row (identified by the NER model) that also contain the disease category in each column (identified via word search). NER: named entity recognition.

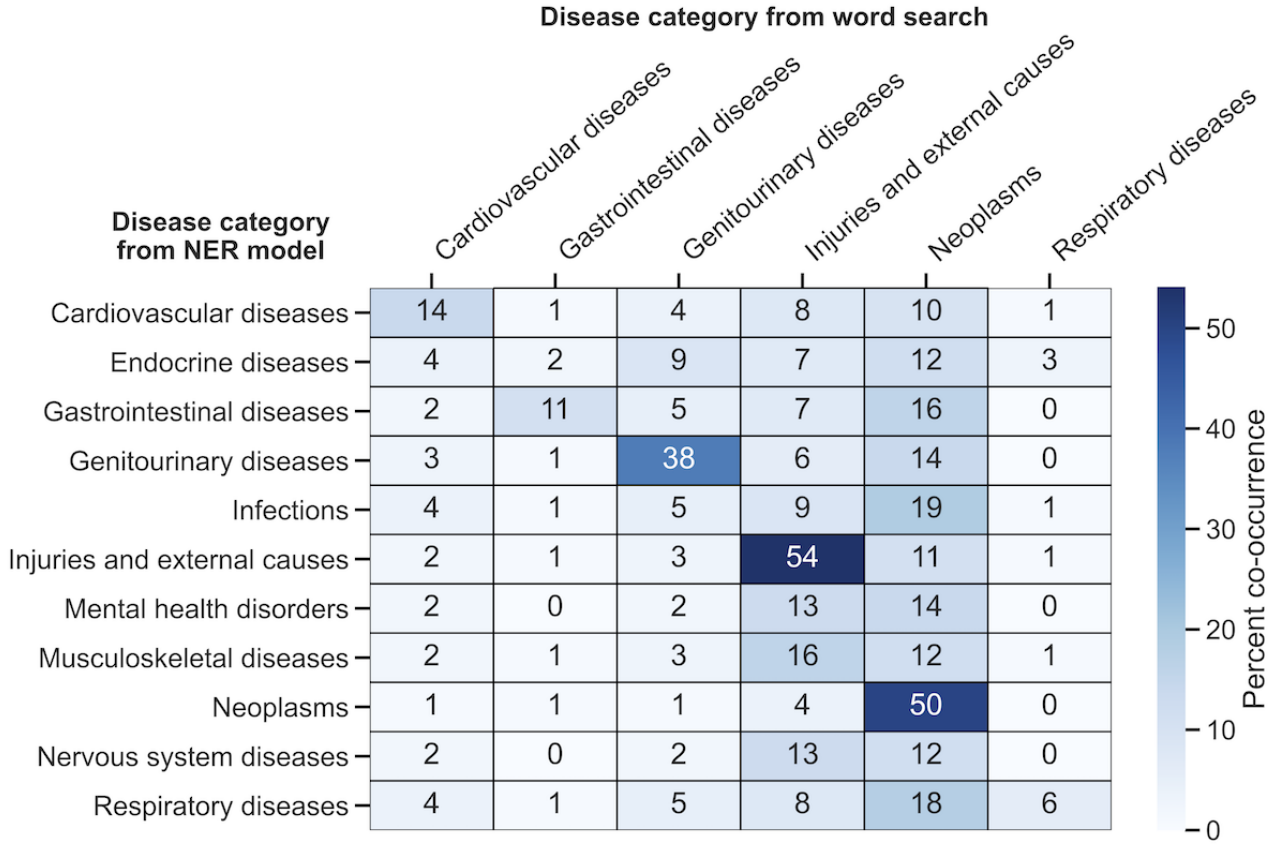


Table 2. Classification performance of the disease identification algorithm[a].

Disease category	Campaigns in the reference set that mention disease category, n	Precision (95% CI)	Recall (95% CI)	F_1 score	Accuracy (95% CI)
Cardiovascular diseases	82	0.92 (0.86-0.99)	0.74 (0.65-0.84)	0.82	0.94 (0.91-0.96)
Endocrine diseases	19	0.75 (0.54-0.96)	0.63 (0.41-0.85)	0.69	0.97 (0.96-0.99)
Gastrointestinal diseases	18	0.56 (0.33-0.79)	0.56 (0.33-0.79)	0.56	0.96 (0.94-0.98)
Genitourinary diseases	35	0.97 (0.90-1.03)	0.8 (0.67-0.93)	0.88	0.98 (0.97-0.99)
Infections	30	0.56 (0.41-0.71)	0.77 (0.62-0.92)	0.65	0.94 (0.91-0.96)
Injuries and external causes	53	0.69 (0.58-0.80)	0.92 (0.85-1.00)	0.79	0.94 (0.91-0.96)
Mental health disorders	20	0.48 (0.30-0.66)	0.7 (0.50-0.90)	0.57	0.95 (0.93-0.97)
Musculoskeletal diseases	45	0.64 (0.48-0.80)	0.51 (0.37-0.66)	0.57	0.91 (0.88-0.94)
Neoplasms	162	0.95 (0.91-0.98)	0.98 0.96-1.00)	0.96	0.97 (0.95-0.99)
Nervous system diseases	66	0.88 (0.76-0.99)	0.42 (0.31-0.54)	0.57	0.90 (0.86-0.93)
Respiratory diseases	29	0.92 (0.81-1.03)	0.76 (0.60-0.91)	0.83	0.98 (0.96-0.99)

[a]The average precision, recall, F_1 score, and accuracy values are 0.83, 0.77, 0.78, and 0.95, respectively. Classification performance is based on a comparison to 400 campaigns that were annotated by a team of expert coders. The averages are weighted by the number of campaigns in the reference set that mention each disease category.

Discussion

Principal Results

We found that a disease identification algorithm using pretrained NER and entity resolution models linked to disease categories based on ICD-10-CM codes was able to detect 11 disease categories in crowdfunding campaigns with high precision and accuracy. Our analysis considered disease categories that represented a broad range of medical conditions. To our knowledge, this methodology is able to identify more disease

XSL•FO
RenderX

categories in web-based medical crowdfunding campaigns than those identified by methods used in previous studies.

Our approach overcomes several limitations of previous work for identifying clinical populations in crowdfunding data. Crowdfunding campaigns contain many misspelled medical terms and informal synonyms (eg, *heart attack* vs *myocardial infarction*). Rule-based methods, such as keyword searches, require all acceptable terms for a given medical condition to be defined beforehand. Therefore, failing to account for alternative phrasings of medical conditions, which are expected in a large corpus, could significantly undermine the sensitivity of this approach. In contrast, NER models predict the probability of a certain word or phrase representing a medical condition and can account for variations in spelling and syntax. Another shortcoming of previous work is treating disease categories as mutually exclusive [6-9]. For example, one campaign may be exclusively categorized into the *neoplasms* category even if the campaign also mentions cardiovascular conditions. In our exploratory reading, we found that campaigns often mentioned multiple medical conditions across disease categories. To reflect the co-occurrence of disease categories, our approach treats disease categories as mutually inclusive. There is no external performance benchmark against which to evaluate our results, and to our knowledge, we are the first to report comprehensive evaluation metrics for a method that allows for multi-class disease category labeling and is scalable to medical crowdfunding text.

Based on our team's exploratory reading of crowdfunding campaigns, we incorporated a word search alongside the NER model to identify additional disease categories that were not explicitly medical diagnoses. In general, the unique contribution of word search to the disease identification algorithm was modest. Although the word search identified additional campaigns for which the NER model did not detect any medical diagnoses, these campaigns represented a small proportion of the total campaigns in our sample. Furthermore, campaigns for which a disease category was found via the word search often had the same disease category detected by the NER model. Because we considered the presence or absence of disease categories at the level of entire campaigns, the word search results were often redundant to those from the NER model.

We examined how often a given disease category was detected in a campaign by both the NER model and word search. The co-occurrence rates corroborate the observation that multiple disease categories are often mentioned in the same campaign and underline the limitations of single-class disease category categorization, which has been used in previous work. In addition, while some disease categories (including genitourinary diseases, neoplasms, and injuries and external causes) were frequently found by both the NER model and word search in the same campaign, lower co-occurrence rates were observed among other disease categories. This may reflect the fact that our word search included a relatively narrow set of procedures or treatments when compared with those for the broad scope of medical diagnoses on which the NER model was trained. For example, it is not surprising that among the campaigns that were identified to mention cardiovascular diseases by the NER model, only 13.8% (1506/10,912) were found to contain mentions of heart surgery or a transplant.

A word search is, by definition, a rule-based approach and is therefore subject to the limitations discussed above. Although including a word search did enable the detection of additional campaigns and disease categories, it is fundamentally limited by the scope of included search terms. Therefore, while the search terms included in our algorithm were informed by an exploratory reading, future work should explore the use of the NER-based detection of procedures and treatments to capitalize on the flexibility of such methods for detecting additional clinical entities in patient-authored text.

Using pretrained NER and entity resolution models and disease categories based on ICD-10-CM codes provides a convenient and scalable method for structuring medical crowdfunding data. To our knowledge, there are no pretrained NER models that can detect a broad range of medical conditions in a corpus authored by members of the general public. Most medical NER models are trained on clinical documentation from electronic health records [18], though the particular NER model used in our study was trained on proprietary data [15]. Nevertheless, we found that one such NER model can be successfully applied to nonclinical texts, suggesting that similar approaches are likely to be effective across a much broader range of free text, such as social media posts.

Limitations

Our study has several limitations. First, several disease categories were relatively infrequent in our reference set. This may have limited the classification performance for those disease categories (eg, gastrointestinal diseases). Second, the resolution of medical diagnoses to ICD-10-CM codes was often imperfect, resulting in clearly stated diagnoses sometimes being translated to an incorrect code. Third, an additional challenge with using ICD-10-CM codes was the lack of consistent formatting among the CCSR codes and the entity resolution model outputs. Codes without an exact match in corresponding data make it difficult to preserve diagnosis-level accuracy, but categorizing codes into broad disease categories largely avoids this problem. Fourth, we excluded several disease categories from our analysis, including conditions associated with pregnancy, ocular and otologic diseases, hematologic and immune disorders, and chromosomal abnormalities. Future work should focus on identifying these disease categories. Fifth, we were unable to distinguish between incidental mentions of medical conditions and those directly related to a beneficiary's expenses. Sixth, we reported accuracy as a part of standard model evaluation metrics, but this should be interpreted with caution, given the class imbalance in the reference set.

Conclusions

To address the challenges of identifying medical conditions in crowdfunding text, we leveraged pretrained NER and entity resolution models to predict the presence or absence of broad disease categories in medical crowdfunding campaign text. We evaluated the algorithm against a rigorously established reference set and provided transparent classification metrics. This algorithm precisely and accurately detects disease

XSL•FO
RenderX

categories representing a broad range of pathologies and addresses key limitations of previous work.

Acknowledgments

Funding for this work was provided by the Patrick and Catherine Weldon Donaghue Medical Research Foundation. SSD and DA were supported by the Margolis Scholars Program of the Margolis Center for Health Policy at Duke University. The funders had no role in the design or execution of this study.

Authors' Contributions

SSD and FC were responsible for data collection and analysis. SSD and DA manually reviewed campaigns for the reference set. MME and PAU provided statistical guidance. All authors contributed to the writing, revision, and final approval of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Reassigned Clinical Classifications Software Refined categories.
[PDF File (Adobe PDF File), 35 KB - jmir_v24i6e32867_app1.pdf]

Multimedia Appendix 2
Disease category assignment from diagnosis chapters.
[PDF File (Adobe PDF File), 35 KB - jmir_v24i6e32867_app2.pdf]

Multimedia Appendix 3
Assignment of *International Classification of Diseases, 10th Revision, Clinical Modification* codes to disease categories.
[XLSX File (Microsoft Excel File), 2183 KB - jmir_v24i6e32867_app3.xlsx]

Multimedia Appendix 4
Reference set indicating the presence (1) or absence (0) of each disease category in the corresponding campaign URL.
[XLSX File (Microsoft Excel File), 31 KB - jmir_v24i6e32867_app4.xlsx]

Multimedia Appendix 5
Results of word search for procedures and treatments. The left panel shows the number of campaigns identified by each search term on the y-axis for which no other disease category was found by the named entity recognition (NER) model (ie, found exclusively by word search). The right panel shows the number of campaigns identified by each search term on the y-axis among campaigns that also had at least 1 disease category found by the NER model.
[PDF File (Adobe PDF File), 28 KB - jmir_v24i6e32867_app5.pdf]

Multimedia Appendix 6
Interrater reliability for detecting broad disease categories in campaign descriptions.
[PDF File (Adobe PDF File), 25 KB - jmir_v24i6e32867_app6.pdf]

Multimedia Appendix 7
Classification concordance between the disease identification algorithm and the annotated reference set.
[PDF File (Adobe PDF File), 29 KB - jmir_v24i6e32867_app7.pdf]

Multimedia Appendix 8
Examples of false-positive and false-negative disease category assignments by the disease identification algorithm.
[PDF File (Adobe PDF File), 46 KB - jmir_v24i6e32867_app8.pdf]

References

1. Ranard BL, Werner RM, Antanavicius T, Schwartz HA, Smith RJ, Meisel ZF, et al. Yelp reviews of hospital care can supplement and inform traditional surveys of the patient experience of care. Health Aff (Millwood) 2016 Apr;35(4):697-705 [FREE Full text] [doi: 10.1377/hlthaff.2015.1030] [Medline: 27044971]

XSL•FO
RenderX

2. MacKinlay A, Aamer H, Yepes AJ. Detection of adverse drug reactions using medical named entities on Twitter. AMIA Annu Symp Proc 2018 Apr 16;2017:1215-1224 [FREE Full text] [Medline: 29854190]
3. Cocos A, Fiks AG, Masino AJ. Deep learning for pharmacovigilance: recurrent neural network architectures for labeling adverse drug reactions in Twitter posts. J Am Med Inform Assoc 2017 Jul 01;24(4):813-821 [FREE Full text] [doi: 10.1093/jamia/ocw180] [Medline: 28339747]
4. Cohen AJ, Brody H, Patino G, Ndoye M, Liaw A, Butler C, et al. Use of an online crowdfunding platform for unmet financial obligations in cancer care. JAMA Intern Med 2019 Dec 01;179(12):1717-1720 [FREE Full text] [doi: 10.1001/jamainternmed.2019.3330] [Medline: 31498408]
5. Loeb S, Taneja S, Walter D, Zweifach S, Byrne N. Crowdfunding for prostate cancer and breast cancer. BJU Int 2018 Nov;122(5):723-725. [doi: 10.1111/bju.14408] [Medline: 29786946]
6. Angraal S, Zachariah AG, Raaisa R, Khera R, Rao P, Krumholz HM, et al. Evaluation of internet-based crowdsourced fundraising to cover health care costs in the United States. JAMA Netw Open 2021 Jan 04;4(1):e2033157 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.33157] [Medline: 33427882]
7. Snyder J, Turner L. Crowdfunding for stem cell-based interventions to treat neurologic diseases and injuries. Neurology 2019 Aug 06;93(6):252-258. [doi: 10.1212/WNL.0000000000007838] [Medline: 31227615]
8. Saleh SN, Ajufo E, Lehmann CU, Medford RJ. A comparison of online medical crowdfunding in Canada, the UK, and the US. JAMA Netw Open 2020 Oct 01;3(10):e2021684 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.21684] [Medline: 33104206]
9. Silver ER, Truong HQ, Ostvar S, Hur C, Tatonetti NP. Association of neighborhood deprivation index with success in cancer care crowdfunding. JAMA Netw Open 2020 Dec 01;3(12):e2026946 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.26946] [Medline: 33270122]
10. Durand WM, Peters JL, Eltorai AEM, Kalagara S, Osband AJ, Daniels AH. Medical crowdfunding for organ transplantation. Clin Transplant 2018 Jun;32(6):e13267. [doi: 10.1111/ctr.13267] [Medline: 29683220]
11. Durand WM, Johnson JR, Eltorai AEM, Daniels AH. Medical crowdfunding for patients undergoing orthopedic surgery. Orthopedics 2018 Jan 01;41(1):e58-e63. [doi: 10.3928/01477447-20171114-04] [Medline: 29156070]
12. Kocaman V, Talby D. Biomedical named entity recognition at scale. arXiv Preprint posted online on November 12, 2020. [FREE Full text] [doi: 10.1007/978-3-030-68763-2_48]
13. GoFundMe sitemap. GoFundMe. URL: https://www.gofundme.com/sitemap.xml [accessed 2021-04-26]
14. sdoerstling/medical_crowdfunding_methods: Source code to construct and validate a disease identification algorithm for online medical crowdfunding campaigns. GitHub. URL: https://github.com/sdoerstling/medical_crowdfunding_methods [accessed 2021-07-31]
15. Detect clinical entities (jsl_ner_wip_clinical)- Spark NLP model. John Snow Labs. URL: https://nlp.johnsnowlabs.com/2021/01/18/jsl_ner_wip_clinical_en.html [accessed 2021-04-26]
16. Sentence entity resolver for ICD10-CM (augmented)- Spark NLP model. John Snow Labs. URL: https://nlp.johnsnowlabs.com/2020/12/13/sbiobertresolve_icd10cm_augmented_en.html [accessed 2021-04-26]
17. Clinical Classifications Software Refined (CCSR). Healthcare Cost and Utilization Project. URL: https://www.hcup-us.ahrq.gov/toolssoftware/ccsr/ccs_refined.jsp [accessed 2021-04-26]
18. Uzuner Ö, South BR, Shen S, DuVall SL. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. J Am Med Inform Assoc 2011;18(5):552-556 [FREE Full text] [doi: 10.1136/amiajnl-2011-000203] [Medline: 21685143]

Abbreviations

CCSR: Clinical Classifications Software Refined
ICD-10-CM: *International Classification of Diseases, 10th Revision, Clinical Modification*
NER: named entity recognition

Edited by R Kukafka; submitted 09.09.21; peer-reviewed by M Torii, S Doan; comments to author 17.11.21; revised version received 11.03.22; accepted 20.04.22; published 21.06.22.

Please cite as:
Doerstling SS, Akrobetu D, Engelhard MM, Chen F, Ubel PA
A Disease Identification Algorithm for Medical Crowdfunding Campaigns: Validation Study
J Med Internet Res 2022;24(6):e32867
URL: https://www.jmir.org/2022/6/e32867
doi:10.2196/32867
PMID:35727610

XSL•FO
RenderX

©Steven S Doerstling, Dennis Akrobetu, Matthew M Engelhard, Felicia Chen, Peter A Ubel. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.06.2022. 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

Medication Use and Clinical Outcomes by the Dutch Institute for Clinical Auditing Medicines Program: Quantitative Analysis

Rawa Kamaran Ismail[1*], PhD, PharmD; Jesper van Breeschoten[2*], PharmD; Silvia van der Flier[3], PhD; Caspar van Loosen[4], MSc; Anna Maria Gerdina Pasmooij[5], PhD; Maaike van Dartel[5], PhD; Alfons van den Eertwegh[6], PhD, MD; Anthonius de Boer[1], MD, PhD; Michel Wouters[7], MD, PhD; Doranne Hilarius[8], PhD, PharmD

[1]Division of Pharmacoepidemiology and Clinical Pharmacology, University of Utrecht, Utrecht, Netherlands

[2]Dutch Institute for Clinical Auditing, Leiden, Netherlands

[3]Stichting Volksgezondheidszorg (VGZ), Arnhem, Netherlands

[4]Dutch Association of Hospital Pharmacists, De Meern, Netherlands

[5]Medicines Evaluation Board, Utrecht, Netherlands

[6]Department of Medical Oncology, Amsterdam University Medical Center, location VUmc, Amsterdam, Netherlands

[7]Department of Biomedical Data Sciences, Leiden University Medical Centre, Leiden, Netherlands

[8]Rode Kruis Ziekenhuis, Beverwijk, Netherlands

[*]these authors contributed equally

Corresponding Author:
Jesper van Breeschoten, PharmD
Dutch Institute for Clinical Auditing
Rijnsburgerweg 10
Leiden, 2333AA
Netherlands
Phone: 31 613382358
Email: j.vanbreeschoten@dica.nl

Abstract

Background: The Dutch Institute for Clinical Auditing (DICA) Medicines Program was set up in September 2018 to evaluate expensive medicine use in daily practice in terms of real-world effectiveness using only existing data sources.

Objective: The aim of this study is to describe the potential of the addition of declaration data to quality registries to provide participating centers with benchmark information about the use of medicines and outcomes among patients.

Methods: A total of 3 national population-based registries were linked to clinical and financial data from the hospital pharmacy, the Dutch diagnosis treatment combinations information system including in-hospital activities, and survival data from health care insurers. The first results of the real-world data (RWD) linkage are presented using descriptive statistics to assess patient, tumor, and treatment characteristics. Time-to-next-treatment (TTNT) and overall survival (OS) were estimated using the Kaplan-Meier method.

Results: A total of 21 Dutch hospitals participated in the DICA Medicines Program, which included 7412 patients with colorectal cancer, 1981 patients with metastasized colon cancer, 3860 patients with lung cancer, 1253 patients with metastasized breast cancer, and 7564 patients with rheumatic disease. The data were used for hospital benchmarking to gain insights into medication use in specific patient populations, treatment information, clinical outcomes, and costs. Detailed treatment information (duration and treatment steps) led to insights into differences between hospitals in daily clinical practices. Furthermore, exploratory analyses on clinical outcomes (TTNT and OS) were possible.

Conclusions: The DICA Medicines Program shows that it is possible to gather and link RWD about medicines to 4 disease-specific population-based registries. Since these RWD became available with minimal registration burden and effort for hospitals, this method can be explored in other population-based registries to evaluate real-world efficacy.

(J Med Internet Res 2022;24(6):e33446) doi:10.2196/33446

KEYWORDS

real-world data; quality of care; medicines; cancer

XSL•FO
RenderX

Introduction

Regulatory authorities approve the majority (76%) of new cancer drugs based on evidence provided by randomized controlled trials (RCTs) [1]. These RCTs have high internal validity and are widely considered the gold standard for establishing the efficacy of new drugs [2]. Many new cancer drugs have been recently approved based on very specific patient groups, surrogate outcomes, and lower patient numbers; these drugs are increasingly approved in accelerated tracks [3,4]. The selected patient groups and well-controlled setting of these RCTs has led to criticisms of their external validity [5]. In addition, recent research has shown that almost one-half of RCTs that applied for marketing authorization for new cancer drugs in Europe had a high risk of bias. This increased risk of bias was caused by their design, conducted analyses, and conduct deficits [1]. Further, due to the increase in newly approved cancer and rheumatic disease drugs, health care costs have increased. The total expenditures by hospitals on expensive medicines in the Netherlands reached €2.1 billion (US $2.2 billion) in 2019 [6].

Following market entry, new cancer drugs are prescribed to a broader group of patients with different characteristics. This leads to a gap in clinical outcomes evidence between RCTs and the real world [7,8]. During routine clinical practice, real-world data (RWD) are generated and registered in validated population-based cancer registries. Clinical quality registries are an important tool for quality assessment and improvement in hospitals, consequently leading to demonstrable improvements in patient outcomes [9]. Comparing the quality of care across hospitals results in insights into differences in outcomes, which can lead to improvements in care [9,10]. Furthermore, data from quality registries are used for outcomes research and to study practice variation between centers using quality indicators [11]. Besides clinical quality registries, detailed administrative and declaration data are available specifically on the use of (expensive) drug treatments. The combination of these data in clinical quality registries, hospital administrative data, and declaration data of drugs used in these indications could be valuable to bridge the efficacy-effectiveness gap.

Previous initiatives linked various databases on drugs to clinical data. This linkage made it feasible to study drug use, health resource use, costs, effectiveness, and the safety of medicines [12]. However, a gap remains for recently approved expensive cancer drugs.

To better understand the effectiveness of expensive cancer medicines in a real-world population, the Dutch Institute for Clinical Auditing (DICA) initiated the Medicines Program in 2018. The program aims to identify variation in use and clinical outcomes of expensive medicines, provide postmarketing authorization data, provide a tool for clinicians to benchmark their practice on the use of expensive medicines, and stimulate interactions between clinicians to share best practices. In this program existing data sources were used. This study aims to describe the potential for the addition of declaration data to quality registries to provide participating centers with benchmark information about the use of medicines and associated outcomes.

Methods

Ethics Approval

In compliance with Dutch regulations, the DICA quality registries were approved by the medical ethical committee of the Leiden University Medical Center and was not subject to the Medical Research Involving Human Subjects Act.

Data Sources

Different existing data sources were used in the DICA Medicines Program; these data sources were linked. The first data sources were national population-based registries that are managed by the DICA. The DICA is a nonprofit organization that facilitates 23 population-based registries on different disciplines and diseases. These registries include information on clinical characteristics but contain limited data on the use of medicines. The DICA Medicines Program uses the Dutch Colorectal Audit (DCRA) [10], the Dutch Lung Cancer Audit [13], and the National Breast Cancer Organization Breast Cancer Audit (NBCA) [14]. These quality registries include information on patient, tumor, and treatment characteristics, and are used to compare hospitals on structure, processes, and clinical outcomes [15,16]. A previous study has shown that the data entered in the DICA registries are accurate and complete [17].

The second data source was financial and administrative data, including hospital pharmacies' declarations of expensive medicines for health insurers. These expensive medicines are listed as expensive (>€1000 per patient per year, equivalent to >US $1058.39) by the Dutch Healthcare Authority [18]. This data source includes precise and valid information about the diagnosis, date of prescription, dose, and quantity of a prescribed drug. Administrative data from hospitals include declarations for the reimbursement of expensive medicines. Only expensive medicines that were relevant and related to the diagnosis were linked to the clinical data.

The third data source includes the Dutch diagnosis treatment combinations (DBC) information system, which contains information on in-hospital activities, such as computerized tomography (CT) scans, infusions, hospital admissions, day treatments, and radiology treatments. The DBC information system is used for the registration and reimbursement of hospital and medical specialist care. This system was introduced in the Netherlands to increase the transparency of care. Furthermore, DBC information systems were initiated to create a supply-led system, increase efficiency, and facilitate competition between health care providers [19]. Because the DCRA and NBCA quality registries only include patients undergoing surgical operations, patients with metastasized cancers who do not undergo surgical operations are missing. To include patients with metastasized colorectal and metastasized breast cancer, the DBC data were used and linked to the fourth data source.

The fourth data source was survival data from the national claims database (VEKTIS) from health insurers [20]. VEKTIS is the national insurance database, which contains administrative data from Dutch national health care insurers, covering approximately 17 million individuals. By adding this data source, we could assess overall survival (OS) from diagnosis

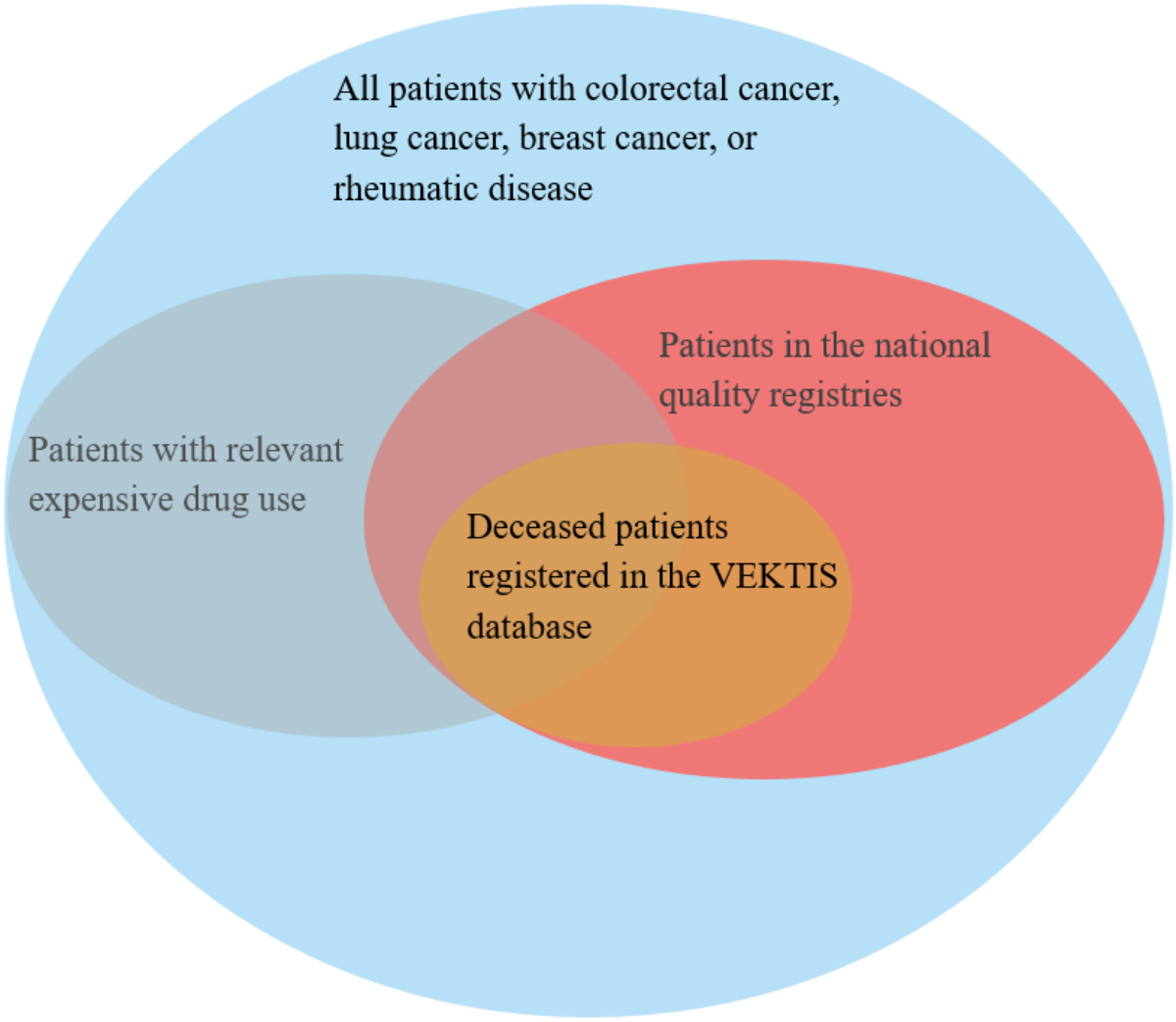
XSL•FO
RenderX

and the start of systemic therapy. Data were retrospectively collected from patients treated from 2017 to 2020. Although the DICA Medicines Program was established in 2018, data from 2017 were available from the hospitals and were therefore linked.

Data Linkage and Privacy

The first step in data linkage was identifying patients diagnosed with colorectal cancer, lung cancer, breast cancer, and rheumatic disease using the DBC information system. The DBC information system is used for the registration and reimbursement of health care in the Netherlands. The second step was to identify whether these patients used relevant expensive drugs, and the third step was to determine whether these patients are registered in the national quality registry. Information on the date of death from the VEKTIS database was added for deceased patients (Figure 1). Data were linked based on hospital patients' ID. A third party pseudonymized patient IDs. The results were visualized in dynamic web-based dashboards in which (systemic) treatments were linked to clinical parameters. Filters on patient and tumor characteristics, clinical outcomes, and therapy varied for the different diagnoses, depending on relevance. Furthermore, participating hospitals were compared, and practice variation was visualized and discussed to share knowledge on medical treatment differences.

Figure 1. Visualization of the patients included in our study and the different data sources used.



Statistical Analysis

The analyses in this manuscript are exploratory. Descriptive statistics were used to assess patient, tumor, and treatment characteristics. Time-to-next-treatment (TTNT) and OS were estimated with the Kaplan-Meier method. Survival times were calculated from the start of a systemic therapy until subsequent treatment (TTNT) or until death from any cause (OS). Patients who were alive or lost to follow-up were right censored at the time of their last registered expensive medicine use. All the statistical data were analyzed using R (version 4.0.2; R Foundation for Statistical Computing) within the RStudio environment (version 3.5.2; RStudio PB; packages tidyverse [21], TableOne [22], Survminer [23]).

Results

Database

A total of 21 Dutch hospitals participated in the DICA Medicines Program and were included in this study. Of these hospitals, 9 were top clinical hospitals, 11 were peripheral

XSL•FO
RenderX

hospitals, and 1 was an academic hospital. The geographic location of these hospitals is shown in Figure 2, which indicates they are spread across the country. The DICA Medicines database included a total of 7412 patients with colorectal cancer, 1981 patients with metastasized colon cancer, 3860 patients with lung cancer, 1253 patients with metastasized breast cancer, and 7564 patients with rheumatic disease.

Figure 2. A map of the Netherlands including the geographic location of the participating hospitals in the Dutch Institute for Clinical Auditing Medicines Program (red dots).

Benchmarking

The DICA Medicines Program provides the ability to compare results between hospitals to improve the quality of care provided. Hospitals were provided with web-based dynamic dashboards (Multimedia Appendix 1), continuously comparing their data to the benchmark. The benchmark consisted of all other participating hospitals. An example of benchmarking is the use of systemic therapies at the end of life in patients with metastatic colorectal cancer. This varied between hospitals from 4.2% (5/119) to 27.8% (5/18), with a median of 13.4%. The dashboards also provide information on the type of systemic therapy used at the end of life. A signaling function is included in the dashboard if hospitals deviate from the benchmark (Multimedia Appendix 2). Deviation from the benchmark was defined as a ranked average calculated as follows: (Percentage of cases within hospital X – Percentage of cases within the benchmark)^2 + Total number of patients in the benchmark.

Use of Medicines and Patient Characteristics

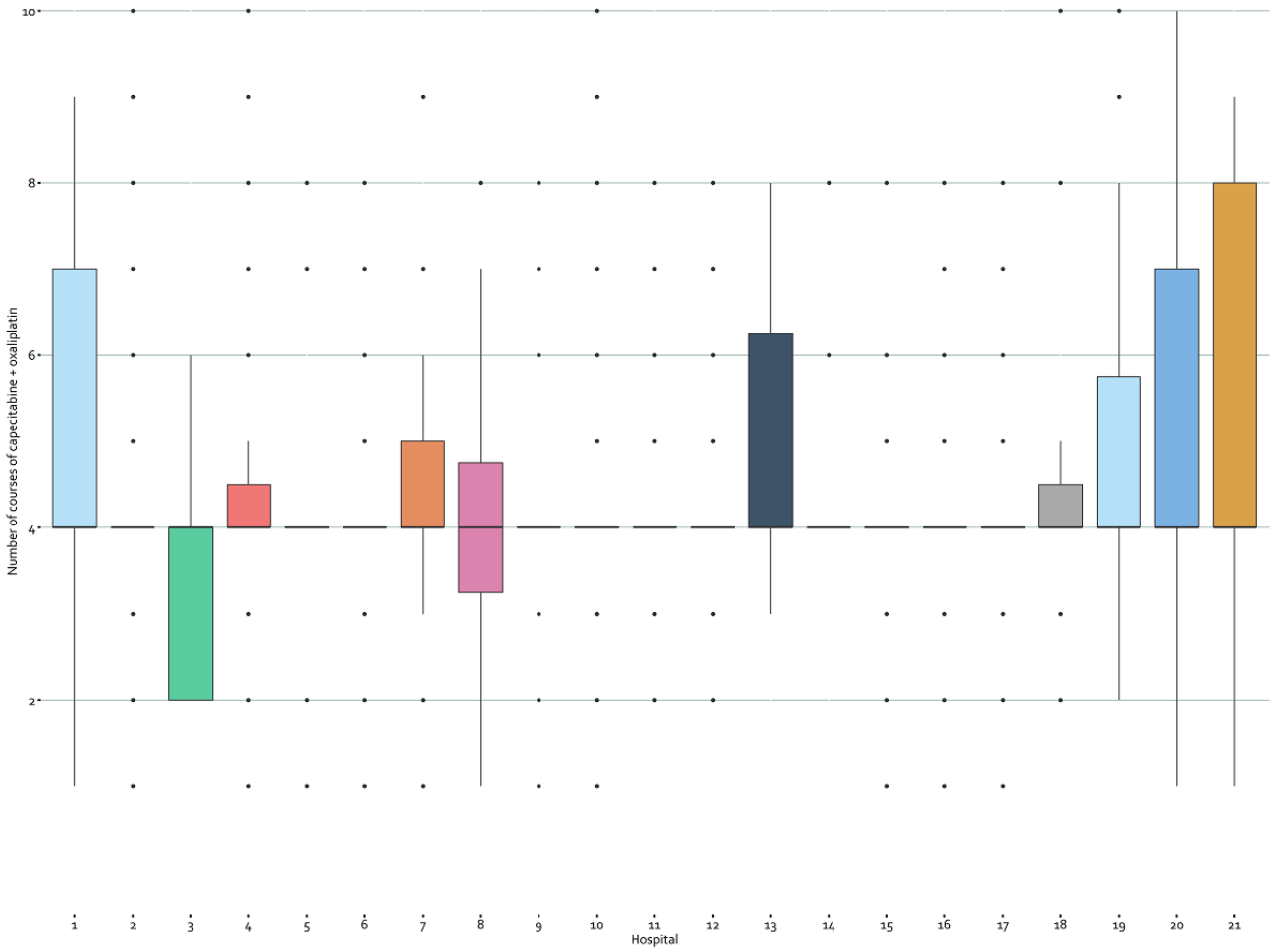
The linkage of different data sources led to new insights into hospitals' use of medicines and patient populations. The patient and tumor characteristics are listed for each medicine in the dashboard as a table that hospitals can compile with available variables. One of the participating hospitals discovered a deviation from the benchmark in the percentage of mesothelioma using the dashboard (Multimedia Appendix 3). This was 9.2% (14/153) for the specific hospital, compared to only 18% (105/3480) in the benchmark.

XSL•FO
RenderX

Treatment Information

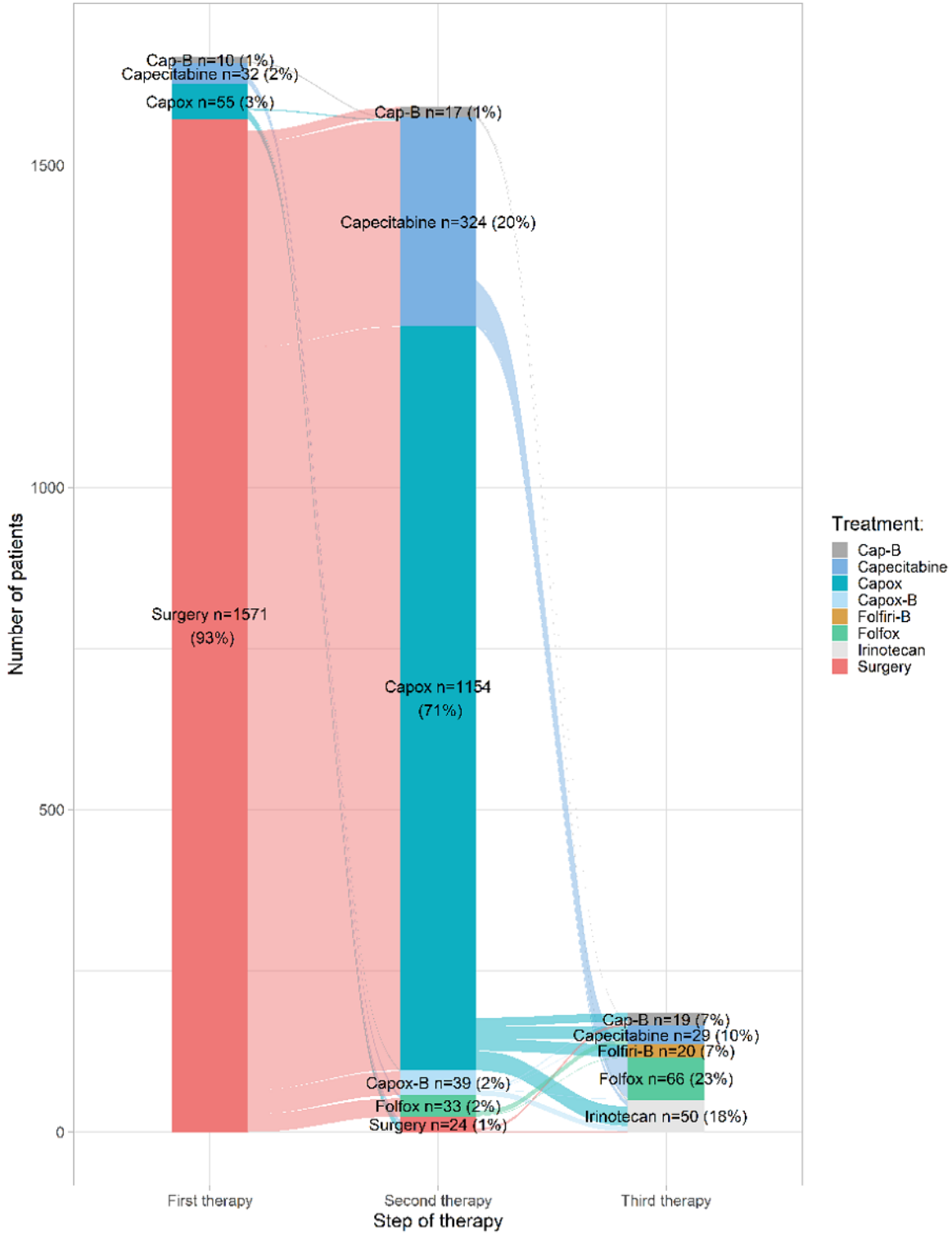
Linking clinical data to systemic treatment information also led to detailed information for each medicine, such as treatment duration in months and the number of cycles per patient. An example is the number of courses of capecitabine and oxaliplatin for the adjuvant treatment of colorectal cancer (Figure 3). Furthermore, administrative data were used to visualize treatment steps in Sankey diagrams in the dashboard, which can be adjusted for specific filters on patient, tumor, and treatment characteristics. Figure 4 shows the Sankey diagram for patients with metastasized colon cancer who were treated between 2017 and 2020. The dashboards also contain detailed information on diagnostic imaging (CT and magnetic resonance imaging scans), the number of consults (or teleconsults), clinical admissions, and emergency room visits pre- and post treatment for each medicine (Multimedia Appendix 4).

Figure 3. Number of courses of capecitabine + oxaliplatine for the adjuvant treatment of colorectal cancer patients per hospital.



XSL•FO
RenderX

Figure 4. Treatment patterns in patients with stage III colon cancer treated between 2017 and 2020 (N=1668). The Sankey diagram shows the flow of patients from the first treatment step to the second treatment step and from the second treatment step to the third treatment step. The width of the lines corresponds with the number of patients. Systemic therapies with less than 5 patients are not displayed in this graph. Cap-B: Capecitabine plus bevacizumab; Capox: Capecitabine plus oxaliplatin; Capox-B: Capecitabine plus oxaliplatin plus bevacizumab; Folfiri-B: Fluorouracil plus irinotecan plus bevacizumab; Folfox: Fluorouracil plus oxaliplatin.



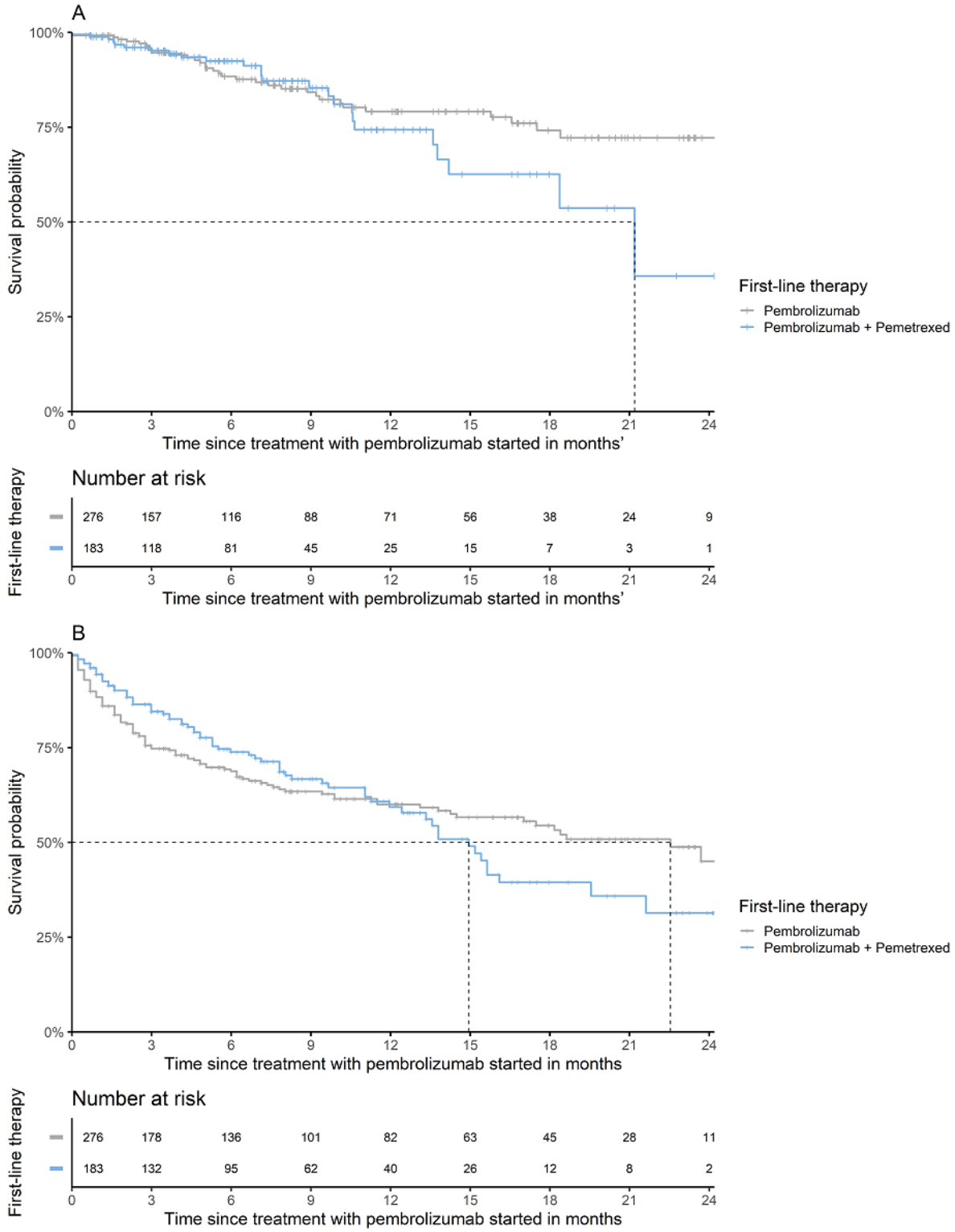
Clinical Outcomes

The DICA Medicines Program also provides hospitals with data on clinical outcomes, such as TTNT and OS. Figure 5 shows the TTNT of patients with metastasized lung cancer treated with first-line pembrolizumab or pembrolizumab and pemetrexed combination therapy. The median TTNT was 22.5 (95% CI 17.0, upper range not available) months and 14.9 (95% CI 12.4-21.6) months for pembrolizumab monotherapy and the combination of pembrolizumab and pemetrexed, respectively. The OS of these treatments is presented in Figure 5. Each hospital's outcomes are compared to the benchmark. It is also

XSL•FO
RenderX

possible to compare clinical outcomes between treatments in exploratory head-to-head comparisons for hospitals specifically or the benchmark in specific patient populations.

Figure 5. (A) Time-to-next-treatment of lung cancer patients treated with first-line pembrolizumab or pembrolizumab + pemetrexed between 2017 and 2020 and (B) overall survival of lung cancer patients treated with first-line pembrolizumab or pembrolizumab + pemetrexed between 2017 and 2020.



Costs

Use of the financial and administrative database from hospital pharmacies provided us with access to detailed information on the costs of systemic therapies (total costs per treatment and costs per patient) in certain subgroups. Hospitals can upload their paid prices to the dashboard, which is then connected to the medicine and patient information (Multimedia Appendix 5). Prices paid by other hospitals are not shown due to

XSL•FO
RenderX

confidential agreements between pharmaceutical companies and hospitals.

Discussion

Principal Findings

This paper reports on the initial results on the potential applications of data from the DICA Medicines Program; in this program, RWD are generated by linking 4 data sources, including data from quality registries, financial pharmacy data, in-hospital activities systems data, and reimbursement data from 21 Dutch hospitals. In this paper, we reported on the potential of this program in terms of benchmarking, treatment information, clinical outcomes, and costs. To be able to use the data as benchmark information, the data were visualized in web-based dashboards available to clinicians, insurers, and researchers; this led to insights on medication use, clinical outcomes, and costs without any additional registration burden for hospitals. Benchmarking hospital performance is relatively uncommon in the field of medical oncology in contrast to surgical oncology, where many quality registries exist that monitor the quality of care in every hospital [10]. Benchmarking information can support hospital pharmacists, oncologists, and other medical professions involved in the systemic treatment of patients to reach a certain level of care. RWD on the use and efficacy of systemic therapies are needed in daily clinical practice. As the real-world setting differs from the RCT setting, these data are needed after marketing authorization. This project provides real-world evidence, for which there is growing interest. One should be cautious when making definitive conclusions based on observational data. Minor observed differences could be the result of unknown confounding factors [24]. Other initiatives on the linkage of administrative data are similar and link patient-centered health data such as patient-reported outcome measures and clinical laboratory measurements but involve small patient groups [25] or limited patient and tumor characteristics [12].

Strengths

First, data are validated at the time of delivery from the hospitals with the clinicians. A lot of effort is put into the validation of the algorithms that are used in the dashboards, for example, in building Sankey diagrams for treatment sequences in specific patient populations. The second strength of the DICA Medicines Program is the use of existing data sources, thereby minimizing the extra registration burden for medical specialists. This strategy could also be used by other parties to minimize registration burden and maximize the value of available RWD sources. Variables that could easily be derived from the declaration data were the number of expenses, start dates of medications, and the total dosages. Third, the program consists of many participating hospitals within a widespread geographic location, resulting in the inclusion of many patients, who are representative of the Dutch population. Another strength is the linkage of survival data to the other data sources. The database from the national health insurers is a valid source as health care insurance coverage stops when a patient dies. The final strength is that the data are up-to-date and representative of the current situation. This is especially valuable in situations such as the COVID-19 pandemic, where the systemic treatment of some patients with cancer was adjusted. Since the data are updated quarterly, it was possible to monitor the impact of COVID-19 in certain subgroups of patients in the dashboard. The DICA Medicines Program led to various insights into medication use. Questions related to the use of (expensive) medicines can be answered using the dashboards, in which users can select patient populations or treatments of interest.

Limitations and Future Perspectives

In the clinical registries used for this study, some indications had incomplete data. The DCRA only includes patients undergoing surgical operations, which leads to incomplete clinical data on patients with metastasized cancers and colorectal cancers. This was also the case for patients with metastasized breast cancer. In this subpopulation of patients with breast cancer, essential tumor information, such as receptor status, is lacking. In addition, we are unable to extract information about weight, response status, date of progression, or toxicities from the declaration data. These are mostly data registered in unstructured text in electronical medical records. However, our intention is to complement the clinical data of these patient groups with other techniques that do not lead to further registration burden, such as text mining. Second, due to privacy regulations in the Netherlands [26], it is not permitted to follow-up on patients when they are referred to other hospitals for treatment. An individual patient may seek a second opinion from another hospital. This may have led to incomplete treatment information and individual patients being included twice in the database. Especially for university hospitals, where many patients are referred, it is necessary to have the complete treatment information. Previous analyses on the entire population of patients with lung cancer showed this was the case in <5% of all patients in the Netherlands. In this study, there may be an overestimate in the number of patients but not the number of prescriptions as these are validated declarations made by the hospitals. Third, more information on patient and tumor characteristics is needed to allow for head-to-head comparisons of medicines. Registries should therefore include information on response status and detailed treatment-related toxicity within each line of treatment. At this moment, emergency room visits and hospital admissions are linked to the use of medications and presented in the dashboards. However, these are only surrogate outcomes and do not give insight into the exact response or toxicity. Adding more outcomes of systemic therapies will also be an opportunity for surgical quality registries to become multidisciplinary, where both surgeons and medical oncologists register specific patients' characteristics and outcomes. We are currently exploring text mining opportunities to add information on toxicities and response statuses to the quality registries.

Presently, hospitals use dashboards to benchmark their results against those of other hospitals and gain insights into the use of medications and patient populations, as we showed in this study. The dashboards can also be used for multiple other purposes and by different stakeholders in the future. First, dashboards and RWD can serve as communication tools between physicians and their patients. Based on specific patient and tumor characteristics, clinical outcomes can help patients better

XSL•FO
RenderX

understand their disease course and improve shared decision-making. Second, registration authorities can also benefit from data as presented in this study. Data on newly approved medicines used in clinical practice are included in financial pharmacy data and can be linked to population-based registries. Especially for postapproval measurements, this information is valuable in monitoring the safety and effectiveness of medicines [27]. This can, in certain cases, eventually lead to the replacement of postapproval clinical studies, which will save time and financial resources. The European Medicines Agency and US Food and Drug Agency are increasingly interested in RWD for the evaluation of medicines [28,29]. Furthermore, health care insurers are interested in these data for reimbursement and effective use of expensive medications in the real-world setting [30].

In the future, accurate data from DBC's and financial information could automatically prefill quality registry items. The DICA quality registry items are now entered manually, which is time-consuming and prone to registration errors. Reusing these data sources will lower the registration burden, reduce missing data, and validate data. These data can be used to complete registries and reduce hospital differences. Furthermore, RWD can also be used in health technology assessment decisions. This will be explored in the near future within European Union programs [31]. However, other data sources, such as pathology databases, must be linked to enrich the data. This additional data on histopathology and mutation status are essential as certain medications targeting specific mutations can influence outcomes. To improve shared decision-making, additional data sources, including patient-reported outcome measurements, must be linked to existing data sources.

Conclusions

The DICA Medicines Program has shown that it is possible to gather and link RWD sources pertaining to medicines. In addition, these data became available with minimal registration burden and effort for hospitals. This method of providing RWD can be used in other population-based registries. The DICA Medicines Program provided participating centers with benchmark information and tools to evaluate the effectiveness of expensive medicines in real-world settings.

Acknowledgments

The DICA Medicines program was supported by the Dutch Association for Hospital Pharmacists (NVZA) and the Foundation of Funds for Medical Specialists (grant 53052287) and by a Dutch health insurance company, Institute of Public Healthcare (VGZ). These grants were used for the initial costs of developing the dashboards. We would like to thank the NVZA and VGZ for their support.

Conflicts of Interest

AE has advisory relationships with Amgen Inc, Bristol Myers Squibb, F. Hoffmann-La Roche AG, Novartis International AG, MSD, Laboratoires Pierre Fabre, Sanofi SA, Pfizer Inc, Ipsen SA, and Merck & Co Inc, and has received research grants not related to this paper from Sanofi SA, F. Hoffmann-La Roche AG, Bristol Myers Squibb, Idera Inc, and Teva Pharmaceutical Industries Ltd. AE has received travel expenses from MSD Oncology, F. Hoffmann-La Roche AG, Pfizer Inc, and Sanofi SA, and has received speaker honoraria from Bristol Myers Squibb and Novartis International AG. All other authors declare no conflicts of interest.

Multimedia Appendix 1
The medication page of the dashboard shows the active ingredients in a selected patient population and the percentage of patients treated with that active ingredient versus the benchmark. This page also shows the trend over the years. This overview can be adjusted with the filters to show the use of medicines in a specific population or year. Furthermore, it is also possible to receive an overview of the medicines with the number of courses instead of usage. The patients list contains the data of all patients that are selected for the shown results.
[PNG File , 96 KB - jmir_v24i6e33446_app1.png]

Multimedia Appendix 2
The signals page of the dashboard shows the insights of the dashboard in which a hospital deviates from the benchmark. There are also some specific signals to stimulate improving patient care.
[PNG File , 105 KB - jmir_v24i6e33446_app2.png]

Multimedia Appendix 3
There is a deep-dive function for every medicine in the dashboard that shows detailed information about the use of these medicines. The deep-dive also shows baseline patient and tumor characteristics of patients treated with this specific medicine, compared to the benchmark.
[PNG File , 131 KB - jmir_v24i6e33446_app3.png]

XSL•FO
RenderX

Multimedia Appendix 4
Details on the diagnostics are included per medicine in the dashboard for the hospital versus the results in the benchmark. These are related to the moment of medicine use (pre- or post treatment).
[PNG File , 143 KB - jmir_v24i6e33446_app4.png]

Multimedia Appendix 5
The costs page of the dashboard shows the total costs and the costs per patient for the hospital and the benchmark. This page can be adjusted for specific patient populations by using the filters.
[PNG File , 116 KB - jmir_v24i6e33446_app5.png]

References

1. Naci H, Davis C, Savović J, Higgins JPT, Sterne JAC, Gyawali B, et al. Design characteristics, risk of bias, and reporting of randomised controlled trials supporting approvals of cancer drugs by European Medicines Agency, 2014-16: cross sectional analysis. BMJ 2019 Sep 18;366:l5221. [doi: 10.1136/bmj.l5221] [Medline: 31533922]
2. Jones DS, Podolsky SH. The history and fate of the gold standard. Lancet 2015 May 18;385(9977):1502-1503. [doi: 10.1016/S0140-6736(15)60742-5] [Medline: 25933270]
3. Hatswell AJ, Baio G, Berlin JA, Irs A, Freemantle N. Regulatory approval of pharmaceuticals without a randomised controlled study: analysis of EMA and FDA approvals 1999-2014. BMJ Open 2016 Jun 30;6(6):e011666 [FREE Full text] [doi: 10.1136/bmjopen-2016-011666] [Medline: 27363818]
4. Booth CM, Eisenhauer EA. Progression-free survival: meaningful or simply measurable? J Clin Oncol 2012 May 01;30(10):1030-1033. [doi: 10.1200/JCO.2011.38.7571] [Medline: 22370321]
5. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet 2005;365(9453):82-93. [doi: 10.1016/S0140-6736(04)17670-8] [Medline: 15639683]
6. Uitgaven aan dure geneesmiddelen stijgen naar 2,1 miljard. Phamaceutisch Weekblad. 2019. URL: https://www.pw.nl/nieuws/2019/uitgave-dure-medicatie-stijgt-naar-20ac-2-1-miljard [accessed 2022-05-27]
7. Cramer-van der Welle CM, Peters BJ, Schramel FM, Klungel OH, Groen HJ, van de Garde EM, Santeon NSCLC Study Group, Santeon NSCLC study group. Systematic evaluation of the efficacy-effectiveness gap of systemic treatments in metastatic nonsmall cell lung cancer. Eur Respir J 2018 Dec;52(6):1801100 [FREE Full text] [doi: 10.1183/13993003.01100-2018] [Medline: 30487206]
8. van Zeijl MCT, Ismail RK, de Wreede LC, van den Eertwegh AJM, de Boer A, van Dartel M, et al. Real-world outcomes of advanced melanoma patients not represented in phase III trials. Int J Cancer 2020 Dec 15;147(12):3461-3470. [doi: 10.1002/ijc.33162] [Medline: 32559817]
9. Beck N, van Bommel AC, Eddes E, van Leersum NJ, Tollenaar R, Wouters M, Dutch Clinical Auditing Group. The Dutch Institute for Clinical Auditing: achieving Codman's dream on a nationwide basis. Ann Surg 2020 Apr;271(4):627-631. [doi: 10.1097/SLA.0000000000003665] [Medline: 31972639]
10. Van Leersum N, Snijders H, Henneman D, Kolfschoten N, Gooiker G, ten Berge M, Dutch Surgical Colorectal Cancer Audit Group, et al. The Dutch surgical colorectal audit. Eur J Surg Oncol 2013 Oct;39(10):1063-1070. [doi: 10.1016/j.ejso.2013.05.008] [Medline: 23871573]
11. Hoeijmakers F, Beck N, Wouters MWJM, Prins HA, Steup WH. National quality registries: how to improve the quality of data? J Thorac Dis 2018 Oct;10(Suppl 29):S3490-S3499. [doi: 10.21037/jtd.2018.04.146] [Medline: 30510784]
12. van Herk-Sukel MP, van de Poll-Franse LV, Lemmens VE, Vreugdenhil G, Pruijt JF, Coebergh JWW, et al. New opportunities for drug outcomes research in cancer patients: the linkage of the Eindhoven Cancer Registry and the PHARMO Record Linkage System. Eur J Cancer 2010 Jan;46(2):395-404. [doi: 10.1016/j.ejca.2009.09.010] [Medline: 19811904]
13. Ismail R, Schramel F, van Dartel M, Hilarius D, de Boer A, Wouters M, Dutch Lung Cancer Audit Scientific Committee. The Dutch Lung Cancer Audit: nationwide quality of care evaluation of lung cancer patients. Lung Cancer 2020 Nov;149:68-77 [FREE Full text] [doi: 10.1016/j.lungcan.2020.08.011] [Medline: 32979634]
14. van Bommel AC, Spronk PE, Vrancken Peeters MT, Jager A, Lobbes M, Maduro JH, NABON Breast Cancer Audit. Clinical auditing as an instrument for quality improvement in breast cancer care in the Netherlands: the national NABON Breast Cancer Audit. J Surg Oncol 2017 Mar;115(3):243-249. [doi: 10.1002/jso.24516] [Medline: 27885679]
15. Spiegelhalter DJ. Funnel plots for comparing institutional performance. Stat Med 2005 May 30;24(8):1185-1202. [doi: 10.1002/sim.1970] [Medline: 15568194]
16. Rakow T, Wright RJ, Spiegelhalter DJ, Bull C. The pros and cons of funnel plots as an aid to risk communication and patient decision making. Br J Psychol 2015 May;106(2):327-348. [doi: 10.1111/bjop.12081] [Medline: 25123852]
17. van der Werf LR, Voeten SC, van Loe CMM, Karthaus EG, Wouters MWJM, Prins HA. Data verification of nationwide clinical quality registries. BJS Open 2019 Dec;3(6):857-864 [FREE Full text] [doi: 10.1002/bjs5.50209] [Medline: 31832593]
18. NZA. Nederlandse Zorgautoriteit. URL: https://www.nza.nl/english [accessed 2022-05-31]
19. Oostenbrink JB, Rutten FFH. Cost assessment and price setting of inpatient care in The Netherlands. The DBC case-mix system. Health Care Manag Sci 2006 Aug;9(3):287-294. [doi: 10.1007/s10729-006-9096-y] [Medline: 17016935]

XSL•FO
RenderX

20. Vektis. 2021. URL: https://www.vektis.nl/ [accessed 2022-05-27]
21. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. J Open Source Software 2019 Nov;4(43):1686. [doi: 10.21105/joss.01686]
22. Yoshida K, Bartel A. tableone: create 'Table 1' to describe baseline characteristics with or without propensity score weights. The Comprehensive R Archive Network. 2020. URL: https://cran.r-project.org/package=tableone [accessed 2022-05-27]
23. Kassambra A, Kosinski M, Biecek P. survminer: drawing survival curves using 'ggplot2'. The Comprehensive R Archive Network. 2020. URL: https://cran.r-project.org/package=survminer [accessed 2022-05-27]
24. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-world evidence - what is it and what can it tell us? N Engl J Med 2016 Dec 08;375(23):2293-2297. [doi: 10.1056/NEJMsb1609216] [Medline: 27959688]
25. Dhruva SS, Ross JS, Akar JG, Caldwell B, Childers K, Chow W, et al. Aggregating multiple real-world data sources using a patient-centered health-data-sharing platform. NPJ Digit Med 2020;3:60. [doi: 10.1038/s41746-020-0265-z] [Medline: 32352038]
26. Quintel T. Article 29 Data Protection Working Party Opinion on the Law Enforcement Directive. European Commission. 2018. URL: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp131_en.pdf [accessed 2022-05-27]
27. Ismail R, Sikkes N, Wouters M, Hilarius D, Pasmooij M, van den Eertwegh F, et al. 1120P Post-approval trials versus patient registries: Comparability of advanced melanoma patients with brain metastases. Ann Oncol 2020 Sep 01;31(1):S754-S766 [FREE Full text] [doi: 10.1016/j.annonc.2020.08.1243]
28. Patient registries. European Medicines Agency. 2015 Sep. URL: https://www.ema.europa.eu/en/human-regulatory/post-authorisation/patient-registries [accessed 2022-05-27]
29. Use of real-world evidence to support regulatory decision-making for medical devices: guidance for industry and food and drug administration staff. US Food and Drug Administration. 2017. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-real-world-evidence-support-regulatory-decision-making-medical-devices [accessed 2022-05-27]
30. Roberts MH, Ferguson GT. Real-world evidence: bridging gaps in evidence to guide payer decisions. Pharmacoecon Open 2021 Mar;5(1):3-11. [doi: 10.1007/s41669-020-00221-y] [Medline: 32557235]
31. New methods for the effective use of real-world data and/or synthetic data in regulatory decision-making and/or in health technology assessment. European Commission. URL: https://www.ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/horizon-hlth-2022-tool-11-02 [accessed 2022-05-27]

Abbreviations

CT: computerized tomography
DBC: diagnosis treatment combinations
DCRA: Dutch Colorectal Audit
DICA: Dutch Institute for Clinical Auditing
NBCA: National Breast Cancer Organization Breast Cancer Audit
NVZA: Dutch Association for Hospital Pharmacists
OS: overall survival
RCT: randomized controlled trial
RWD: real-world data
TTNT: time-to-next-treatment
VGZ: Institute of Public Healthcare

Edited by JMIRPE Office, G Eysenbach; submitted 29.09.21; peer-reviewed by Z Su, R Halkes; comments to author 09.02.22; revised version received 06.04.22; accepted 30.04.22; published 23.06.22.

Please cite as:
Ismail RK, van Breeschoten J, van der Flier S, van Loosen C, Pasmooij AMG, van Dartel M, van den Eertwegh A, de Boer A, Wouters M, Hilarius D
Medication Use and Clinical Outcomes by the Dutch Institute for Clinical Auditing Medicines Program: Quantitative Analysis
J Med Internet Res 2022;24(6):e33446
URL: https://www.jmir.org/2022/6/e33446
doi:10.2196/33446
PMID:35737449

XSL•FO
RenderX

©Rawa Kamaran Ismail, Jesper van Breeschoten, Silvia van der Flier, Caspar van Loosen, Anna Maria Gerdina Pasmooij, Maaike van Dartel, Alfons van den Eertwegh, Anthonius de Boer, Michel Wouters, Doranne Hilarius. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 23.06.2022. 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

Effectiveness of Recruitment Strategies of Latino Smokers: Secondary Analysis of a Mobile Health Smoking Cessation Randomized Clinical Trial

Evelyn Arana-Chicas[1*], DrPH; Francisco Cartujano-Barrera[2*], MD; Katherine K Rieth[3*], MD, MA; Kimber K Richter[4*], MPH, PhD; Edward F Ellerbeck[4*], MD, MPH; Lisa Sanderson Cox[4*], PhD; Kristi D Graves[5*], PhD; Francisco J Diaz[6*], PhD; Delwyn Catley[7*], PhD; Ana Paula Cupertino[2*], PhD

[1]Department of Surgery, University of Rochester School of Medicine & Dentistry, Rochester, NY, United States

[2]Department of Public Health Sciences, University of Rochester School of Medicine & Dentistry, Rochester, NY, United States

[3]Department of Otolaryngology, University of Rochester School of Medicine & Dentistry, Rochester, NY, United States

[4]Department of Population Health, The University of Kansas Medical Center, Kansas City, KS, United States

[5]Department of Oncology, Georgetown University Medical Center, Washington DC, DC, United States

[6]Department of Biostatistics & Data Science, The University of Kansas Medical Center, Kansas City, KS, United States

[7]Center for Children's Healthy Lifestyle and Nutrition, Children's Mercy Kansas City, Kansas City, MO, United States

[*]all authors contributed equally

Corresponding Author:
Evelyn Arana-Chicas, DrPH
Department of Surgery
University of Rochester School of Medicine & Dentistry
265 Crittenden Blvd
Rochester, NY, 14642
United States
Phone: 1 585 287 4217
Email: Evelyn_Arana@urmc.rochester.edu

Abstract

Background: Latinos remain disproportionately underrepresented in clinical trials, comprising only 2%-3% of research participants. In order to address health disparities, it is critically important to increase enrollment of Latino smokers in smoking cessation trials. There is limited research examining effective recruitment strategies for this population.

Objective: The purpose of this study was to compare the effectiveness of direct versus mass and high- versus low-effort recruitment strategies on recruitment and retention of Latino smokers to a randomized smoking cessation trial. We also examine how the type of recruitment might have influenced the characteristics of enrolled participants.

Methods: Latino smokers were enrolled into *Decídetexto* from 4 states—New Jersey, Kansas, Missouri, and New York. Participants were recruited from August 2018 until March 2021. Mass recruitment strategies included English and Spanish advertisements to the Latino community via flyers, Facebook ads, newspapers, television, radio, church bulletins, and our *Decídetexto* website. Direct, high-effort strategies included referrals from clinics or community-based organizations with whom we partnered, in-person community outreach, and patient registry calls. Direct, low-effort strategies included texting or emailing pre-existing lists of patients who smoked. A team of trained bilingual (English and Spanish) recruiters from 9 different Spanish-speaking countries of origin conducted recruitment, assessed eligibility, and enrolled participants into the trial.

Results: Of 1112 individuals who were screened, 895 (80.5%) met eligibility criteria, and 457 (457/895, 51.1%) enrolled in the trial. Within the pool of screened individuals, those recruited by low-effort recruitment strategies (both mass and direct) were significantly more likely to be eligible (odds ratio [OR] 1.67, 95% CI 1.01-2.76 and OR 1.70, 95% CI 0.98-2.96, respectively) and enrolled in the trial (OR 2.60, 95% CI 1.81-3.73 and OR 3.02, 95% CI 2.03-4.51, respectively) compared with those enrolled by direct, high-effort strategies. Among participants enrolled, the retention rates at 3 months and 6 months among participants recruited via low-effort strategies (both mass and direct) were similar to participants recruited via direct, high-effort methods. Compared with enrolled participants recruited via direct (high- and low-effort) strategies, participants recruited via mass strategies were less likely to have health insurance (44.0% vs 71.2% and 71.7%, respectively; P<.001), lived fewer years in the United

XSL•FO
RenderX

States (22.4 years vs 32.4 years and 30.3 years, respectively; *P*<.001), more likely to be 1st generation (92.7% vs 76.5% and 77.5%, respectively; *P*=.007), more likely to primarily speak Spanish (89.3% vs 65.8% and 66.3%, respectively), and more likely to be at high risk for alcohol abuse (5.8 mean score vs 3.8 mean score and 3.9 mean score, respectively; *P*<.001).

Conclusions: Although most participants were recruited via direct, high-effort strategies, direct low-effort recruitment strategies yielded a screening pool more likely to be eligible for the trial. Mass recruitment strategies were associated with fewer acculturated enrollees with lower access to health services—groups who might benefit a great deal from the intervention.

Trial Registration: ClinicalTrials.gov identifier: NCT03586596; https://clinicaltrials.gov/ct2/show/NCT03586596

International Registered Report Identifier (IRRID): RR2-10.1016/j.cct.2020.106188

(J Med Internet Res 2022;24(6):e34863) doi:10.2196/34863

KEYWORDS

smoking cessation; Latino health, Latino recruitment; health disparities; participant recruitment

Introduction

Latinos constitute the largest minority group in the United States, representing 18.5%, or 55 million, of the current US population [1], and this group is projected to grow to 30% by 2060 [2]. An estimated 6 million Latinos (9.8%) in the United States are current cigarette smokers [3,4]. Although the National Institutes of Health (NIH) Revitalization Act of 1993 [5,6] called for the inclusion of minorities in clinical research, Latinos remain disproportionately underrepresented in clinical trials, comprising only 2%-3% of research participants [7-9].

Increasing enrollment of Latino smokers in smoking cessation trials is critical for addressing health disparities; however, there is limited research examining effective strategies for recruiting this population [10,11]. Common obstacles to recruitment may include language barriers and health literacy [12,13], and some Latinos may have concerns or mistrust of government-funded research related to privacy or deportation concerns [7,13-15]. Increased burden from social conditions such as poverty [16], low education levels [16,17], and immigration issues [18] also contribute to low participation in clinical trials. These reported barriers may lead to the perception that recruitment of Latinos into clinical trials is difficult. However, despite these barriers, when invited to participate in research, enrollment rates of Latinos are comparable to those of non-Latino Whites [7,14]. Indeed, Latinos are interested in enrolling in research when recruitment strategies are culturally and linguistically tailored to them.

Literature on the recruitment of Latinos into clinical trials has described the use of different recruitment strategies [19-23]. Some studies have recruited Latinos through proactive recruitment in which study staff directly contact individual potential participants [20,24] and reactive recruitment in which studies disseminate information via mass media and potential participants must contact the study themselves [19,21,25]. Often, recruitment studies emphasize including Latino researchers, fostering community connections to build trust, and using culturally and linguistically tailored recruitment materials delivered through culturally appropriate outlets such as Latino newspapers [19,22,23].

Traditional categorizations of recruitment approaches (eg, into proactive versus reactive) do not capture the complexity of current recruitment strategies. Although proactive recruitment strategies involving personal outreach to individuals have historically necessitated relatively high effort compared with reactive outreach efforts such as mass advertising, the advent of electronic communications such as text messages and emails now allows direct, personalized outreach with relatively low effort. To date, there has been a lack of research distinguishing the effects of direct versus mass outreach and level of effort on recruitment success. Furthermore, there are limited data available on the retention of Latinos who were recruited via different strategies in clinical trials. One study compared ethnic-specific retention rates in various clinical trials and found that Latino adults have a retention rate of ~54% in clinical trials, and this did not significantly differ compared with other ethnic groups [26]. Only one study has analyzed the effects of recruitment type on retention; however, the recruitment strategies used in that study were limited to newspapers, posters on buses and subways, study flyers at community organizations, and in-person recruitment and community organizations [24].

It is also possible that different recruitment strategies will yield participants with different characteristics. For example, compared with direct recruitment, mass media recruitment (eg, radio, flyers) may yield more inherently motivated participants since little outreach or encouragement is provided; those who reactively join the study following mass media exposure may have higher commitment to behavior change [27]. This study, therefore, calculated the associations of mass versus direct recruitment strategies, involving high and low study staff effort, with characteristics of Latino smokers who were screened, enrolled, and retained in a randomized smoking cessation trial—*Decídetexto* [28].

Methods

Study Design

This study is a secondary data analysis of *Decidetexto*, a mobile health (mHealth) smoking cessation randomized clinical trial. It compares the efficiency ratios for eligibility, enrollment, and retention (at 3 months and 6 months) of Latino smokers recruited via direct versus mass and high- versus low-effort recruitment strategies.

Ethical Approval

The details of the study intervention and protocol are described elsewhere [28]. Study procedures were approved and monitored

XSL•FO
RenderX

by Hackensack University Medical Center (#Pro2017-0528), the University of Rochester Medical Center (IRB #STUDY00005080), and the University of Kansas Medical Center Institutional Review Boards (IRB # KUMC IRB #STUDY00004475).

Recruitment

Latino smokers were enrolled into *Decídetexto* from multiple communities (both urban and rural) in 4 states—New Jersey, Kansas, Missouri, and New York. Participants were recruited from August 2018 until March 2021. Direct recruitment strategies involved one-on-one communication with identified Latino smokers and were dichotomized as either "direct, high-effort" or "direct, low-effort" strategies. Recruitment and eligibility were conducted by a team of trained bilingual (English and Spanish) recruiters from different countries of origin (eg, Cuba, Dominican Republic, Ecuador, El Salvador, Mexico, Nicaragua, Peru, Puerto Rico, Venezuela).

In this study, "recruitment method" refers broadly to either mass, direct high-effort, or direct low-effort recruitment methods. "Recruitment strategies" refer to the specific recruitment strategy implemented. Mass recruitment strategies did not rely on interpersonal communication but instead included bilingual (English and Spanish) advertisements of the study to the larger Latino community via flyers, Facebook ads, newspapers, television, radio, church bulletins, and the *Decídetexto* website. Direct, high-effort strategies required more staff resources to connect with potential participants and included personal calls based on referrals from clinics or community-based organizations (CBOs), in-person community outreach, and personal calls made to patients on patient registries. Furthermore, as reported in previous research [23], research staff adhered to important cultural values in their interactions with potential participants by communicating with *personalismo* (initiating warm conversations that conveyed care and understanding of the patient's circumstances), *simpatía* (not criticizing the patient), and *confianza* (establishing trust). Direct, low-effort strategies demanded less time and effort from the research team. Direct, low-effort strategies included sending emails and texts to patients on patient registries and referrals from family and friends. Direct, low-effort and mass strategies were similar in that interested participants had to take the step of contacting the study for screening and follow-up. In this sense, they are both "reactive" recruitment strategies. However, in this study, they are differentiated by whether the recruitment strategy used mass communication to the Latino community or was directly sent to an identified Latino smoker.

Measures

Research staff administered all study assessments either in person or via telephone. Prior to completing the eligibility questionnaire, participants were asked the open-ended question "How did you learn about the study?" The baseline assessment collected data on demographics (eg, gender, education, age, income, health insurance status, marital status), smoking characteristics (eg, cigarettes smoked per day; the number of past quit attempts), and biopsychosocial variables: eg, depressive symptoms via the Patient Health Questionnaire-2 (PHQ-2) scale [29], alcohol use via the Alcohol Use Disorders Identification Test-2 (AUDIT-2) [30], anxiety via the Generalized Anxiety Disorder-2 (GAD-2) [31], self-efficacy [32], and acculturation measures including years lived in the United States, primary language, generation, and region of origin.

Analyses

Logistic regression analyses were used to calculate odds ratios (ORs; efficiency ratios) and 95% CIs for associations (1) between recruitment method and obtaining eligible individuals among screened individuals, (2) between recruitment method and enrolling the screened participants, and (3) between recruitment method and retaining the enrolled participants at the 3-month and 6-month follow-up visits. Rates of eligibility, enrollment, and retention across the 3 recruitment methods and recruitment strategies were compared using chi-square tests. For each recruitment method, characteristics of enrolled participants were summarized with percentages for categorical variables and with means and SDs for continuous variables. Differences in categorical variables were exploratorily compared using Pearson chi-square tests while differences in continuous variables were compared using 1-way ANOVA tests. Reasons for ineligibility were compared between participants who were recruited via direct, high-effort; direct, low-effort; and mass recruitment methods using Pearson chi-square tests or Fisher exact tests. Data were analyzed using SPSS version 25.

Results

Overview

Of 1112 individuals who completed screening, 895 (80.5%) met eligibility criteria, and 457 (457/895, 51.1%) enrolled in the trial. The majority of participants were enrolled via direct, high-effort strategies (300/457, 65.6%). Table 1 lists the numbers screened, eligible, enrolled, and retained at 3 months and 6 months by recruitment method and includes efficiency ratios for eligibility, enrollment, and retention at 3 months and 6 months.

Table 2 shows the efficiency of specific recruitment strategies. Overall, eligibility efficiency ratios were lowest for Facebook ads (66.7%), followed by in-person community outreach (74.7%), our Decidetexto website (75.0%), and patient registry calls (79.1%). Enrollment efficiency ratios were lowest for Facebook ads (33.3%), followed by television (41.9%), in-person community outreach (31.9%), and patient registry calls (35.4%). The 3-month retention efficiency ratios were lowest for the Decidetexto website (66.7%), television (69.2%), and patient registry text (70.6%). The 6-month retention efficiency ratios were lowest for television (76.9%) and patient registry text (76.5%).

Compared with the direct, high-effort recruitment method, individuals screened in both the mass and direct, low-effort recruitment methods were significantly more likely to be eligible (OR 1.67, 95% CI 1.01-2.76 and OR 1.70, 95% CI 0.98-2.96, respectively) and enrolled (OR 2.60, 95% CI 1.81-3.73 and OR 3.02, 95% CI 2.03-4.51, respectively; Table 3). Of participants enrolled, those recruited via mass and direct, low-effort methods were just as likely to be retained at 3 months and 6 months compared with participants recruited via the direct, high-effort

XSL•FO
RenderX

method. Furthermore, given that 45.5% (208/457) of all enrolled participants were recruited via patient registry calls and that 69.3% (208/300) of all participants who were recruited via direct, high-effort strategies were recruited via patient registry calls, a logistic regression model was run to identify any differences in efficiency ratios between patient registry calls and other direct, high-effort strategies. No differences were found between the 2 (data not shown).

Table 1. Efficiency ratios for personalized and nonpersonalized recruitment methods.

Recruitment method	Number screened[a]	Number eligible[b]	Number enrolled	Number retained at 3 months	Number retained at 6 months	Eligibility efficiency ratio[c], %	Enrollment efficiency ratio[d], %	3-month retention efficiency ratio[e], %	6-month retention efficiency ratio[f], %
Mass	144	127	84	69	73	88.2	58.3	82.1	86.9
Direct, low effort	117	101	73	60	61	86.3	62.4	82.2	83.6
Direct, high effort	847	666	300	261	261	78.6	35.4	87.0	87.0

[a]The total is not 1112 because of missing data on the recruitment strategy.

[b]The total is not 895 because of missing data on the recruitment strategy.

[c]Ratio of number eligible to number screened.

[d]Ratio of number enrolled to number screened.

[e]Ratio of number retained at 3 months to number enrolled.

[f]Ratio of number retained at 6 months to number enrolled.

Table 2. Recruitment efficiency of specific recruitment strategies.

Recruitment method	Proportion for the recruitment strategies, n (%)	Number screened	Number eligible	Number enrolled	Number retained at 3 months	Number retained at 6 months	Eligibility efficiency ratio, %	Enrollment efficiency ratio, %	3-month retention efficiency ratio, %	6-month retention efficiency ratio, %
Mass (n=144)										
Church bulletin	4 (2.8)	4	4	4	4	3	100	100	100	75.0
Newspaper	22 (15.3)	22	22	16	14	14	100	72.7	87.5	87.5
Radio	32 (22.2)	32	28	18	14	15	87.5	56.3	77.8	83.3
Flyer	48 (33.3)	48	40	29	25	27	83.3	60.4	86.2	93.1
Decídetexto website	4 (2.8)	4	3	3	2	3	75.0	75.0	66.7	100
Television	31 (21.5)	31	24	13	9	10	77.4	41.9	69.2	76.9
Facebook ads	3 (2.1)	3	2	1	1	1	66.7	33.3	100	100
Direct, low effort (n=117)										
Clinic or CBO[a] email	10 (8.5)	10	9	7	7	6	90	100	100	85.7
Patient registry text	35 (29.9)	35	28	17	12	13	80.0	48.6	70.6	76.5
Friend or family referral	72 (61.6)	72	64	49	41	42	88.9	68.1	83.7	85.7
Direct, high effort (n=847)										
Clinic or CBO referral	81 (9.6)	81	68	35	30	29	83.9	43.2	85.7	82.3
In-person community outreach	182 (21.5)	182	136	58	48	52	74.7	31.9	82.3	89.7
Patient registry call	584 (68.9)	584	462	207	183	180	79.1	35.4	88.4	86.9

[a]CBO: community-based organization.

XSL•FO
RenderX

Table 3. Results of the logistic regression analysis using recruitment method to predict eligibility, enrollment, and retention.

Recruitment method	Eligible[a] (n=895)		Enrolled[a] (n=459)		Retained at 3 months[b] (n=390)		Retained at 6 months[b] (n=395)	
	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value
Mass recruitment	1.67 (1.01-2.76)	.04	2.60 (1.81-3.73)	<.001	1.4 (0.76-2.80)	.26	1.01 (0.49-2.1)	.98
Direct, low effort	1.70 (0.98-2.96)	.06	3.02 (2.03-4.51)	<.001	1.4 (0.73-2.9)	.29	1.32 (0.65-2.66)	.44
Direct, high effort	1.0	N/A[c]	1.0	N/A	1.0	N/A	1.0	N/A

[a]Denominator for recruitment method is number screened.

[b]Denominator for recruitment method is number enrolled.

[c]N/A: not applicable.

Differences in Participant Characteristics

The characteristics of enrolled participants (Table 4) were compared across recruitment methods. Participants recruited via mass recruitment strategies were significantly less likely to have health insurance (44.0% vs 71.2% and 71.7%, respectively; *P*<.001), lived significantly fewer years in the United States (22.4 years vs 32.4 years and 30.3 years, respectively; *P*<.001), significantly more likely to be 1st generation (92.7% vs 76.5% and 77.5%, respectively; *P*=.007), significantly more likely to primarily speak Spanish (89.3% vs 65.8% and 66.3%, respectively), and significantly more likely to be at high risk for alcohol abuse (5.8 mean score vs 3.8 mean score and 3.9 mean score, respectively; *P*<.001) compared with those recruited via direct, low-effort and direct, high-effort strategies. Participants recruited via mass recruitment strategies were significantly more likely to come from Mexico (45.2% vs 20.5% and 8.3%, respectively; *P*<.001), while participants from Central America were more likely to be recruited via direct, low-effort strategies and direct, high-effort strategies (13.1% vs 32.9% and 35.0%, respectively; *P*<.001) compared with mass recruitment strategies. Participants born in the United States were significantly more likely to be recruited via both direct low-effort strategies and direct high-effort strategies (9.5% vs 28.8% and 25.8%, respectively; *P*<.001) compared with mass recruitment strategies. Moreover, Latino smokers recruited via direct, high-effort strategies were more likely to have depressive symptoms (1.7 mean score vs 1.1 mean score and 1.4 mean score, respectively; *P*=.02) and anxiety (1.8 mean score vs 1.1 mean score and 1.6 mean score, respectively; *P*=.004).

With respect to the 3-month retention rate, participants were significantly more likely to primarily speak English (68.7% vs 80.6%; *P*=.06) and to be a second or higher generation American (75.8% vs 89.6%; *P*=.02) compared with participants who did not complete their 3-month follow-up assessment. With respect to 6-month retention, particpants were significantly older (49.7 years vs 46.5 years; *P*=.02) and reported less self-efficacy (1.9 mean score vs 2.2 mean score; *P*=.008) compared with participants who did not complete their 6-month follow-up assessment (Table 4).

Of participants who were ineligible (n=217), the most frequent reasons for ineligibility were planning to move in the next 6 months, not willing to come to all study visits, smoking on average less than 3 cigarettes per day, and not knowing how to send or read text messages (Table 5). Ineligible participants identified via direct, high-effort strategies were significantly more likely to plan to move in the next 6 months (69.4% vs 13.9% and 16.7%, respectively; *P*=.04) compared with mass and direct, low-effort strategies.

XSL•FO
RenderX

Table 4. Baseline characteristics of enrolled participants who were recruited using proactive and reactive strategies and who were retained at 3 months and 6 months.

Characteristic	Recruitment method				Retained at 3 months (n=391)			Retained at 6 months (n=395)		
	Mass (n=144), n (%)	Direct, low effort (n=117), n (%)	Direct, high effort (n=847), n (%)	*P* value	Yes, n (%)	No, n (%)	*P* value	Yes, n (%)	No, n (%)	*P* value
Female	54 (64.3)	39 (53.4)	157 (52.3)	.15	211 (54.1)	39 (58.2)	.60	212 (53.7)	38 (61.3)	.28
Greater than a high school education	25 (29.8)	29 (39.7)	117 (39.0)	.28	152 (39.0)	19 (28.4)	.10	153 (38.7)	18 (29.0)	.16
Has health insurance	37 (44.0)	52 (71.2)	215 (71.7)	<.001	264 (68.2)	40 (59.7)	.21	266 (67.9)	38 (61.3)	.31
Married	52 (61.9)	35 (47.9)	157 (52.3)	.22	207 (53.4)	37 (56.1)	.69	206 (52.4)	38 (62.3)	.17
Employed full time	55 (65.5)	36 (49.3)	151 (50.3)	.04	212 (54.4)	30 (44.8)	.19	209 (52.9)	33 (53.2)	.99
Annual income (US $)										
0-29,000	33 (39.3)	29 (39.7)	125 (41.7)	.34	159 (42.0)	29 (46.8)	.53	165 (43.8)	22 (36.1)	.45
30,000-59,000	33 (39.3)	22 (30.1)	86 (28.7)		125 (33.2)	16 (25.8)		120 (31.8)	21 (34.4)	
≥60,000	13 (15.5)	20 (27.4)	77 (25.7)		93 (24.7)	17 (27.4)		92 (24.4)	18 (29.5)	
Age (years)[a]	46.6 (11.1)	50.1 (12.5)	48.9 (10.7)	.11	48.9 (11.1)	47.5 (11.2)	.35	49.7 (11.1)	45.6 (10.9)	.02
Number of cigarettes per day[a]	10.3 (7.6)	12.8 (9.3)	11.7 (7.8)	.17	11.5 (7.9)	12.0 (8.0)	.68	11.5 (7.9)	12.4 (8.5)	.37
Number of prior quit attempts[a]	5.0 (9.4)	2.7 (5.6)	3.5 (7.8)	.16	3.8 (8.2)	2.7 (4.8)	.30	3.9 (8.3)	2.4 (3.4)	.17
Alcohol score[b]	5.8 (2.6)	3.8 (2.4)	3.9 (2.5)	<.001	4.2 (2.6)	4.5 (2.7)	.44	4.2 (2.6)	4.5 (2.8)	.38
Depressive symptoms[b]	1.1 (1.6)	1.4 (1.5)	1.7 (1.8)	.02	1.5 (1.7)	1.4 (1.8)	.65	1.6 (1.7)	1.3 (1.6)	.29
Anxiety[b]	1.1 (1.4)	1.6 (1.7)	1.8 (1.7)	.004	1.7 (1.7)	1.6 (1.7)	.60	1.7 (1.7)	1.6 (1.6)	.60
Self-efficacy[a]	2.1 (0.82)	2.0 (0.87)	1.9 (0.71)	.13	1.9 (0.7)	2.0 (1.7)	.72	1.9 (0.7)	2.2 (0.9)	.008
Years in the United States[a]	22.4 (14.3)	32.4 (16.9)	30.3 (16.7)	<.001	29.4 (17.0)	27.8 (13.7)	.43	29.2 (16.9)	29.1 (14.7)	.96
Language, Spanish	77 (89.3)	48 (65.8)	199 (66.3)	<.001	268 (68.7)	54 (80.6)	.06	275 (69.6)	47 (75.8)	.37
1st generation	76 (92.7)	52 (76.5)	224 (77.5)	.007	292 (75.8)	60 (89.6)	.02	305 (78.0)	47 (77.0)	.87
Region of birth										
Mexico	38 (45.2)	15 (20.5)	25 (8.3)	<.001	64 (16.5)	14 (20.9)	.07	67 (17.0)	11 (17.7)	.04
Caribbean	11 (13.1)	24 (32.9)	105 (35.0)		120 (30.9)	20 (29.9)		126 (31.9)	14 (22.6)	
South America	18 (21.4)	11 (15.1)	84 (28.0)		91 (23.5)	22 (32.8)		98 (24.8)	15 (24.2)	
Central America	9 (10.7)	2 (2.7)	8 (2.7)		15 (3.9)	4 (6.0)		12 (3.0)	7 (11.3)	
United States	8 (9.5)	21 (28.8)	76 (25.8)		98 (25.3)	7 (10.4)		90 (22.8)	15 (24.2)	

[a]Mean (SD).

[b]Score (sum score).

XSL•FO
RenderX

Table 5. Major reasons for ineligibility by recruitment method (only ineligibility criteria that included a total of ≥10 individuals are reported).

Reasons	Mass[a] (n=20), n (%)	Direct, low effort[a] (n=16), n (%)	Direct, high effort[a] (n=180), n (%)	*P* value
Not willing to come to all study visits	6 (30.0)	1 (6.7)[b]	41 (25.6)[c]	.22
Does not know how to send or read text messages	4 (20.0)	1 (6.7)[b]	36 (21.4)[d]	.51
Smokes cigarettes less than 3 days/week	4 (20.0)	3 (18.8)	36 (20.2)[e]	.20
Planning to move in the next 6 months	5 (25.0)	6 (37.5)	25 (14.7)[f]	.04
Uses other tobacco products more than 1 day/week	5 (25.0)	4 (25.0)	24 (13.9)[g]	.23
Not interested in quitting in 30 days	2 (10.0)	1 (6.7)	9 (5.3)[h]	.41
Has not smoked cigarettes for at least 6 months	2 (10.0)	1 (6.7)	7 (4.0)[i]	.29

[a]The denominator is the difference across ineligibility criteria because of missing data.

[b]n=15.

[c]n=160.

[d]n=168.

[e]n=178.

[f]n=170.

[g]n=172.

[h]n=82.6.

[i]n=174.

Discussion

Principal Findings

This paper compared mass; direct, low-effort; and direct, high-effort recruitment methods on Latino eligibility, enrollment, and retention at 3 months and 6 months for a smoking cessation clinical trial, *Decídetexto*. Results showed that, although direct, high-effort methods yielded the highest total number of enrollees, eligibility and enrollment were significantly lower when compared with the mass and direct, low-effort methods. However, when considering retention at 3 months and 6 months, there is no evidence that method of recruitment impacted retention once participants were enrolled in the study. Thus, although the eligibility and enrollment rates were low for direct, high-effort strategies, participants are just as likely to be retained after they are enrolled when compared with mass and direct low-effort strategies.

It is important to note that, although mass and direct, low-effort strategies are efficient methods and do not demand much staff time, they are unlikely to reach the recruitment goal for a randomized clinical trial without contribution from direct, high-effort strategies. Future studies should include cost-effectiveness to determine whether highly funded mass and direct, low-effort strategies can recruit equal numbers in a cost-effective manner. This is especially important to consider as mass recruitment strategies seemed to yield fewer acculturated enrollees with lower access to health services—groups that might benefit a great deal from the intervention.

Our research corroborates a study that tested the efficiency of strategies to recruit Latino male smokers. That study found that reactive recruitment was more efficient than proactive recruitment but yielded significantly fewer participants and was costlier per participant enrolled [11]. As noted by Harris et al [27], reactive recruitment may be more effective at identifying eligible individuals because it reaches a wider audience and individuals who take the trouble to respond are likely to be more ready and motivated to quit. Furthermore, individuals who learn about research via mass media may have more time to collect information about the study and consider the pros and cons of enrolling before calling the study phone number to complete eligibility. This, in turn, might prevent less motivated individuals from contacting the study for screening. It should be noted that this study's advertisements did not include all of our eligibility criteria. Potential participants were able to self-screen for some criteria using the Latino identity and current smoker criteria that were noted in the advertisments. There were a number of additional criteria they had to meet (Table 5). Future research should consider enhancing advertisements (eg, flyers, posters) to yield higher response rates using theoretical constructs such as self-efficacy, social norms, and rewards. Moreover, additional research should consider assessing which method of recruitment yielded a higher rate of participants who quit smoking.

Of the individual mass recruitment strategies, Facebook ads yielded the lowest efficiency ratios for eligibility and enrollment. This contradicts previous research that found Facebook was a useful recruitment tool for smokers [33,34]. Of direct low-effort strategies, patient registry texts yielded the lowest efficiency ratios for eligibility and enrollment and yielded among the lowest ratios for retention at 3 months and 6 months. This is consistent with previous research reporting an ~34% enrollment rate for patients recruited via text messages [35]. It is interesting to note that referring friends and family members were either (1) Latino smokers on a patient registry who received a call from us but no longer smoked or were ineligible or (2) study

XSL•FO
RenderX

participants who had completed the study. Of direct, high-effort strategies, in-person community outreach yielded the lowest efficiency ratios for eligibility and enrollment. This is important to note, as this is the recruitment strategy that demanded the most staff resources. Although patient registry calls yielded the highest number of participants, it was among the lowest efficiency ratio for eligibility and enrollment. This is in contrast to previous research that has reported the feasibility and cost-effectiveness of recruiting participants via calls from patient registries via a research associate program [20,33].

With respect to the characteristics of participants recruited via the different recruitment methods, mass recruitment yielded less acculturated participants (eg, more likely to speak Spanish, to be 1st generation, fewer years lived in the United States) who were more likely to be at risk for alcohol abuse than participants recruited via direct, low-risk and direct, high-risk methods. It is possible that low-acculturation Latino smokers face unique barriers that limit the effectiveness of direct recruitment strategies.

We also found that Latinos from different Latin American regions appear to respond differently to different recruitment methods. Mexicans were more likely to be recruited via mass recruitment strategies compared with all other Latin American regions, while Latinos from the Caribbean were more likely to be recruited using direct strategies. Thus, recruitment approaches that researchers choose to employ should be determined by their population of interest and the desired participant characteristics. Furthermore, Latino smokers experiencing depression or anxiety were less likely to respond to mass recruitment. This may be because they are less motivated to quit smoking or they have less energy to reach out to inquire about the study [36]. The psychosocial finding corroborates findings from a study that compared reactive versus proactive recruitment strategies in recruiting African American smokers [27], in which it was found that participants recruited proactively were more likely to report indicators of depression. Taken together, these findings suggest that both mass and direct recruitment strategies should be implemented for studies interested in recruiting Latino participants across the socioeconomic, acculturation, and country of origin spectra. Additional research is also needed to examine differences in clinical outcomes based on recruitment method.

Limitations

The *Decídetexto* clinical trial was not designed to test the efficiency of recruitment strategies; therefore, this study has several limitations. Given the broad reach of our advertisements, it is possible that participants were exposed to multiple advertisement strategies. It is possible that individuals responding to a mass strategy may have been exposed to a direct strategy. Therefore, some cross-contamination effect is likely to have occurred. However, no participants in this study mentioned that they learned about the study through more than one strategy. Moreover, we were unable to conduct a cost analysis in this study given that (1) most of our mass recruitment strategies were free of cost or paid in an unusual way (eg, paid a graphic artist in Mexico) and (2) we had volunteers aid in personalized recruitment for this study. We did not collect data on barriers to participant retention. Despite these limitations, this study has high representation of a heterogeneous group of Latino smokers representing different Latin American regions of origin, making it generalizable to Latinos nationwide. The study team consisted of Latino researchers and interns from different Latin American countries of origin, several of whom were native Spanish speakers. Our recruitment efforts involved working collaboratively with a community advisory board, collaborating closely with local CBOs, and culturally and linguistically tailoring all materials to Latino smokers. Furthermore, the bulk of our recruitment occurred prior to the onset of the COVID-19 pandemic. No recruitment activities occurred from March 2020 through July 2020. From August 2020 to March 2021 we recruited 21 additional participants, 62% of whom were recruited via mass recruitment strategies. It is possible that the effect of recruitment strategies will be different during and after the COVID-19 pandemic. Moreover, although this study is specific to a tobacco treatment trial, the findings are relevant to health research and clinical trials broadly.

Conclusion

This study compared the eligibility, enrollment, and retention efficiency ratios of recruiting Latino smokers via mass; direct, low-effort; and direct, high-effort strategies utilized in *Decídetexto*, a mobile smoking cessation randomized clinical trial. A heterogeneous sample of Latino smokers was enrolled in the trial. Results show that, although direct, high-effort recruitment strategies yielded the highest total number of enrollees, eligibility and enrollment were significantly lower when compared with mass and direct, low-effort recruitment strategies. Yet, when considering retention at 3 months and 6 months, there is no evidence that method of recruitment impacted retention once participants were enrolled in the study. Participants recruited via mass recruitment strategies were less acculturated, of lower socioeconomic status, and more likely to be Mexican than those recruited via other strategies. These findings suggest that these 3 recruitment methods should be implemented for studies interested in recruiting Latino participants across the socioeconomic, acculturation, and country of origin spectra. These findings provide further insight into effective recruitment strategies for Latino smokers.

Acknowledgments

The authors would like to thank the recruitment research staff and volunteers. This includes Ileana Cepeda, Arlette Chavez, Carlos Coronado, Katia Gallegos-Carillo, Stephanie Horton, Rudy Lara, Denise Maldonado, Joanna Mansfield, Maria Mansfield, Arturo Ponce, Dora Ponce, Rosibel Rodriguez Bolaños, Lizbeth Sandoval, Ciara Torres, Yamilet Torres, and Victoria Uceda.

Special acknowledgment goes to Marco Polo Ramírez Bonilla at In Touch Productions for the development of the graphic materials for recruitment. Finally, we thank the people and organizations that facilitated the recruitment of this study: (1) Angel Vazquez,

XSL•FO
RenderX

Bill Tipacti, Charles Neugebauer, Marlon Mazo, Roberto Yanez, Contigo en la Comunidad New York, and Noticiero Univisión Edición Digital at Univision Local Media, New York; (2) Aidè Almeida and Ricardo Salinas at El Gallito Radio; (3) Ida Pérez at Radio Poder 97.1 FM, part of Ibero American Action League; (4) Enrique Rodriguez Franz at KSMM-FM La Mexicana radio in Liberal, Kansas; and (5) Anthony Ibarria and Johnny Torres at El Especialito Hispanic Weekly in New York and New Jersey.

This work was supported by the National Cancer Institute (NCI) under grant R01CA212189 (principal investigator: APC) and T32CA102618 for author EAC. The content is the sole responsibility of the authors and does not necessarily represent the official views of the NCI.

Authors' Contributions

EAC, FCB, KK Rieth, EFE, LSC, KDG, FJD, DC, and APC conceptualized the study. EAC curated the data, performed the formal analysis, supervised the study, performed project administration, and drafted the original manuscript. APC acquired the funding. EAC, FCB, and APC designed the methodology. EAC, FCB, KK Rieth, KK Richter, EFE, LSC, KDG, FJD, DC, and APC reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

References

1. Quick Facts: United States. United States Census Bureau. URL: https://www.census.gov/quickfacts/fact/table/US/PST045219 [accessed 2021-03-23]
2. Colby SL, Ortman JM. Projections of the size and composition of the U.S. Population: 2014 to 2060. United States Census Bureau. 2015 Mar 03. URL: https://www.census.gov/library/publications/2015/demo/p25-1143.html [accessed 2022-06-12]
3. Creamer MR, Wang TW, Babb S, Cullen KA, Day H, Willis G, et al. Tobacco Product Use and Cessation Indicators Among Adults - United States, 2018. MMWR Morb Mortal Wkly Rep 2019 Nov 15;68(45):1013-1019 [FREE Full text] [doi: 10.15585/mmwr.mm6845a2] [Medline: 31725711]
4. Bethel J, Schenker M. Acculturation and smoking patterns among Hispanics: a review. Am J Prev Med 2005 Aug;29(2):143-148. [doi: 10.1016/j.amepre.2005.04.014] [Medline: 16005811]
5. Chen MS, Lara PN, Dang JHT, Paterniti DA, Kelly K. Twenty years post-NIH Revitalization Act: enhancing minority participation in clinical trials (EMPaCT): laying the groundwork for improving minority clinical trial accrual: renewing the case for enhancing minority participation in cancer clinical trials. Cancer 2014 Apr 01;120 Suppl 7:1091-1096 [FREE Full text] [doi: 10.1002/cncr.28575] [Medline: 24643646]
6. Freedman LS, Simon R, Foulkes MA, Friedman L, Geller NL, Gordon DJ, et al. Inclusion of women and minorities in clinical trials and the NIH Revitalization Act of 1993--the perspective of NIH clinical trialists. Control Clin Trials 1995 Oct;16(5):277-85; discussion 286. [doi: 10.1016/0197-2456(95)00048-8] [Medline: 8582146]
7. Ellington L, Wahab S, Sahami Martin S, Field R, Mooney KH. Factors that influence Spanish- and English-speaking participants' decision to enroll in cancer randomized clinical trials. Psychooncology 2006 Apr;15(4):273-284. [doi: 10.1002/pon.943] [Medline: 15973647]
8. Wallington SF, Luta G, Noone A, Caicedo L, Lopez-Class M, Sheppard V, et al. Assessing the awareness of and willingness to participate in cancer clinical trials among immigrant Latinos. J Community Health 2012 Apr 31;37(2):335-343 [FREE Full text] [doi: 10.1007/s10900-011-9450-y] [Medline: 21805372]
9. Vickers S, Fouad M. An overview of EMPaCT and fundamental issues affecting minority participation in cancer clinical trials: enhancing minority participation in clinical trials (EMPaCT): laying the groundwork for improving minority clinical trial accrual. Cancer 2014 Apr 01;120 Suppl 7:1087-1090 [FREE Full text] [doi: 10.1002/cncr.28569] [Medline: 24643645]
10. Thompson TP, Greaves CJ, Ayres R, Aveyard P, Warren FC, Byng R, et al. Lessons learned from recruiting socioeconomically disadvantaged smokers into a pilot randomized controlled trial to explore the role of Exercise Assisted Reduction then Stop (EARS) smoking. Trials 2015 Feb 12;16:1 [FREE Full text] [doi: 10.1186/1745-6215-16-1] [Medline: 25971836]
11. Graham AL, Lopez-Class M, Mueller NT, Mota G, Mandelblatt J. Efficiency and cost-effectiveness of recruitment methods for male Latino smokers. Health Educ Behav 2011 Jun 01;38(3):293-300 [FREE Full text] [doi: 10.1177/1090198110372879] [Medline: 21460176]
12. George S, Duran N, Norris K. A Systematic Review of Barriers and Facilitators to Minority Research Participation Among African Americans, Latinos, Asian Americans, and Pacific Islanders. Am J Public Health 2014 Feb;104(2):e16-e31. [doi: 10.2105/ajph.2013.301706]
13. London L, Hurtado-de-Mendoza A, Song M, Nagirimadugu A, Luta G, Sheppard VB. Motivators and barriers to Latinas' participation in clinical trials: the role of contextual factors. Contemp Clin Trials 2015 Jan;40:74-80 [FREE Full text] [doi: 10.1016/j.cct.2014.11.013] [Medline: 25433203]

XSL•FO
RenderX

14. Larkey L, Ogden S, Tenorio S, Ewell T. Latino recruitment to cancer prevention/screening trials in the Southwest: setting a research agenda. Appl Nurs Res 2008 Feb;21(1):30-39. [doi: 10.1016/j.apnr.2006.09.003] [Medline: 18226761]
15. Hildebrand J, Billimek J, Olshansky E, Sorkin DH, Lee JA, Evangelista LS. Facilitators and barriers to research participation: perspectives of Latinos with type 2 diabetes. Eur J Cardiovasc Nurs 2018 Dec;17(8):737-741 [FREE Full text] [doi: 10.1177/1474515118780895] [Medline: 29886773]
16. Flores A, Lopes G, Radford J. Hispanic Population in the United States Statistical Portrait. Pew Research Center. 2017 Sep 18. URL: https://www.pewresearch.org/hispanic/2017/09/18/2015-statistical-information-on-hispanics-in-united-states-trend-data/ [accessed 2022-06-12]
17. Krogstad JM. 5 facts about Latinos and education. Pew Research Center. 2016 Jul 28. URL: https://www.pewresearch.org/fact-tank/2016/07/28/5-facts-about-latinos-and-education/ [accessed 2022-06-12]
18. Arbona C, Olvera N, Rodriguez N, Hagan J, Linares A, Wiesner M. Acculturative Stress Among Documented and Undocumented Latino Immigrants in the United States. Hisp J Behav Sci 2010 Aug 19;32(3):362-384 [FREE Full text] [doi: 10.1177/0739986310373210] [Medline: 25484488]
19. Sheppard V, Cox L, Kanamori M, Cañar J, Rodríguez Y, Goodman M, Latin American Cancer Research Coalition. Brief report: if you build it, they will come: methods for recruiting Latinos into cancer research. J Gen Intern Med 2005 May;20(5):444-447 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.0083.x] [Medline: 15963169]
20. Arana-Chicas E, Cartujano-Barrera F, Ogedegbe C, Ellerbeck EF, Cox LS, Graves KD, et al. Feasibility and Effectiveness of Recruiting Latinos in -A Smoking Cessation Clinical Trial from an Emergency Department Patient Registry. Int J Environ Res Public Health 2021 Oct 15;18(20):10859 [FREE Full text] [doi: 10.3390/ijerph182010859] [Medline: 34682601]
21. Kiernan M, Phillips K, Fair J, King AC. Using direct mail to recruit hispanic adults into a dietary intervention: An experimental study. Ann Behav Med 2000 Mar;22(1):89-93. [doi: 10.1007/bf02895172]
22. Rhodes S, Alonzo J, Mann-Jackson L, Tanner AE, Vissman AT, Martinez O, et al. Selling the product: Strategies to increase recruitment and retention of Spanish-speaking Latinos in biomedical research. J Clin Transl Sci 2018 Jun;2(3):147-155 [FREE Full text] [doi: 10.1017/cts.2018.314] [Medline: 30510779]
23. García AA, Zuñiga JA, Lagon C. A Personal Touch: The Most Important Strategy for Recruiting Latino Research Participants. J Transcult Nurs 2017 Jul;28(4):342-347 [FREE Full text] [doi: 10.1177/1043659616644958] [Medline: 27114390]
24. Collins BN, Wileyto EP, Hovell MF, Nair US, Jaffe K, Tolley NM, et al. Proactive recruitment predicts participant retention to end of treatment in a secondhand smoke reduction trial with low-income maternal smokers. Transl Behav Med 2011 Sep 6;1(3):394-399 [FREE Full text] [doi: 10.1007/s13142-011-0059-6] [Medline: 24073063]
25. 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]
26. Sangi-Haghpeykar H, Meddaugh HM, Liu H, Grino P. Attrition and retention in clinical trials by ethnic origin. Contemp Clin Trials 2009 Nov;30(6):499-503. [doi: 10.1016/j.cct.2009.06.004] [Medline: 19573625]
27. Harris K, Ahluwalia J, Catley D, Okuyemi KS, Mayo MS, Resnicow K. Successful recruitment of minorities into clinical trials: The Kick It at Swope project. Nicotine Tob Res 2003 Aug;5(4):575-584. [doi: 10.1080/1462220031000118540] [Medline: 12959796]
28. Cartujano-Barrera F, Arana-Chicas E, Catley D, Cox LS, Diaz FJ, Ellerbeck EF, et al. Decídetexto: Mobile cessation support for Latino smokers. Study protocol for a randomized clinical trial. Contemp Clin Trials 2020 Dec;99:106188 [FREE Full text] [doi: 10.1016/j.cct.2020.106188] [Medline: 33080379]
29. Löwe B, Kroenke K, Gräfe K. Detecting and monitoring depression with a two-item questionnaire (PHQ-2). J Psychosom Res 2005 Feb;58(2):163-171. [doi: 10.1016/j.jpsychores.2004.09.006] [Medline: 15820844]
30. 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]
31. Donker T, van Straten A, Marks I, Cuijpers P. Quick and easy self-rating of Generalized Anxiety Disorder: validity of the Dutch web-based GAD-7, GAD-2 and GAD-SI. Psychiatry Res 2011 Jun 30;188(1):58-64. [doi: 10.1016/j.psychres.2011.01.016] [Medline: 21339006]
32. Cartujano-Barrera F, McIntosh S, Cox LS, Arana-Chicas E, Catley D, Ellerbeck EF, et al. Translation and Examination of the Reliability and Validity of the Spanish Version of the Smoking Self-Efficacy Questionnaire Among Latino Smokers. Tob Use Insights 2021 Jul 29;14:1179173X2110353 [FREE Full text] [doi: 10.1177/1179173x211035366]
33. Abar B, Ogedegbe C, Dalawari P, Freeman K, Boudreaux ED, Illuzzi F, et al. Promoting tobacco cessation utilizing pre-health professional students as research associates in the emergency department. Addict Behav 2015 Jan;40:73-76 [FREE Full text] [doi: 10.1016/j.addbeh.2014.08.014] [Medline: 25226592]
34. Cook S, Jerome R, Dunagan J, Kennedy N, Edwards T, Minnix JA, et al. Engaging smokers in research: Utility of Facebook in facilitating recruitment to a smoking cessation study. Contemp Clin Trials 2021 Aug;107:106461 [FREE Full text] [doi: 10.1016/j.cct.2021.106461] [Medline: 34098038]

XSL•FO
RenderX

35. Alsumidaie M. Text Messagingnhancing Clinical Trial Patient Recruitment and Enrollment. Applied Clinical Trials. 2014 Nov 3. URL: https://www.appliedclinicaltrialsonline.com/view/text-messaging-enhancing-clinical-trial-patient-recruitment-and-enrollment [accessed 2022-06-12]
36. Haukkala A, Uutela A, Vartiainen E, Mcalister A, Knekt P. Depression and smoking cessation. Addictive Behaviors 2000 Mar;25(2):311-316. [doi: 10.1016/s0306-4603(98)00125-7]

Abbreviations

AUDIT-2: Alcohol Use Disorders Identification Test-2
CBO: community-based organization
GAD-2: Generalized Anxiety Disorder-2
mHealth: mobile health
NCI: National Cancer Institute
PHQ-2: Patient Health Questionnaire-2

Edited by A Mavragani; submitted 12.11.21; peer-reviewed by S Morgan, C Bullen; comments to author 12.12.21; revised version received 30.03.22; accepted 18.04.22; published 27.06.22.

Please cite as:
Arana-Chicas E, Cartujano-Barrera F, Rieth KK, Richter KK, Ellerbeck EF, Cox LS, Graves KD, Diaz FJ, Catley D, Cupertino AP Effectiveness of Recruitment Strategies of Latino Smokers: Secondary Analysis of a Mobile Health Smoking Cessation Randomized Clinical Trial
J Med Internet Res 2022;24(6):e34863
URL: https://www.jmir.org/2022/6/e34863
doi:10.2196/34863
PMID:35759320

©Evelyn Arana-Chicas, Francisco Cartujano-Barrera, Katherine K Rieth, Kimber K Richter, Edward F Ellerbeck, Lisa Sanderson Cox, Kristi D Graves, Francisco J Diaz, Delwyn Catley, Ana Paula Cupertino. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.06.2022. 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

Practices and Attitudes of Bavarian Stakeholders Regarding the Secondary Use of Health Data for Research Purposes During the COVID-19 Pandemic: Qualitative Interview Study

Stuart McLennan[1], PhD; Sarah Rachut[2], Ass Jur; Johannes Lange[1], Dipl soz; Amelia Fiske[1], PhD; Dirk Heckmann[2], Dr Jur; Alena Buyx[1], Dr med

[1]Institute of History and Ethics in Medicine, TUM School of Medicine, Technical University of Munich, Munich, Germany

[2]TUM Center for Digital Public Services, Department Governance, TUM School of Social Sciences and Technology, Technical University of Munich, Munich, Germany

Corresponding Author:
Stuart McLennan, PhD
Institute of History and Ethics in Medicine
TUM School of Medicine
Technical University of Munich
Ismaninger Straße 22
Munich, 81675
Germany
Phone: 49 089 4140 404
Email: stuart.mclennan@tum.de

Abstract

Background: The COVID-19 pandemic is a threat to global health and requires collaborative health research efforts across organizations and countries to address it. Although routinely collected digital health data are a valuable source of information for researchers, benefiting from these data requires accessing and sharing the data. Health care organizations focusing on individual risk minimization threaten to undermine COVID-19 research efforts, and it has been argued that there is an ethical obligation to use the European Union's General Data Protection Regulation (GDPR) scientific research exemption during the COVID-19 pandemic to support collaborative health research.

Objective: This study aims to explore the practices and attitudes of stakeholders in the German federal state of Bavaria regarding the secondary use of health data for research purposes during the COVID-19 pandemic, with a specific focus on the GDPR scientific research exemption.

Methods: Individual semistructured qualitative interviews were conducted between December 2020 and January 2021 with a purposive sample of 17 stakeholders from 3 different groups in Bavaria: researchers involved in COVID-19 research (n=5, 29%), data protection officers (n=6, 35%), and research ethics committee representatives (n=6, 35%). The transcripts were analyzed using conventional content analysis.

Results: Participants identified systemic challenges in conducting collaborative secondary-use health data research in Bavaria; secondary health data research generally only happens when patient consent has been obtained, or the data have been fully anonymized. The GDPR research exemption has not played a significant role during the pandemic and is currently seldom and restrictively used. Participants identified 3 key groups of barriers that led to difficulties: the wider ecosystem at many Bavarian health care organizations, legal uncertainty that leads to risk-adverse approaches, and ethical positions that patient consent ought to be obtained whenever possible to respect patient autonomy. To improve health data research in Bavaria and across Germany, participants wanted greater legal certainty regarding the use of pseudonymized data for research purposes without the patient's consent.

Conclusions: The current balance between enabling the positive goals of health data research and avoiding associated data protection risks is heavily skewed toward avoiding risks; so much so that it makes reaching the goals of health data research extremely difficult. This is important, as it is widely recognized that there is an ethical imperative to use health data to improve care. The current approach also creates a problematic conflict with the ambitions of Germany, and the federal state of Bavaria, to be a leader in artificial intelligence. A recent development in the field of German public administration known as *norm screening*

XSL•FO
RenderX

(*Normenscreening*) could potentially provide a systematic approach to minimize legal barriers. This approach would likely be beneficial to other countries.

(J Med Internet Res 2022;24(6):e38754) doi:10.2196/38754

KEYWORDS

COVID-19; data sharing; General Data Protection Regulation; GDPR; research exemption; public health; research; digital health; electronic health records

Introduction

Background

The COVID-19 pandemic is a threat to global health and requires collaborative health research efforts across organizations and countries to address it. However, lack of integrated, comprehensive, and accessible patient-level data has been identified as a key barrier to COVID-19 research across the globe [1].

A valuable source of information for researchers is the large amount of digital health data collected by health care organizations through electronic health records. Indeed, health care systems worldwide are increasingly using this routinely collected digital health data for biomedical research, enabling large-scale and multidimensional aggregation and analysis of heterogeneous data sources [2]. The increase of such digital data has also created significant opportunities for artificial intelligence (AI) in health care [3]. With the ability to learn from large sets of clinical data, health care AI applications have the potential to support a wide range of activities [4-11], and public and private sector investment in the field continues to grow [12-14]. If data-intensive medicine is able to realize the continuous improvement of health care quality and thereby reduce patient harm, increase health, empower patient decision-making, and improve equity, it would fulfill the core ethical principles of health care [15,16].

However, benefiting from digital health data requires the ability to access and share the data. Single-center databases are also somewhat limited and sharing data across institutions and countries has various potential advantages, including allowing cross-validation of models across institutions to determine which findings are institution specific and which are generalizable and for knowledge discovery to be accelerated [17]. Efforts to create and link databases for secondary-use research, however, can be undermined by concerns about data protection; concerns that are only likely to intensify available data for research become higher resolution and more diverse (eg, medical images and physiological waveforms) [18].

Patients have legitimate interests in controlling access to and use of their health data, and their consent is often required for the use of their personal data if it was not collected for specific research purposes [18]. However, requiring consent for pseudonymized data to be used in secondary-use research cannot only lead to significant administrative and financial hurdles that delay or even impede important research but can also create major selection biases that undermine data representativeness [19]. Although fully anonymized data typically fall outside data protection laws around the world and can thus be freely used and shared, full anonymization is increasingly difficult to achieve given the use of models that can correctly reidentify people in anonymized data sets [20]. Furthermore, irreversible anonymization involves removing essential information required for most large collaborative research projects [21].

The European Union's (EU) General Data Protection Regulation (GDPR) is a key legal framework for the use and exchange of European digital health data for research purposes [18]. The GDPR entered into force in May 2016 but was only applied from May 25, 2018. Although early drafts of the GDPR raised concerns that the regulation may severely restrict data research [22], the final text adopted a more research-friendly approach, and it was thought that the GDPR would have little negative impact on data research overall [23]. However, concerns remain that the GDPR has made many organizations very risk-averse in terms of data sharing, even if the regulation permits such sharing; for example, via scientific research exemption [1]. When health care organizations are overly concerned with individual risk minimization, it threatens to undermine COVID-19 research efforts. It has been argued that there is an ethical obligation to use the GDPR scientific research exemption, particularly during a crisis such as the COVID-19 pandemic, to support collaborative health research [1].

However, the integration of clinical care and clinical research as part of a learning health care system can often conflict with the current regulatory system and raise a number of important ethical, legal, and social implications [24]. The need for more work on determining when patient notification and consent are required has been particularly highlighted, and investigating the views of patients and other stakeholders has been identified as essential to this work [25,26]. Previous international empirical research with patients regarding the secondary use of their data has found widespread support for such activities and willingness to share their data; however, it has also highlighted variations in patient's wishes regarding notification and consent [27-38]. Previous research with other stakeholders is rather limited. However, a systematic review found that although researchers and health care professionals were generally supportive of data sharing, they raised concerns about access to data, data storage infrastructure, and consent [39]. Research with other stakeholders has highlighted the challenge of balancing the benefits and risks of secondary research [40-42].

Although Germany is known for its strict approach to data protection, it is currently attempting to make health data more useful and meaningful, such as through the Medical Informatics Initiative [43]. During the COVID-19 pandemic, there have also been large research corporations, such as the National University Medicine Research Network. On April 15, 2020, a nationwide standardized template document for patient consent was

XSL•FO
RenderX

approved, enabling researchers across Germany to obtain broad consent for the use of pseudonymized health data in accordance with the GDPR. Germany has also taken a number of regulatory efforts to support digital health care. In 2019, it enacted a new Digital Healthcare Act in 2019, which allows digital health applications to be prescribed and reimbursed under statutory health insurance [44,45]. In 2020, the Patient Data Protection Act was passed, and a step-by-step plan to implement electronic health records and complementary applications such as electronic prescriptions was announced [46]. Despite these efforts, however, the successful implementation of such digital health applications has experienced significant delays and is yet to be achieved [47,48]. Nevertheless, Germany, and the federal state of Bavaria in particular, has also set the goal of becoming a leading hot spot and innovation location for AI [49]. Although recent research involving German patients indicates that abolishing consent for secondary research use of clinical data will likely be acceptable to a large majority of patients [28,29], we are not aware of empirical research with other stakeholders such as researchers, data protection officers, and research ethics committee members, regarding their views and use of the GDPR scientific research exemption for secondary-use health data research either during or before the pandemic.

Objectives

The first German COVID-19 case appeared in Bavaria in January 2020, and Bavaria was one of the most affected states in Germany during the pandemic. In August 2020, the Bavarian State Ministry of Science and the Arts funded the Technical University of Munich's Faculty of Medicine for COVID-19 research projects. As a part of this program, this project aimed to explore the practices and attitudes of Bavarian stakeholders regarding the secondary use of health data for research purposes in a time of particular need for fast, data-rich research, namely during the COVID-19 pandemic. It was particularly interested in exploring stakeholders' views and use of the GDPR scientific research exemption for secondary-use health data research, either during or before the pandemic. Such research, even when performed at the local and regional levels to assess attitudes within a specific legal, cultural, and national context, can illuminate and inform the wider challenge of balancing the goals of furthering health research and improving public health with the goal of responsible data use, a challenge that is relevant across the globe.

Methods

The methods of the study are presented in accordance with the COREQ (Consolidated Criteria for Reporting Qualitative Research) [50].

Research Team and Reflexivity

Personal Characteristics

Interviews were conducted by JL, a male PhD student in sociology. JL, SM, AF, and AB have long-standing experience with qualitative research [24,51-65].

Relationship With Participants

No relationship was established between the interviewer and participants before the study, and the participants received limited information about the interviewer. There was no hierarchical relationship between the interviewers and study participants.

Study Design

Theoretical Framework

The theoretical framework used in this study was conventional content analysis [66].

Participant Selection

Stakeholders were primarily selected through purposive sampling [67] to ensure that the participants involved in COVID-19 data sharing for scientific research were from different backgrounds. Additional participants were identified using snowball sampling [68]. Participants were contacted by email and provided with information about the study design and aims and rights as participants. Suitable dates for an interview were found for those willing to participate. Verbal consent was obtained from all participants directly before the interview and audio recorded. A total of 17 Bavarian stakeholders agreed to participate in the study and were recruited from 3 groups: researchers involved in COVID-19 research (n=5, 29%), data protection representatives (n=6, 35%), and research ethics committee representatives (n=6, 35%). A total of 6 people who were contacted did not respond to emails.

Setting

The interviews were conducted between December 2020 and January 2021. All interviews were conducted via a telephone or video call in German. Only the participant and researcher were present during the interview. Overall, 71% (12/17) of the stakeholders were male, and 29% (5/17) were female.

Data Collection

A researcher-developed semistructured interview guide was developed for each group to guide the discussions (Multimedia Appendix 1). On the basis of the first 2 interviews that did not show any problems, it was decided that no further piloting or adaptation of the interview guides was necessary. No repeat interviews were conducted. Interviews were audio recorded, and no field notes were taken. The interviews lasted an average of 32 minutes (range 20-41 minutes). The interviews were transcribed in full, checked for accuracy, and subsequently pseudonymized. After 17 interviews, a question about data saturation arose, and it was concluded that saturation was reached in the content and attitudes expressed by the participants [69]. The transcripts of the interviews were returned to all participants with an invitation to review the transcription and send any corrections or clarifications; a total of 6 responses were received with minor corrections to syntax.

Analysis and Findings

Using the interview transcriptions in their original language, JL and SM performed conventional content analysis with the assistance of the qualitative software MAXQDA (version 11; VERBI Software). The analysis commenced after the interviews

XSL•FO
RenderX

were completed. Initial themes identified that were common across participants, as well as those unique to individuals, were labeled using a process of open coding. The findings are presented as higher and lower level categories. The other investigators (SR, AF, DH, and AB) reviewed the initial analysis to clarify and refine codes, and conversations among the investigators continued until coding differences were resolved and consensus was achieved. Selected quotes have been translated into English by the researchers using back translation.

Ethics Approval

This study received a waiver from the Technical University of Munich's Research Ethics Committee.

Results

Current Practices

Participants identified systemic challenges in conducting collaborative secondary-use health data research in Bavaria, particularly research that involves sharing health data outside the institution it was collected. These were reported to be preexisting challenges that were independent of the COVID-19 pandemic but were often brought into sharp focus during the pandemic.

Participants described strict handling of patient data in Bavaria, which led to collaborative secondary health data research being conducted primarily only when patient consent (individual or broad) had been obtained, or the data had been fully anonymized. Although patient data could be used within the hospital where it was collected for research or educational purposes without consent, it was reported that sharing pseudonymized data for research purposes outside the hospital where it was collected was generally not possible under Bavarian law without consent. Participants reported that this could make the use of Bavarian patient data in multicenter studies highly bureaucratic and time consuming. Although many participants thought it was important that the GDPR research exemption existed, they reported that it is currently used seldom and very restrictively:

> *I welcome the fact that this exception exists. Ultimately, it says that the common good takes precedence over individual rights under certain circumstances. That also has to be weighed up. It is also right that ethics committees are called upon to weigh up such things.* [P18, ethics committee]

> *And that is the attitude of our ethics committee. We are very restrictive [with the research exemption]. One can ask the patient.* [P02, ethics committee]

> *So we have never actually applied this article. And it's also questionable. So I think that the supervisory authorities will apply very strict standards when it comes to research privilege. So as I said, we have never applied it. It's really only ever consent that comes into question.* [P12, data protection]

Participants reported that the strict handling of patient data in Bavaria continued during the COVID-19 pandemic, although COVID-19 project applications were assessed more urgently than other research applications. COVID-19 projects were primarily conducted based on patient consent or anonymization, and participants felt that the GDPR research exemption did not play a significant role during the pandemic. Nevertheless, some research ethics committee representatives reported instances during the pandemic where their committees had allowed the use of pseudonymized data for research purposes without patient consent, as they wanted to allow valuable research to be conducted; however, they saw themselves at risk of breaking the law:

> *It [the GDPR research exemption] played practically no role for us. It did not have to be forced, because as an ethics committee we made it possible from the outset to work with the data. The patients who were COVID positive and able to give consent were, as far as I know, very willing to agree to this, and those who were not able to give consent because they were too ill, we as an ethics committee stuck our necks out so that their data and samples could also be used.* [P10, ethics committee]

For many participants, balancing the protection of patient privacy with health research for the common good was at the core of many of these challenges, which was particularly pronounced during the COVID-19 pandemic. Research ethics committee representatives saw it as their responsibility to consider how to best balance these issues:

> *Of course, there are always situations where two fundamental rights conflict with each other. This is precisely what we have now with this COVID situation [...] Then it is also clearly a matter for society to discuss where we stand. In case of doubt, which of these fundamental rights is more important to us, and in what form, and how can we take the other into account accordingly? And that's one of the points we have here. I think that's the case with many ethics committees, that they say, well, the right to data protection, and the right to research, that's also a right. If there is an extreme contradiction, then we as an ethics committee are authorized, or there is a social consensus that the ethics committees are authorized, to simply look at how strongly the personal right is restricted and how strongly do we restrict the research project. If we had a research project that did not yield any knowledge, then it would not matter. Then the right to privacy always applies. But if we have an emergency situation, then you can probably also say, these retrospective analyses, you might not necessarily need patient consent.* [P2, ethics committee]

Nevertheless, a number of participants thought that the current strict handling of patient data in Bavaria and Germany generally undermined important health research:

> *If you ask me personally, I actually find the handling of patient data in Germany too strict. We would need to create legal regulations as other countries have done. Health data protection laws or research data protection laws, for example [...] That would be*

XSL•FO
RenderX

> *feasible. But if you follow the discussion about this patient data protection law for [statutory health insurance] patients within the framework of the telematics infrastructure, we Germans, or many Germans, have a fundamental distrust of state institutions and are therefore not prepared to make data available that would really be very helpful for medical research. As a legislator, you probably have to accept that. And in this respect, yes, that is my personal opinion. I think it's a pity, because we fall behind many other countries in the context of medical research, but yes, that's a decision of the legislator.* [P5, data protection]

Barriers to Collaborative Health Data Research in Bavaria

Participants identified three key groups of barriers that led to difficulties in conducting collaborative health data research in Bavaria: (1) the wider ecosystem at many Bavarian health care organizations, (2) legal uncertainty and risk minimization, and (3) participants' ethical views.

Wider Ecosystem

A number of participants identified issues in the wider ecosystem at many Bavarian health care organizations as underlying barriers to collaborative secondary-use health data research.

Medical Informatics

Although participants noted increasing pressure from the German-wide Medical Informatics Initiative to use health data, it was reported that many Bavarian health care organizations were still not using the valuable data they possess. Medical informatics systems were often reported to be inadequate and that there were insufficient people with the right knowledge and skills to implement such systems:

> *However, this is more likely to come about as a result of pressure from the higher goals of the medical informatics initiative. Simply saying, people, you have to get your data usable at all. And that means there is already an interest among clinics as well. Because they know what a treasure trove of information they have that they don't even use. Even to the detriment of patients, they don't use it or can't use it. Because the information that is available is not used. That is a disadvantage. And I believe that the clinics are already working on this, but there are not enough people to implement it. There is a lack of computer scientists who can implement this, they are being swept off the market because everyone needs one. And data protection experts. So that's all being built up right now. So I think that's why it's a difficult time right now, because I think they're all being built up on a voluntary basis. There are structures being created. I think medical informatics is a big driver to systematically create structures nationwide, that's why I say that so often. And then the hospital boards say, yes, finally something uniform. The others are doing the same. Maybe that's something that helps a little bit.* [P20, ethics committee]

GDPR Implementation

Some participants felt that the basic implementation of the GDPR is still lacking in many Bavarian health care organizations; smaller institutions, in particular, were reported to have insufficient financial and personal resources to adequately implement it. Some participants were also not in favor of using the research exemption until sufficient implementation of the GDPR was achieved:

> *Well, funnily enough, so after we have been standing here since 2018, three years later, we are still lacking the basic implementation of the GDPR, I think it is the last remaining bastion that is being taken care of here. [Interviewer: So it's sort of, we need to get this place up and running first?] Right. Yes, but actually it's like that. So already data protection per se. So to somehow take all these requirements into account, that's already difficult in itself. And then there's the question of how to implement it, especially technically. And then somehow this research exemption, that would be the crowning glory, so to speak. To be honest, that may be different at other universities that have more money available and are also larger, but not at our university. I think we are basically too small for that and we are not...It's also still a matter of manpower. So you also have to have time and capacity for it somehow. First of all, the basic technical possibilities for complying with the GDPR are lacking at every turn. So I am no longer in favour of this kind of research exemption.* [P27, ethics committee]

Strict interpretation of the GDPR also made life very challenging at the beginning of the pandemic, with a participant calling for more flexibility and proportion during the pandemic:

> *The GDPR made life difficult for us in the first few weeks. Because we were partly inhibited in our interaction with the health authorities. That is, so documents that in the Stone Age could only be sent back and forth by fax. Then, when it came to discharging patients to home isolation, etc., we had to do it by fax. That was very tedious. And there, of course, one would like to see a little sense of proportion in the pandemic. And, as the saying goes, the church should be left in the village and not just read the letter of the law. Because sometimes things have to move very quickly in pandemic times. And we really have to adapt requirements to the situational context.* [P03, researcher]

Legal Uncertainty and Risk Minimization

Participants perceived legal uncertainty regarding a number of issues were leading people to be risk adverse in relation to collaborative secondary-use health data research in Bavaria.

XSL•FO
RenderX

Bavarian Hospital Act

Participants identified the Bavarian Hospital Act's Article 27 on Data Protection as a significant barrier. Although participants reported that the act permitted patient data to be used within the hospital for education and research purposes, they repeatedly noted the challenges raised by the requirement in Article 27 that patient data must remain in the custody of the hospital. Participants felt that the provisions of the Bavarian Hospital Act prevented the GDPR research exemption from being used and that patient consent is required if pseudonymized data are shared with third parties for research purposes:

> *However, this is not based on the research exemption that you allude to in the GDPR. And in my opinion, it is also not possible in Bavaria, because the Bavarian Hospital Act contains special regulations that prevent this.* [P05, data protection]

> *That is one point and the other, which I must of course make clear, according to Article 27, Paragraph 4, of the Bavarian Hospital Act, I may indeed conduct research in the hospital with the data as a treating physician, I may commission others, but the data, if I do not have consent, may only leave the house anonymized, which brings us back to the vexed topic of what is anonymised?* [P04, data protection]

Nevertheless, participants reported that the GDPR research exemption had been incorporated into the Bavarian Data Protection Act but that Article 27 of the Bavarian Hospital Act had not changed:

> *And in the Bavarian Data Protection Act, such things are partly taken up. [...] In as much, the Bavarian legislator, and it explicitly says above the processing for research purposes, has more or less taken reference to this, has created a regulation, for processing data for research purposes, but has not attached or has not changed Article 27 of the Bavarian Hospital Act.* [P05, data protection officer]

Participants described how this situation created a great deal of legal uncertainty in Bavaria and led to a general unwillingness to share pseudonymized patient data for secondary-use data research without consent.

Vagueness of Law

Regarding the GDPR research exemption, participants reported that as there was significant perceived vagueness in the law in Germany, people avoided using the research exemption to reduce their legal risk:

> *So they would still need a national law that really allows this. And the laws that currently allow this are very general. So they say that if the research objective cannot be achieved in any other way, then you can also use personal data. [...] The deeper this encroachment is, the less help general clauses and laws are. That's the case throughout German law. And it's the same here. So you can't create a large research project with a deep intrusion into the rights of people and say that I have a general clause that says, if there's no other way, then that's how we'll do it. So as I said, in the Bavarian hospital law it is explicit. It says, yes, you are allowed to do research with the data. And there is also an exception in the data protection law, which gives a suitable guarantee, so to speak. This does not mean that the data may not leave the custody of the hospital. And such, such a clause or such a regulation does not exist nationally, and therefore we have a bit of a problem. So the opening clause GDPR, yes, but national law, very shaky.* [P08, data protection]

> *Yes, we have with this [research exemption]. However, I understood the lawyers to say that this exemption is so vague that in case of doubt, if the patient sues, the physician is poorly advised if they do not have the consent. Because we simply have more sensitive data here. So we don't invoke this research exemption in research or in the ethics committee. In any case, so far, I think there is a consensus among all ethics committees that this [research exemption] is not sufficient.* [P02, ethics committee]

> *Of course, we would prefer that or that there would be a concrete definition of how to deal with the research exemption of the GDPR. As far as I know, also from colleagues, we have actually avoided all of this so far and taken the standard route. We see to it that we get consent.* [P04, data protection]

Variations in the Interpretation of the Law

Participants also reported that wide variations in the interpretation of the law could create significant uncertainty and cause researchers to take a very conservative approach to minimize risk. At the local level, participants reported that a lot depended on the makeup of the research ethics committee, whether it was more medically or legally oriented, with a legally oriented ethics committee seen to be more complicated:

> *The other problem is that ethics committees are structured differently. Whether they are more medically oriented or more legally oriented. So that has to be said honestly. And the more legalistic it becomes, the more complicated it usually is. Because it depends on whether you are looking for a solution or whether you say, I'll make it easy for myself, I'll forbid it for now or I'll put up some kind of hurdle and then I'll have peace, I'll close it now. As such it's very difficult to balance in between them. Because both sides are right, but it's always a trade-off.* [P20, ethics committee]

Local ethics committees and data protection officers did not have the same opinion, sometimes resulting in additional hurdles for the researchers:

> *But even there, there was and is a lot to clarify with data protection officers. Because sometimes there are two hurdles. One is the ethics committees, which are then called upon, and the other is the data protection officers, who do not necessarily always have identical ideas.* [P20, ethics committee]

XSL•FO
RenderX

Collaborative research at the national and international levels was reported to be generally difficult, with a lack of consensus among data protection commissioners and variations in local laws:

> *However, in the overarching sense of the exemption regulation, they have to deal with the individual data protection commissioners of the federal states. This makes supra-regional studies extremely difficult because there are so many different opinions of the data protection commissioners and no general opinion can be reached.* [P20, ethics committee]

> *Even that doesn't necessarily help you, because of course data protection is also covered by many area-specific regulations such as the state hospital laws. Yes. This means that what Bavaria now applies does not necessarily apply equally in the case of Rhineland-Palatinate.* [P04, data protection]

> *And yet we have 16 or 17 supervisory authorities with different ways of applying the law. Because it is somehow difficult when the data protection commissioner in Baden-Württemberg says something different about the same facts as the data protection commissioner in Berlin. And the further north, the stricter. That has to be said quite clearly.* [P05, data protection]

One area cited by the participants as resulting in significantly different interpretations was the distinction between pseudonymized data and fully anonymized data:

> *Is it absolutely anonymised or is it relatively anonymised? Is the recipient of the data anonymous, because he doesn't know that 150 is Mrs Meier? Or does someone who has passed it on still have a list somewhere that says 150 is Mrs Meier? There are also different interpretations. To be fair, it has to be said that with the strict interpretation it will never be possible to create anonymised data. But here too, as we have learned, there is no agreement among the data protection commissioners, neither in Germany nor throughout Europe.* [P20, ethics committee]

> *When sharing data, I have to inform the person responsible, in our case the board of directors, about the legal risk he is taking, because there is no secure interpretation of when pseudonymised data, which cannot be identified by another person, i.e. which are in fact anonymous in the old way of speaking, fall under the GDPR or not. And there are both interpretations in the literature. And research projects that you absolutely want to have, let's say, then perhaps you tend to say that this is not your own legal basis for dealing with de facto anonymous data. And if at some point the ice becomes too thin, then perhaps in other cases one will say that it is personal data, we cannot do it without consent. [...] I point out every time that there is a certain risk here when you share this data. However, I consider the actual risk of a supervisory authority in Covid times taking action against a research institution that actually exchanges anonymous data to be really small.* [P08, data protection]

Ethical Views

The reluctance to conduct health data research without patient consent also reflected participants' ethical views. Some participants felt that patient consent ought to be obtained whenever possible to respect patient autonomy, and consequently, that the use of the research exemption should be very limited and as an option of last resort:

> *So you're right. I'm a bit reluctant to take this as a license for all kinds of things. Fortunately, it has to be said that many researchers don't even know this recital. Even within the Ethics Committee, probably not every member knows it in detail either. But yes, I have already said a few times that I would support and welcome something like this. If this is really useful and actually advances science, and I see myself as a scientist in the same way. Then I think, yes, we should use it. But if it serves the laziness somewhere and he says, no, I can do it much more elegantly and it's all so time-consuming and always inform the poor patient and in the end he doesn't agree. Then to take refuge in that, I think, is not correct.* [P18, ethics committee]

Some participants perceived a risk that the research exemption could lead to a *carte blanche* to use patient data and questioned why COVID-19 research should be treated differently from other types of research:

> *As far as the [research exemption] is concerned, this is viewed somewhat more cautiously, because otherwise it can degenerate into a carte blanche. We do research on humans. I mean, why is Corona now higher-ranking than cancer research or something?* [P20, ethics committee]

Facilitators for Collaborative Health Data Research in Bavaria

To improve health data research in Bavaria, participants wanted greater legal certainty regarding the use of pseudonymized data for research purposes without patient consent. In the short term, some participants felt that a clear statement from relevant local authorities clarifying the application of the law would be helpful:

> *I would suggest the following. Briefly, at least in the short term, a stipulation by the relevant state supervisory authorities that project data and really, I would call it project pseudonymised data, which are not traceable for the recipient of this data under any reasonable conditions and using any normally applicable means or methods, are treated the same as fully anonymised data as far as the transfer of data is concerned and this transfer of data does not require consent. It is equivalent to fully anonymised data as far as data transfer is concerned and this data transfer does not require consent. In the long term, I would like to see an amendment to Article 27 of the Bavarian Hospital Act to the effect that I can say that data may leave the hospital anonymised and*

XSL•FO
RenderX

> *correspondingly properly pseudonymised without requiring consent.* [P04, data protection]
>
> *Well, there could be a clear statement by the Bavarian legislator, so to speak, about what data is affected and how, from the point of view of...well, if you could give people more legal certainty and perhaps also a public statement of this information... By making this statement officially, a lot would be gained. If we had something to refer to, whereby, as I said, the Bavarian hospital laws already have this research possibility, but it is always limited in the sharing.* [P20, ethics committee]

However, most participants ultimately thought that Article 27 of the Bavarian Hospital Act needed to be amended to allow appropriately pseudonymized data to leave the hospital without the explicit consent of the patient:

> *In my opinion, this can only be done if the legislator amends Article 27, Paragraph 4 of the Bavarian Hospital Act. And not only for Covid data, but I think for research into...So Covid is of course important and currently very high. But there are many other diseases [...] and if there were a legal basis for using this data for research purposes across the board, that would certainly, as I briefly mentioned before, for me personally, yes, that wouldn't be bad.* [P05, data protection]
>
> *In the long term, I would like to see an amendment to Article 27 of the Bavarian Hospital Act to the effect that data may leave the hospital anonymised and correspondingly properly pseudonymised without requiring consent.* [P04, data protection]

However, participants also pointed out that this situation cannot be improved by Bavaria alone and saw the need for a federal law for the handling of research data:

> *To be honest, I don't see Bavaria as the decisive factor for advancing research in an area like this. You have to talk to more specialised centres, and the German university hospitals are already a good cluster. Although at the level of a university hospital, I would say that a Bavarian regulation would perhaps be easier to implement in parliament, but all the projects of the National University Medicine Initiative alone would not be helped here, because they are all coordinated via Charité.* [P08, data protection]
>
> *I would like to see something like a research law or at least a binding specification of the requirements and a binding harmonisation of the requirements at federal level. We should say that we are enacting a research law and that data should be handled in such and such a way. Taking into account the GDPR, data protection regulations, hospital regulations and other regulations.* [P04, data protection]

Discussion

Principal Findings

This is one of the first qualitative studies examining European stakeholders' views and use of the GDPR scientific research exemption for secondary-use health data research, either during or before the COVID-19 pandemic. It also aims to add empirical insight to the global debate about the conflicting goals of furthering public health and health research on the one hand and protecting individual privacy and ensuring responsible data use on the other. This study has resulted in two key findings: (1) stakeholders in the German federal state of Bavaria were generally unwilling to use scientific research exemption owing to legal and ethical concerns and (2) stakeholders felt that the current strict handling of patient data is undermining important health research. This study suggests that the balance between enabling the positive goals of health data research and avoiding associated data protection risks can often be heavily skewed toward avoiding risks; thus making it extremely difficult to reach the goals of health data research. This is important as it is widely recognized that there is an ethical imperative to use health data to improve care. The current approach also creates a problematic conflict with Germany's, and the federal state of Bavaria's, ambitions to be a leader in AI. However, this is also a challenge for many other countries.

Despite recent research indicating that abolishing consent for secondary research use of clinical data will likely be acceptable to a large majority of German patients [28], Germany and many other countries, including in the EU, are still pursuing a *consent or anonymize* approach. Various authors have argued for several years that this approach undermines data-intensive medicine and that there is a need to specify the appropriate conditions for using a research exemption from consent [21]. Article 9(2)(j) of the GDPR sets out a scientific research exemption for processing sensitive personal data, which could occur without consent if subject to appropriate safeguards and if such rights would render impossible or seriously impair the achievement of the research purpose—see Article 89(1). Such a research exemption has a number of advantages in the context of the secondary use of health data for research purposes. The data from a large number of patients not requiring consent can be covered by the same provision. The existence of the research purpose is also relatively independent of further developments. In contrast, for example, if the data processing is based on Article 9(2)(i) GDPR, which explicitly allows the processing of sensitive personal data if it is "necessary for reasons of public interest in the area of public health," the processing can only take place as long as it is necessary to protect against the risk. If the situation stops being so dangerous, it would have the consequence that the data could not be used any further, not even to avert more abstract or possible future dangers [70].

However, as this study highlights many countries across the world continue to pursue a restrictive approach regarding the secondary use of patient data and seldom allow the use of data without consent or anonymization. As the participants of this study noted, various regulatory regimes must be considered, both at the national and European levels, with regard to the legal

XSL•FO
RenderX

use of data in EU countries. Owing to the primacy of the application of EU law, the lawfulness of data processing is governed by the GDPR. This is supplemented by further general data protection laws at the national level (for Germany, the Federal Data Protection Act), as well as the state level (for Bavaria, the Bavarian Data Protection Act). However, Article 9(2)(i) of the GDPR—in conjunction with Article 89(1)—does not constitute separate authorization for data processing for research purposes [71]. Rather, the member states must make use of this opening clause through their own law. For Germany, this means that because of the federal system and the different competences of the federal government and federal states, the competence is initially based on the national competence regulations. Corresponding regulations can be found in section 27 of the Federal Data Protection Act; Article 25 of the Bavarian Data Protection Act, Act 25, section 27; and in Article 27 of the Bavarian Hospital Act. However, the decisive competence in this area usually lies with the federal states (exceptions may occur in individual cases, but in principle, hospital acts apply to all hospitals in a federal state, regardless of whether they are run publicly or privately or church-run). Furthermore, the regulations of the more specific Bavarian Hospital Act take precedence over those of the general Bavarian Data Protection Act (principle lex specialis derogat legi generali). Article 27(4) of the Bavarian Hospital Act states the following:

1. "Hospital physicians may use patient data insofar as this is necessary within the framework of the hospital medical treatment relationship, for initial, further and continuing training in the hospital, for research purposes in the hospital or in the research interests of the hospital."
2. "They may instruct other persons in the hospital to do so, insofar as this is necessary for the fulfillment of these tasks; for the purposes of research in accordance with sentence 1, they may permit other persons to use patient data if this is necessary for the implementation of the research project and the patient data remain in the custody of the hospital."
3. "Such persons shall be bound to secrecy."
4. "The hospital administration may use patient data to the extent necessary for the administrative processing of patient treatment."

The regulations are narrower than the general data protection requirements in Article 25 of the Bavarian Data Protection Act. In particular, the Hospital Act requires that the data remain in the custody of the hospital. This restriction is not provided for by the Data Protection Act. However, it is up to the member states to decide in which form and under which further conditions they make use of the opening clause. If the requirements of Article 27(4) of the Bavarian Hospital Act cannot be met, one of the other justification options for data processing from Article 9(2) GDPR must be used. Therefore, the current legal requirements in Bavaria hinder the use of patient data for research purposes. Other German federal states have similar regulations to Bavaria; however, some states (eg, Bremen or North Rhine-Westphalia) have opted for more detailed regulations. There is a need to examine such local regulations in more detail, not only in Germany but also in other countries and how these are affecting the secondary use of patient data.

Recent developments in the field of German public administration known as *norm screening* (*Normenscreening*) could potentially provide a systematic approach to minimize such legal barriers. To encourage digital transformation and reduce existing barriers, some German federal states (eg, Schleswig-Holstein and Saarland) reviewed their entire public state laws to identify all norms that could act as barriers to further digitalization. As a result, the federal state of Saarland identified and categorized more than 1000 rules, and those preventing, or at least complicating digitalization, were removed or at least changed into more moderate forms. Such a screening process could be used to identify all types of legislative obstacles, sort them, and enable fundamental changes. Therefore, it is suggested that a screening process could be used to harmonize and minimize barriers in regulations regarding the use of patient data for research purposes. The focus would be on regulations that are narrower than those required by EU or constitutional law. This screening process would likely be beneficial to all EU members, as member states are required to make use of the GDPR scientific research exemption through their own national law and will therefore likely face similar challenges as those described in Germany. This approach could also, in principle, be applied outside Europe.

In an era of increasing global collaborative health research efforts, however, significant variations in laws regarding this issue are not only a problem within countries but also across countries [18]. Concerns have long been raised that the GDPR allows too much room for interpretation of the regulation by member states on key aspects of data protection, including sufficient methods of pseudonymization, when data are considered fully nonidentifiable, what further limitations should be set on processing sensitive data for research purposes, and sufficient safeguards and conditions for processing data under research exemption [72]. Although this may help recognize local values and norms, it risks undermining the goal of the GDPR to address the heterogeneity of data protection within the EU. The process of norm screening on a national level could potentially help identify already existing similarities between member states as well as detect best practices, which could support the progress on EU-level.

Germany, and Bavaria in particular, prides itself for research impact and innovation potential. However, their current approach to patient data is often heavily skewed toward avoiding risks; thus making it extremely difficult to reach the goals of health data research. This approach also conflicts with their stated ambitions to be leaders in AI. Benefiting from data-intensive medicine, particularly activities driven by AI technologies, requires first and foremost, having access to data. Being very restrictive with secondary patient data use at the same time as pouring significant amounts of public funds into data-intensive medical and medical AI is inconsistent and arguably unethical, as it constitutes a waste of public resources and, at worse, may end up causing patient harm owing to unrepresentative and biased data sets and models [73]. Politicians and policy makers need to take the issue of data access and sharing more seriously.

XSL•FO
RenderX

Strengths and Limitations

This is a qualitative study that did not collect statistically representative data. However, we included a range of experts who have direct experience with sharing COVID-19–related health data for research purposes in Bavaria, which makes it likely that this study has captured key aspects of a multisided issue. A bias might exist toward the reporting of socially desirable attitudes [74]; however, given that our results are rather critical of current practice, we believe that such a bias is limited. The study was only carried out in Bavaria, and there may be some regional and country-specific differences that might limit the generalizability. Nevertheless, many of the key issues are associated with aspects that are common in other countries (eg, balancing the goals of public benefit of the research with consent and privacy), and these findings are likely to be of wider international interest. The strengths of this study include the fact that it is, to our knowledge, one of the first to investigate stakeholders' views and use of the GDPR scientific research exemption either during or before the pandemic.

Acknowledgments

This study was funded by the Technical University of Munich's Faculty of Medicine via the Special Programme for the Promotion of Corona Research by the Bavarian Council of Ministers. The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Data Availability

Our data include pseudonymized transcripts of interviews, which cannot be made publicly available in their entirety because of (1) the terms of our ethics approval and (2) because participants could be identifiable if the context of the entire transcript. This is in line with the current ethical expectations for qualitative interview research. We provide anonymized quotes within the paper to illustrate our findings (corresponding to transcript excerpts), and the complete interview guide used in the study has been included in the Multimedia Appendix 1.

Conflicts of Interest

AB is the chair of the German Ethics Council.

Multimedia Appendix 1
Interview guides.
[DOCX File , 23 KB - jmir_v24i6e38754_app1.docx]

References

1. McLennan S, Celi LA, Buyx A. COVID-19: putting the general data protection regulation to the test. JMIR Public Health Surveill 2020 May 29;6(2):e19279 [FREE Full text] [doi: 10.2196/19279] [Medline: 32449686]
2. Ienca M, Ferretti A, Hurst S, Puhan M, Lovis C, Vayena E. Considerations for ethics review of big data health research: a scoping review. PLoS One 2018 Oct 11;13(10):e0204937 [FREE Full text] [doi: 10.1371/journal.pone.0204937] [Medline: 30308031]
3. Topol EJ. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. New York, NY, USA: Basic Books; 2019.
4. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit Health 2019 Oct;1(6):e271-e297 [FREE Full text] [doi: 10.1016/S2589-7500(19)30123-2] [Medline: 33323251]
5. 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]
6. Schork NJ. Artificial intelligence and personalized medicine. Cancer Treat Res 2019;178:265-283 [FREE Full text] [doi: 10.1007/978-3-030-16391-4_11] [Medline: 31209850]
7. Woo M. An AI boost for clinical trials. Nature 2019 Sep;573(7775):S100-S102. [doi: 10.1038/d41586-019-02871-3] [Medline: 31554996]
8. Fleming N. How artificial intelligence is changing drug discovery. Nature 2018 May;557(7707):S55-S57. [doi: 10.1038/d41586-018-05267-x] [Medline: 29849160]
9. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019 Jun;6(2):94-98 [FREE Full text] [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]
10. Shibata T, Wada K. Robot therapy: a new approach for mental healthcare of the elderly - a mini-review. Gerontology 2011;57(4):378-386 [FREE Full text] [doi: 10.1159/000319015] [Medline: 20639620]

XSL•FO
RenderX

11. Fiske A, Henningsen P, Buyx A. Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. J Med Internet Res 2019 May 09;21(5):e13216 [FREE Full text] [doi: 10.2196/13216] [Medline: 31094356]
12. Lavender J. Investment in AI for healthcare soars. Klynveld Peat Marwick Goerdeler. 2018. URL: https://home.kpmg/xx/en/home/insights/2018/11/investment-in-ai-for-healthcare-soars.html [accessed 2022-05-28]
13. Taylor NT. Healthcare AI funding hits new high as sector matures. Medtechdive. 2019 Aug 7. URL: https://www.medtechdive.com/news/healthcare-ai-funding-hits-new-high-as-sector-matures/560396/ [accessed 2022-05-28]
14. Insights Team. AI And Healthcare: A Giant Opportunity. Forbes. 2019 Feb 11. URL: https://www.forbes.com/sites/insights-intelai/2019/02/11/ai-and-healthcare-a-giant-opportunity/#22e072d44c68 [accessed 2022-05-28]
15. ABIM Foundation, ACP-ASIM Foundation, European Federation of Internal Medicine. Medical professionalism in the new millennium: a physician charter. Ann Intern Med 2002 Feb 05;136(3):243-246 [FREE Full text] [doi: 10.7326/0003-4819-136-3-200202050-00012] [Medline: 11827500]
16. Beauchamp TL, Childress JF. Principles of Biomedical Ethics. 7th edition. Oxford, UK: Oxford University Press; 2013.
17. Celi LA, Mark RG, Stone DJ, Montgomery RA. "Big data" in the intensive care unit. Closing the data loop. Am J Respir Crit Care Med 2013 Jun 01;187(11):1157-1160 [FREE Full text] [doi: 10.1164/rccm.201212-2311ED] [Medline: 23725609]
18. McLennan S, Shaw D, Celi L. The challenge of local consent requirements for global critical care databases. Intensive Care Med 2019 Feb;45(2):246-248. [doi: 10.1007/s00134-018-5257-y] [Medline: 29922844]
19. Tu JV, Willison DJ, Silver FL, Fang J, Richards JA, Laupacis A, Investigators in the Registry of the Canadian Stroke Network. Impracticability of informed consent in the Registry of the Canadian Stroke Network. N Engl J Med 2004 Apr 01;350(14):1414-1421. [doi: 10.1056/NEJMsa031697] [Medline: 15070791]
20. Rocher L, Hendrickx JM, de Montjoye YA. Estimating the success of re-identifications in incomplete datasets using generative models. Nat Commun 2019 Jul 23;10(1):3069 [FREE Full text] [doi: 10.1038/s41467-019-10933-3] [Medline: 31337762]
21. Mostert M, Bredenoord AL, Biesaart MC, van Delden JJ. Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach. Eur J Hum Genet 2016 Jul;24(7):956-960 [FREE Full text] [doi: 10.1038/ejhg.2015.239] [Medline: 26554881]
22. Nyrén O, Stenbeck M, Grönberg H. The European Parliament proposal for the new EU General Data Protection Regulation may severely restrict European epidemiological research. Eur J Epidemiol 2014 Apr;29(4):227-230 [FREE Full text] [doi: 10.1007/s10654-014-9909-0] [Medline: 24802287]
23. Rumbold JM, Pierscionek B. The effect of the general data protection regulation on medical research. J Med Internet Res 2017 Feb 24;19(2):e47 [FREE Full text] [doi: 10.2196/jmir.7108] [Medline: 28235748]
24. McLennan S, Kahrass H, Wieschowski S, Strech D, Langhof H. The spectrum of ethical issues in a Learning Health Care System: a systematic qualitative review. Int J Qual Health Care 2018 Apr 01;30(3):161-168. [doi: 10.1093/intqhc/mzy005] [Medline: 29394354]
25. Faden RR, Kass NE, Goodman SN, Pronovost P, Tunis S, Beauchamp TL. An ethics framework for a learning health care system: a departure from traditional research ethics and clinical ethics. Hastings Cent Rep 2013;Spec No:S16-S27. [doi: 10.1002/hast.134] [Medline: 23315888]
26. Morain SR, Kass NE. Ethics issues arising in the transition to learning health care systems: results from interviews with leaders from 25 health systems. EGEMS (Wash DC) 2016 Mar 29;4(2):1212 [FREE Full text] [doi: 10.13063/2327-9214.1212] [Medline: 27141521]
27. Richter G, Krawczak M, Lieb W, Wolff L, Schreiber S, Buyx A. Broad consent for health care-embedded biobanking: understanding and reasons to donate in a large patient sample. Genet Med 2018 Jan;20(1):76-82 [FREE Full text] [doi: 10.1038/gim.2017.82] [Medline: 28640237]
28. Richter G, Borzikowsky C, Lieb W, Schreiber S, Krawczak M, Buyx A. Patient views on research use of clinical data without consent: legal, but also acceptable? Eur J Hum Genet 2019 Jun;27(6):841-847 [FREE Full text] [doi: 10.1038/s41431-019-0340-6] [Medline: 30683927]
29. Richter G, Borzikowsky C, Lesch W, Semler SC, Bunnik EM, Buyx A, et al. Secondary research use of personal medical data: attitudes from patient and population surveys in The Netherlands and Germany. Eur J Hum Genet 2021 Mar;29(3):495-502 [FREE Full text] [doi: 10.1038/s41431-020-00735-3] [Medline: 33005018]
30. Kelley M, James C, Alessi Kraft S, Korngiebel D, Wijangco I, Rosenthal E, et al. Patient perspectives on the learning health system: the importance of trust and shared decision making. Am J Bioeth 2015;15(9):4-17 [FREE Full text] [doi: 10.1080/15265161.2015.1062163] [Medline: 26305741]
31. Cho MK, Magnus D, Constantine M, Lee SS, Kelley M, Alessi S, et al. Attitudes toward risk and informed consent for research on medical practices: a cross-sectional survey. Ann Intern Med 2015 May 19;162(10):690-696 [FREE Full text] [doi: 10.7326/M15-0166] [Medline: 25868119]
32. Kraft SA, Cho MK, Constantine M, Lee SS, Kelley M, Korngiebel D, et al. A comparison of institutional review board professionals' and patients' views on consent for research on medical practices. Clin Trials 2016 Oct;13(5):555-565 [FREE Full text] [doi: 10.1177/1740774516648907] [Medline: 27257125]

XSL•FO
RenderX

33. Husedzinovic A, Ose D, Schickhardt C, Fröhling S, Winkler EC. Stakeholders' perspectives on biobank-based genomic research: systematic review of the literature. Eur J Hum Genet 2015 Dec;23(12):1607-1614 [FREE Full text] [doi: 10.1038/ejhg.2015.27] [Medline: 25735479]
34. Page SA, Manhas KP, Muruve DA. A survey of patient perspectives on the research use of health information and biospecimens. BMC Med Ethics 2016 Aug 15;17(1):48 [FREE Full text] [doi: 10.1186/s12910-016-0130-4] [Medline: 27527514]
35. Trinidad MG, Platt J, Kardia SL. The public's comfort with sharing health data with third-party commercial companies. Humanit Soc Sci Commun 2020;7(1):149 [FREE Full text] [doi: 10.1057/s41599-020-00641-5] [Medline: 34337435]
36. Seltzer E, Goldshear J, Guntuku SC, Grande D, Asch DA, Klinger EV, et al. Patients' willingness to share digital health and non-health data for research: a cross-sectional study. BMC Med Inform Decis Mak 2019 Aug 08;19(1):157 [FREE Full text] [doi: 10.1186/s12911-019-0886-9] [Medline: 31395102]
37. Willison DJ, Steeves V, Charles C, Schwartz L, Ranford J, Agarwal G, et al. Consent for use of personal information for health research: do people with potentially stigmatizing health conditions and the general public differ in their opinions? BMC Med Ethics 2009 Jul 24;10:10 [FREE Full text] [doi: 10.1186/1472-6939-10-10] [Medline: 19630941]
38. Caine K, Kohn S, Lawrence C, Hanania R, Meslin EM, Tierney WM. Designing a patient-centered user interface for access decisions about EHR data: implications from patient interviews. J Gen Intern Med 2015 Jan;30 Suppl 1:S7-16 [FREE Full text] [doi: 10.1007/s11606-014-3049-9] [Medline: 25480719]
39. Hutchings E, Loomes M, Butow P, Boyle FM. A systematic literature review of researchers' and healthcare professionals' attitudes towards the secondary use and sharing of health administrative and clinical trial data. Syst Rev 2020 Oct 12;9(1):240 [FREE Full text] [doi: 10.1186/s13643-020-01485-5] [Medline: 33046097]
40. Ballantyne A, Moore A, Bartholomew K, Aagaard N. Points of contention: qualitative research identifying where researchers and research ethics committees disagree about consent waivers for secondary research with tissue and data. PLoS One 2020 Aug 5;15(8):e0235618 [FREE Full text] [doi: 10.1371/journal.pone.0235618] [Medline: 32756563]
41. Ballantyne A, Moore A. Data and tissue research without patient consent: a qualitative study of the views of research ethics committees in New Zealand. AJOB Empir Bioeth 2018;9(3):143-153. [doi: 10.1080/23294515.2018.1518938] [Medline: 30407144]
42. Köngeter A, Jungkunz M, Winkler EC, Schickhardt C, Mehlis K. Sekundärnutzung klinischer daten aus der patientenversorgung für forschungszwecke – eine qualitative interviewstudie zu nutzen- und risikopotenzialen aus sicht von expertinnen und experten für den deutschen forschungskontext. In: Richter G, Loh W, Buyx A, von Kielmansegg SG, editors. Datenreiche Medizin und das Problem der Einwilligung: Ethische, rechtliche und sozialwissenschaftliche Perspektiven. Berlin, Germany: Springer; 2022:185-210.
43. Semler SC, Wissing F, Heyder R. German medical informatics initiative. Methods Inf Med 2018 Jul;57(S 01):e50-e56 [FREE Full text] [doi: 10.3414/ME18-03-0003] [Medline: 30016818]
44. Digital Healthcare Act (DVG): Driving the digital transformation of Germany's healthcare system for the good of patients. German Federal Ministry of Health. URL: https://www.bundesgesundheitsministerium.de/digital-healthcare-act.html [accessed 2021-05-26]
45. Gerke S, Stern AD, Minssen T. Germany's digital health reforms in the COVID-19 era: lessons and opportunities for other countries. NPJ Digit Med 2020 Jul 10;3:94 [FREE Full text] [doi: 10.1038/s41746-020-0306-7] [Medline: 32685700]
46. Heckmann D, Rachut S. Elektronische patientenakte und elektronische gesundheitskarte. In: Rehmann WA, Tillmanns C, editors. E-Health/Digital Health. Munich, Germany: C.H. Beck; 2022:282-312.
47. Einführung des E-Rezeptes auf unbestimmte Zeit verschoben. Deutscher Bundestag. 2022 Feb 14. URL: https://tinyurl.com/yckntm2s [accessed 2022-05-28]
48. Rachut S. Start der elektronischen Patientenakte in Deutschland - Aktueller Stand in der Praxis und der rechtswissenschaftlichen Diskussion. AnwZert ITR 2021;2021:2.
49. Söder M. Hightech Agenda Bayern: Regierungserklärung des Bayerischen Ministerpräsidenten Dr. Markus Söder, MdL, am 10 Oktober 2019 vor dem bayerischen Landtag. Bayerische Staatsregierung. 2019. URL: https://www.bayern.de/wp-content/uploads/2019/10/hightech_agenda_bayern.pdf [accessed 2022-05-28]
50. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
51. Satalkar P, McLennan S, Elger B, von Elm E, Matthias B. Investigators' sense of failure thwarted transparency in clinical trials discontinued for poor recruitment. J Clin Epidemiol 2022 Feb 04;145:136-143 [FREE Full text] [doi: 10.1016/j.jclinepi.2022.01.024] [Medline: 35124187]
52. McLennan S, Nussbaumer-Streit B, Hemkens LG, Briel M. Barriers and facilitating factors for conducting systematic evidence assessments in academic clinical trials. JAMA Netw Open 2021 Nov 01;4(11):e2136577 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.36577] [Medline: 34846522]
53. McLennan S, Griessbach A, Briel M, Making Randomized Trials Affordable (MARTA) Group. Practices and attitudes of Swiss stakeholders regarding investigator-initiated clinical trial funding acquisition and cost management. JAMA Netw Open 2021 Jun 01;4(6):e2111847 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.11847] [Medline: 34076698]

XSL•FO
RenderX

54. McLennan S. Rejected online feedback from a Swiss physician rating website between 2008 and 2017: analysis of 2352 ratings. J Med Internet Res 2020 Aug 03;22(8):e18374 [FREE Full text] [doi: 10.2196/18374] [Medline: 32687479]
55. McLennan S. The content and nature of narrative comments on Swiss physician rating websites: analysis of 849 comments. J Med Internet Res 2019 Sep 30;21(9):e14336 [FREE Full text] [doi: 10.2196/14336] [Medline: 31573918]
56. McLennan S. The ethical oversight of learning health care activities in Switzerland: a qualitative study. Int J Qual Health Care 2019 Oct 31;31(8):G81-G86. [doi: 10.1093/intqhc/mzz045] [Medline: 31066452]
57. McLennan S, Moore J. New Zealand district health boards' open disclosure policies: a qualitative review. J Bioeth Inq 2019 Mar;16(1):35-44. [doi: 10.1007/s11673-018-9894-1] [Medline: 30617731]
58. McLennan S, Strech D, Kahrass H. Why are so few patients rating their physicians on German physician rating websites? A qualitative study. BMC Health Serv Res 2018 Aug 29;18(1):670 [FREE Full text] [doi: 10.1186/s12913-018-3492-0] [Medline: 30157842]
59. McLennan S, Schwappach D, Harder Y, Staender S, Elger B. Patient safety issues in office-based surgery and anaesthesia in Switzerland: a qualitative study. Z Evid Fortbild Qual Gesundhwes 2017 Aug;125:23-29. [doi: 10.1016/j.zefq.2017.06.002] [Medline: 28711421]
60. Pless A, McLennan SR, Nicca D, Shaw DM, Elger BS. Reasons why nurses decline influenza vaccination: a qualitative study. BMC Nurs 2017 Apr 28;16:20 [FREE Full text] [doi: 10.1186/s12912-017-0215-5] [Medline: 28465672]
61. Pless A, Shaw D, McLennan S, Elger BS. Nurses' attitudes towards enforced measures to increase influenza vaccination: a qualitative study. Influenza Other Respir Viruses 2017 May;11(3):247-253 [FREE Full text] [doi: 10.1111/irv.12441] [Medline: 27943585]
62. McLennan SR, Diebold M, Rich LE, Elger BS. Nurses' perspectives regarding the disclosure of errors to patients: a qualitative study. Int J Nurs Stud 2016 Feb;54:16-22. [doi: 10.1016/j.ijnurstu.2014.10.001] [Medline: 25458803]
63. Kuhn E, Müller S, Teusch C, Tanner G, Schümann M, Baur C, et al. Interfaces of occupational health management and corporate social responsibility: a multi-centre qualitative study from Germany. BMC Public Health 2021 Jun 02;21(1):1042 [FREE Full text] [doi: 10.1186/s12889-021-11016-z] [Medline: 34078332]
64. Zimmermann BM, Fiske A, Prainsack B, Hangel N, McLennan S, Buyx A. Early perceptions of COVID-19 contact tracing apps in German-speaking countries: comparative mixed methods study. J Med Internet Res 2021 Feb 08;23(2):e25525 [FREE Full text] [doi: 10.2196/25525] [Medline: 33503000]
65. Samuel G, Roberts SL, Fiske A, Lucivero F, McLennan S, Phillips A, et al. COVID-19 contact tracing apps: UK public perceptions. Crit Public Health 2022 Jan 01;32(1):31-43 [FREE Full text] [doi: 10.1080/09581596.2021.1909707] [Medline: 35221546]
66. 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]
67. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Policy Ment Health 2015 Sep;42(5):533-544 [FREE Full text] [doi: 10.1007/s10488-013-0528-y] [Medline: 24193818]
68. Marshall MN. Sampling for qualitative research. Fam Pract 1996 Dec;13(6):522-525. [doi: 10.1093/fampra/13.6.522] [Medline: 9023528]
69. Fusch PI, Ness LR. Are we there yet? Data saturation in qualitative research. Qual Rep 2015 Sep 8;20(9):1408-1416. [doi: 10.46743/2160-3715/2015.2281]
70. Spitz M, Cornelius K, Jungkunz M, Schickhardt C. Rechtlicher rahmen für eine privilegierte nutzung klinischer daten zu forschungszwecken. Medizinrecht 2021 Jun 29;39(6):499-504 [FREE Full text] [doi: 10.1007/s00350-021-5898-7]
71. Heckmann D, Scheurer M. Datenschutzrecht. In: Heckmann D, Paschke A, editors. Praxiskommentar Internetrecht. Huntington, NY, USA: Juris Publishing; 2021.
72. Shabani M, Borry P. Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation. Eur J Hum Genet 2018 Feb;26(2):149-156 [FREE Full text] [doi: 10.1038/s41431-017-0045-7] [Medline: 29187736]
73. Bak M, Madai VI, Fritsche MC, Mayrhofer MT, McLennan S. You can't have AI both ways: balancing health data privacy and access fairly. Front Genet 2022 Jun 13;13:929453. [doi: 10.3389/fgene.2022.929453]
74. Bergen N, Labonté R. "Everything is perfect, and we have no problems": detecting and limiting social desirability bias in qualitative research. Qual Health Res 2020 Apr;30(5):783-792. [doi: 10.1177/1049732319889354] [Medline: 31830860]

Abbreviations

AI: artificial intelligence
COREQ: Consolidated Criteria for Reporting Qualitative Research
EU: European Union
GDPR: General Data Protection Regulation

XSL•FO
RenderX

Edited by G Eysenbach, T Leung; submitted 14.04.22; peer-reviewed by G Richter, M Raimi; comments to author 09.05.22; revised version received 28.05.22; accepted 29.05.22; published 27.06.22.

Please cite as:
McLennan S, Rachut S, Lange J, Fiske A, Heckmann D, Buyx A
Practices and Attitudes of Bavarian Stakeholders Regarding the Secondary Use of Health Data for Research Purposes During the COVID-19 Pandemic: Qualitative Interview Study
J Med Internet Res 2022;24(6):e38754
URL: https://www.jmir.org/2022/6/e38754
doi:10.2196/38754
PMID:35696598

©Stuart McLennan, Sarah Rachut, Johannes Lange, Amelia Fiske, Dirk Heckmann, Alena Buyx. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.06.2022. 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

Cybersickness Variability by Race: Findings From 6 Studies and a Mini Meta-analysis

Alison Jane Martingano[1], PhD; Ellenor Brown[2], PhD; Sydney H Telaak[1], BA; Alexander P Dolwick[1], BS; Susan Persky[1], PhD

[1]Social and Behavioral Research Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, United States
[2]Office of Science and Engineering Laboratories, US Food and Drug Administration, Silver Spring, MD, United States

Corresponding Author:
Susan Persky, PhD
Social and Behavioral Research Branch
National Human Genome Research Institute
National Institutes of Health
31 Center Drive, B1B36
Bethesda, MD, 20892
United States
Phone: 1 3014430098
Email: perskys@mail.nih.gov

Abstract

Background: With the influx of medical virtual reality (VR) technologies, cybersickness has transitioned from a nuisance experienced during leisure activities to a potential safety and efficacy concern for patients and clinicians. To improve health equity, it is important to understand any potential differences in cybersickness propensity among demographic groups, including racial groups.

Objective: This study aims to explore whether cybersickness propensity differs across racial groups.

Methods: We collected self-reported cybersickness ratings from 6 racially diverse independent samples within 1 laboratory group (N=931). In these studies, the participants were asked to perform tasks in VR such as traversing environments, pointing at and selecting objects, and interacting with virtual humans.

Results: Significant racial differences in cybersickness were found in 50% (3/6) of studies. A mini meta-analysis revealed that, on average, Black participants reported approximately one-third of SD less cybersickness than White participants (Cohen d=−0.31; P<.001), regardless of the nature of the VR experience. There was no overall difference in reported cybersickness between the Asian and White participants (Cohen d=−0.11; P=.51).

Conclusions: Racial differences in cybersickness indicate that researchers, practitioners, and regulators should consider patient demographics when evaluating VR health intervention outcomes. These findings lay the groundwork for future studies that may explore racial differences in cybersickness directly.

(J Med Internet Res 2022;24(6):e36843) doi:10.2196/36843

KEYWORDS

cybersickness; racial differences; virtual reality; head-mounted displays; simulator sickness

Introduction

Background

Cybersickness is a common negative physiological effect of exposure to virtual reality (VR), with symptoms similar to motion sickness, including disorientation, nausea, headache, and eye strain [1]. Recent technological advances have led to the widespread use of low-cost VR technologies. In turn, VR technologies that were developed and primarily used for gaming and entertainment have been applied broadly across areas such as education, industry, and medicine. Medicine, in particular, is a rapidly expanding application space for VR technologies, including new developments in a range of areas such as medical education and training [2], physical therapy and rehabilitation [3], surgical planning [4], pain management [5-7], psychotherapy [8,9], and treatment for ophthalmic disorders [10]. As medical VR technologies have become more

XSL•FO
RenderX

commonplace, cybersickness has transitioned from a nuisance to a potential safety and efficacy concern. Cybersickness concerns have prompted researchers [11-14], professional groups [15], standards organizations [16], and the US Food and Drug Administration [17] to address prevention, assessment, and mitigation strategies.

When researchers and clinicians are making benefit-risk determinations for emerging VR technologies, they should consider cybersickness propensity. Although cybersickness is a well-known response to VR exposure, the full range of causes and risk factors are not well understood. This knowledge gap may be a barrier in assessing the safety and effectiveness of VR technologies for all users. Thus far, the scientific literature has identified several factors that are linked to differences in cybersickness risk. It is well known that elements of VR content, VR hardware, and the interface between them influence cybersickness outcomes [18,19]. Certain demographic and within-person factors have also been connected with propensity to experience cybersickness, such as age [18,20-23], sex and gender [18,22,24-26], BMI [27], and health and health history [28-30]. A large meta-analysis of existing literature (k=137) recently found that various individual differences predict cybersickness propensity, including gender, real-world experience, technological experience, possessing a neurological disorder, and possessing a relevant phobia [31]. However, this meta-analysis did not consider race as a potential moderator.

Racial Differences in Cybersickness and Motion Sickness

Studies investigating potential cybersickness differences by user race could provide valuable new insights into potential inequities in VR accessibility, which is critical for ensuring that this emerging technology is accessible to all in the future. Currently, such studies are lacking in the cybersickness literature. However, early research has suggested that there are racial differences in motion sickness propensity. For example, a series of studies conducted in the United States found that Asian participants reported more motion sickness symptoms than White and Black participants [32-34]. This racial difference was maintained regardless of whether the Asian participants were born in the United States or were recent immigrants [35]. This led the authors to posit an evolutionary and genetic basis for these differences. However, this conclusion is at odds with the modern understanding that race is a social construct rather than a biological or genetic one. As such, the identified differences in motion sickness reporting by race may alternatively reflect cultural and social differences that result, in part, from systemic differential treatment. Various other sociocultural factors may contribute to racial differences in reporting of discomfort such as language, acculturation, learning and cultural conditioning, and attention to uncomfortable stimuli (refer to Lasch [36]). More recent research conducted in Germany found that Asian participants reported less motion sickness than White participants [37]. However, this study found that Asian participants had a shorter tolerance for rotation despite reporting less motion sickness, which may indicate differences in motion sickness reporting that are separate from physical experience. Overall, the existing research on racial differences in motion sickness is limited. Moreover, cybersickness, although related to motion sickness and simulator sickness, is a distinct phenomenon, with disorientation being more common and oculomotor symptoms being less common [38]. Given these differences between motion sickness and cybersickness, racial variability in cybersickness warrants investigation.

In anticipation of evaluating VR-based medical product efficacy alongside VR-associated risks across patient demographics, it is important to understand any potential underlying differences between groups related to these outcomes. Addressing this knowledge gap aligns with an increasing regulatory focus on health equity [39] as well as related efforts to promote diversity in study populations and evaluate potential differential product outcomes by patient demographics [40]. The potential for race-related variability in the performance of medical technologies was recently illustrated by a safety communication on the limitations of pulse oximeter devices, which highlighted the potential accuracy differences between patients with dark and light skin pigmentation [41]. Similarly, medical use of VR may be susceptible to racial inequities in ways that have not yet been uncovered. Thus, although theory and previous literature do not provide a clear path toward hypothesizing racial differences in cybersickness, it is important to explore existing data associated with VR use to determine whether racial variability in cybersickness exists. Understanding the differences in cybersickness propensity based on race is critical to ensuring that this emerging technology is accessible to all in the future.

Currently, studies exploring racial differences in cybersickness are lacking. To address this gap in the literature, we reported data from 6 independent samples collected within 1 laboratory group. In these studies, participants were asked to perform various tasks in VR such as traversing environments, pointing at and selecting objects, and interacting with virtual humans. The analyses compared self-identified Black and Asian participants' reporting of cybersickness to that of self-identified White participants. Comparisons between Black and Asian participants were also included in individual studies, where feasible. These 3 racial groups were chosen for comparison because they were well represented across all study samples and represent groups of interest for potential disparities when evaluating VR health care devices for use in the United States. White participants were chosen as the comparison group in the analyses because they are the most represented racial group in the existing literature. We also report a mini meta-analysis to illustrate the overall trends across all 6 studies. Together these studies are intended to reveal any differences in reported cybersickness between racial groups and lay the groundwork for future studies that may explore these differences directly. To the best of our knowledge, this is the first report of racial differences in cybersickness in the literature and is therefore a critical first step that should be explored in future research. Ultimately, addressing racial differences in cybersickness will help to move toward greater health equity.

XSL•FO
RenderX

Methods

Overview

This analysis included data from 6 experimental trials conducted for other purposes (Table 1). All studies were conducted between 2009 and 2020 through the Immersive Simulation Program at the National Human Genome Research Institute, National Institutes of Health. All research participants were recruited from the local community. VR was used at the program's laboratory facility.

Each study used one of 2 types of VR settings: a buffet restaurant environment called the VR buffet [42] or a clinical examination room environment. Both VR programs were created using the Vizard VR platform [43]. Studies were selected for inclusion because they administered measures of participant cybersickness symptoms using a variant of the Short Symptoms Checklist (SSC) [44], because they recorded participants' self-reported race, and because data were available for analysis. In all studies, the possibility of experiencing cybersickness was communicated to participants both in the consent form and by the research assistant during the study. Participants were told that they were welcome to stop the study if they experienced any cybersickness symptoms. This rarely occurred in practice. More details about each study are available in the original publications [45-50].

Table 1. Characteristics of the virtual reality (VR) environment for each research study.

	Year	Content	Locomotion	Headset	Aim of the study
Study 1	2017	VR buffet	Walking	HTC Vive	Measure the influence of messages about children's diet on parents' feeding behavior
Study 2	2011	VR buffet	Walking	nVisor SX60	Measure the influence of children's risk information provision on parents' feeding behavior
Study 3	2009	Virtual clinic	Walking	nVisor SX60	Assess medical students' reaction to a virtual patient's weight in a clinical scenario
Study 4	2020	Virtual clinic	Seated	HTC Vive Pro	Assess medical students' use of a virtual patient's genomic risk information in a clinical scenario
Study 5	2014	Virtual clinic	Seated	nVisor SX60	Assess reaction of women with overweight to virtual provider's messages
Study 6	2012	Virtual clinic	Seated	nVisor SX60	Assess reaction of women with overweight to virtual provider's messages

VR Environments

The VR Buffet

The VR buffet is a simulated buffet restaurant in which parental food choices for their child are assessed by tracking the parents' virtual food selections. Outcomes for the VR buffet are a validated measure of parental food choices [42]. The participants' physical movements drive the viewpoint in the virtual world, such that walking around the physical room corresponds to walking around the virtual buffet. Participants made food selections in the virtual buffet using a controller. Once all food and drink selections were made, participants selected a virtual cash register to indicate completion. Figure 1 shows the VR buffet environment.

Figure 1. Screenshots of buffet and clinical virtual reality environments.

VR Clinical Simulations

Several VR clinical simulations are included in which participants are immersed in a virtual medical examination room as either the provider or patient and asked to interact verbally with a virtual human playing the opposite role. When medical students (as opposed to patients) are the users, they are also asked to read information about the virtual patient's medical

XSL•FO
RenderX

records on a virtual computer monitor or tablet situated within the VR environment. A research assistant controlled the prerecorded statements of the virtual human interaction partner. In most cases, users are seated in this virtual environment, although there is also a version in which users can walk around and approach their virtual interaction partner. Figure 1 shows a sample VR clinical simulations.

VR Equipment

All studies were conducted within the same physical laboratory environment, which consisted of a room fitted with a 6-dof VR headset system. The headset and equipment used differed across studies (Table 1 provides information on the system used in each study). The earlier VR system included an NVIS nVisor SX60 headset with a WorldViz Precision Point Tracking System. A handheld presentation pointer was modified to provide hand control of the selection tool in the VR buffet environment. Later systems included an HTC Vive headset with an integrated tracking system or an HTC Vive Pro headset with an integrated tracking system. In both cases, the relevant Vive or Vive Pro controllers were used for hand control when needed.

Study Inclusion and Exclusion Criteria

Several inclusion and exclusion criteria (eg, gender, age, and parental status) varied between studies based on the content of the specific research study. All studies also had exclusion criteria related to the use of the VR equipment. In all studies, potential participants were excluded if they reported having epilepsy, seizures, or a vestibular disorder or if they reported having vision or hearing that was neither normal nor corrected to normal. In most studies, a known pregnancy was also an exclusion criterion. Potential participants were excluded if they reported higher levels of propensity to motion sickness. Participants were asked the following question: “How easily would you say that you get motion or car sickness on a scale of 1 to 7 where 1 would be that you ‘never get motion sick’ and 7 would be that you ‘get sick easily’?” Those who answered 6 or 7 on a 7-point scale were deemed ineligible for participation. As such, all participants in the included studies were individuals who did not identify themselves as being particularly vulnerable to motion sickness symptoms.

All participants were encouraged to adjust the VR headset themselves while viewing the VR environment. For example, participants were asked to move the headset up and down on their face and then to use the adjustment that would “move the lenses closer together and further apart, allowing [them] to focus.” They were repeatedly asked whether the virtual room looked blurry and were instructed to make adjustments until they felt that their view of the room was clear. After completing the VR experience, the participants were asked to rate their level of cybersickness as part of a larger questionnaire. Demographic information including self-reported race was collected at pretest or during the laboratory visit.

Measures

All studies included in this analysis administered a variant of the Simulator Sickness Checklist, the SSC [43]. The SSC is a commonly used self-report measure of cybersickness that contains a subset of symptoms used in the longer Simulator Sickness Questionnaire [50]. Most of the included studies used a 5-item version of the SSC that assessed headache, blurred vision, dizziness with eyes open, dizziness with eyes closed, and nausea. Clinical studies involving women with higher weight as participants (studies 5 and 6) used a 6-item version that additionally assessed eyestrain. In all studies, each item was measured on a 4-point Likert-type scale; some studies began this scale at zero, whereas others began at one. The lowest end point was labeled *none* or *not at all*, and the highest end point was labeled *severe* (Table 2). For each study, the composite cybersickness score was calculated by summing the responses for each item on the scale. In all analyses, we retained all original scale items and response options (without transformation), as these may have influenced participant responses [51]. Therefore, we caution readers that it is not possible to compare raw cybersickness scores among the included studies. We performed a mini meta-analysis for this purpose.

The primary predictor in our analysis was the race of participants included in the study. We considered the number of White, Black, and Asian participants based on each participant’s self-reported racial background. For a racial group to be considered for analysis within a given study, at least 10 participants in the study needed to self-identify with that racial group.

Additional variables assessed included self-reported gender and age. BMI was calculated from weight and height, which was self-reported except in the case of the 2 studies of medical students (studies 3 and 4), where it was measured in the laboratory. The time spent in the VR environment was automatically calculated using the VR environment software. The year of the study was determined as the year in which the last participant data collection visit occurred. Table 3 provides a summary of the demographic variables for each study.

XSL•FO
RenderX

Table 2. Self-reported cybersickness symptoms by racial group.

Scale	Racial group	Severity rating of cybersickness symptoms for each item, mean (SD)						Composite cybersickness, mean (SD)
		Headache	Eyestrain	Blurred vision	Dizzy (eyes open)	Dizzy (eyes closed)	Nausea	
Study 1 (0=none, 1=slight, 2=moderate, 3=severe)[a]								
	White	0.15 (0.44)	N/A[b]	0.48 (0.61)	0.39 (0.63)	0.09 (0.33)	0.13 (0.37)	1.23 (1.68)
	Black	0.04 (0.20)	N/A	0.34 (0.60)	0.13 (0.45)	0.11 (0.38)	0.00 (0.00)	0.57 (1.13)
	Asian	0.22 (0.42)	N/A	0.26 (0.45)	0.26 (0.45)	0.07 (0.27)	0.04 (0.19)	0.85 (1.10)
	Total	0.13 (0.39)	N/A	0.40 (0.58)	0.29 (0.57)	0.09 (0.33)	0.07 (0.29)	0.98 (1.49)
Study 2 (0=none, 1=slight, 2=moderate, 3=severe)[a]								
	White	0.18 (0.43)	N/A	0.44 (0.63)	0.33 (0.56)	0.19 (0.46)	0.09 (0.30)	1.23 (1.37)
	Black	0.11 (0.32)	N/A	0.29 (0.48)	0.19 (0.39)	0.05 (0.22)	0.06 (0.24)	0.69 (0.94)
	Total	0.15 (0.39)	N/A	0.38 (0.58)	0.27 (0.50)	0.13 (0.38)	0.08 (0.28)	1.01 (1.23)
Study 3 (0=none, 1=slight, 2=moderate, 3=severe)[a]								
	White	0.12 (0.36)	N/A	0.40 (0.57)	0.36 (0.59)	0.12 (0.43)	0.14 (0.40)	1.14 (1.68)
	Black	0.06 (0.24)	N/A	0.38 (0.55)	0.38 (0.55)	0.00 (0.00)	0.03 (0.17)	0.86 (1.05)
	Asian	0.27 (0.54)	N/A	0.38 (0.61)	0.42 (0.68)	0.15 (0.36)	0.21 (0.46)	1.42 (2.03)
	Total	0.15 (0.40)	N/A	0.39 (0.57)	0.38 (0.61)	0.10 (0.37)	0.14 (0.39)	1.16 (1.69)
Study 4 (0=none, 1=slight, 2=moderate, 3=severe)[a]								
	White	0.12 (0.41)	N/A	1.00 (0.78)	0.12 (0.41)	0.00 (0.00)	0.00 (0.00)	1.19 (1.19)
	Black	0.00 (0.00)	N/A	0.81 (0.91)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.73 (0.88)
	Asian	0.08 (0.28)	N/A	0.56 (0.65)	0.08 (0.28)	0.00 (0.00)	0.00 (0.00)	0.78 (0.90)
	Total	0.08 (0.32)	N/A	0.81 (0.78)	0.08 (0.32)	0.00 (0.00)	0.00 (0.00)	0.96 (1.05)
Study 5 (1=not at all, 4=severe)[c]								
	White	1.10 (0.31)	1.56 (0.66)	1.44 (0.56)	1.08 (0.27)	1.00 (0.00)	1.01 (0.11)	7.19 (1.27)
	Black	1.03 (0.18)	1.33 (0.50)	1.36 (0.55)	1.03 (0.18)	1.03 (0.18)	1.00 (0.00)	6.81 (1.05)
	Total	1.07 (0.25)	1.45 (0.59)	1.40 (0.56)	1.06 (0.23)	1.02 (0.13)	1.01 (0.08)	7.01 (1.18)
Study 6 (1=not at all, 4=severe)[c]								
	White	1.14 (0.39)	1.36 (0.58)	1.32 (0.54)	1.03 (0.18)	1.03 (0.18)	1.02 (0.13)	6.91 (1.39)
	Black	1.13 (0.51)	1.24 (0.49)	1.24 (0.49)	1.06 (0.28)	1.04 (0.19)	1.05 (0.25)	6.75 (1.56)
	Total	1.13 (0.47)	1.28 (0.52)	1.27 (0.51)	1.05 (0.25)	1.04 (0.19)	1.04 (0.22)	6.81 (1.50)

[a]Scale minimum: 0; scale maximum: 15.

[b]N/A: not applicable.

[c]Scale minimum: 6; scale maximum: 24.

XSL•FO
RenderX

Table 3. Demographic variables for each study.

Study	Racial group	Sample, n	Age (years), mean (SD)	Gender (female), n (%)	BMI (kg/m^2), mean (SD)	Time in virtual reality (seconds), mean (SD)
Study 1						
	White	88	38.08 (5.72)	60 (68)	25.87 (5.51)	318 (323)
	Black	48	36.31 (6.35)	33 (69)	31.63 (9.32)	301 (294)
	Asian	27	39.12 (4.76)	18 (67)	25.64 (5.90)	306 (214)
Study 2						
	White	105	38.89 (5.33)	105 (100)	30.18 (4.78)	409 (491)
	Black	75	35.81 (5.62)	75 (100)	31.10 (5.18)	389 (452)
Study 3						
	White	104	26.55 (2.25)	49 (47)	23.92 (2.85)	414 (119)
	Black	34	26.56 (3.74)	21 (62)	26.08 (4.43)	433 (125)
	Asian	48	25.77 (2.48)	24 (50)	22.95 (3.74)	414 (119)
Study 4						
	White	34	26.41 (2.66)	20 (59)	23.26 (3.63)	664 (252)
	Black	16	26.06 (1.84)	12 (75)	25.56 (4.57)	663 (218)
	Asian	25	25.96 (1.57)	15 (60)	23.53 (4.29)	721 (185)
Study 5						
	White	88	35.24 (9.65)	88 (100)	31.25 (5.25)	253 (21)
	Black	85	35.55 (8.16)	85 (100)	35.55 (8.16)	255 (22)
Study 6						
	White	58	35.35 (8.71)	58 (100)	36.33 (7.66)	423 (60)
	Black	109	36.07 (11.24)	109 (100)	32.07 (6.03)	427 (70)

Data Analysis

For each individual study, we conducted an ANOVA to examine the relationship between participant race (2 or 3 groups depending on whether Asian participants were present in sufficient numbers to be included) and cybersickness. When there were 3 racial groups, we examined planned contrasts to assess the differences between individual racial groups. We also examined zero-order correlations between cybersickness and 3 person-level variables: age, BMI, and time spent in the VR environment (Multimedia Appendix 1, Tables S1-S5). When these variables demonstrated significant relationships with cybersickness, they were included as covariates in an additional analysis of covariance.

We also conducted a random effects meta-analysis [52] using Comprehensive Meta-Analysis V3 software [53] to determine the overall difference in self-reported cybersickness between racial groups in our 6 studies. Analyses were performed using Cohen *d* with weighted averages of the effect sizes.

Ethics Approval

Participants were compensated for participation in all studies, and all studies were approved by the relevant institutional review boards. IRB review approval numbers are as follows: 08HG0122, 10HG0076, 11HG0238, 13HG0125, and 16HG0026.

Results

Cybersickness Levels Overall

Self-reported cybersickness was very low in all racial groups (Table 2). On average, participants reported no to slight symptoms, with blurred vision and eyestrain being reported the most and nausea and dizziness (with eyes closed) being reported the least.

Relationships Between Race and Cybersickness in Individual Studies

Study 1: Buffet With Parents

ANOVA revealed a significant difference in cybersickness by racial group (Table 4). Pairwise comparisons showed that Black participants reported lower levels of cybersickness than White participants. There was no significant difference between White and Asian participants. No other person-level variables showed a significant relationship with cybersickness (Multimedia Appendix 1, Table S1).

XSL•FO
RenderX

Table 4. Results from ANOVA within individual studies.

	Effect of racial group, omnibus analysis		Pairwise comparisons					
	F test (*df*)	*P* value	White vs Black			White vs Asian		
			Mean difference	*P* value	Bonferroni corrected *P* value	Mean difference	*P* value	Bonferroni corrected *P* value
Study 1	3.34 (2,157)	.03	0.84	.004	.02	0.44	.19	.38
Study 2	7.75 (1,180)	.006	N/A[a]	N/A	N/A	N/A	N/A	N/A
Study 3	1.12 (2,183)	.33	0.29	.38	.76	−0.27	.36	.72
Study 4	2.07 (2,72)	.13	0.42	.18	.84	0.52	.06	.12
Study 5	5.18 (1,173)	.02	N/A	N/A	N/A	N/A	N/A	N/A
Study 6	0.36 (1,166)	.55	N/A	N/A	N/A	N/A	N/A	N/A

[a]N/A: not applicable (pairwise comparisons are only reported for studies with more than 2 racial groups).

Study 2: Buffet With Mothers Only

The ANOVA revealed a significant difference in cybersickness by racial group (Table 4), wherein Black participants reported lower levels of cybersickness than White participants. There was also a significant relationship between age and cybersickness, with older participants reporting more cybersickness. However, the main effect of race on cybersickness was maintained when age was added as a covariate ($F_{1,175}$=5.33; *P*=.02). No other person-level variables showed a significant relationship with cybersickness (Multimedia Appendix 1, Table S2).

Study 3: Clinical With Medical Students

The ANOVA did not show a significant effect of race on cybersickness. Among the person-level variables, there was a significant relationship between age and cybersickness, with older participants reporting greater cybersickness (Multimedia Appendix 1, Table S3). However, the analysis of covariance also did not show a main effect of race on cybersickness ($F_{1,180}$=2.32; *P*=.44).

Study 4: Clinical With Medical Students

The ANOVA did not show a significant effect of race on cybersickness. No other person-level variables showed a significant relationship with cybersickness (Multimedia Appendix 1, Table S4).

Study 5: Clinical With Women Only

The ANOVA revealed a significant difference in cybersickness by racial group (Table 4), wherein Black participants reported lower levels of cybersickness than White participants. No other person-level variables showed a significant relationship with cybersickness (Multimedia Appendix 1, Table S5).

Study 6: Clinical With Women Only

The ANOVA did not show a significant effect of race on cybersickness. No other person-level variables showed a significant relationship with cybersickness (Multimedia Appendix 1, Table S6).

Mini Meta-analysis

Overview

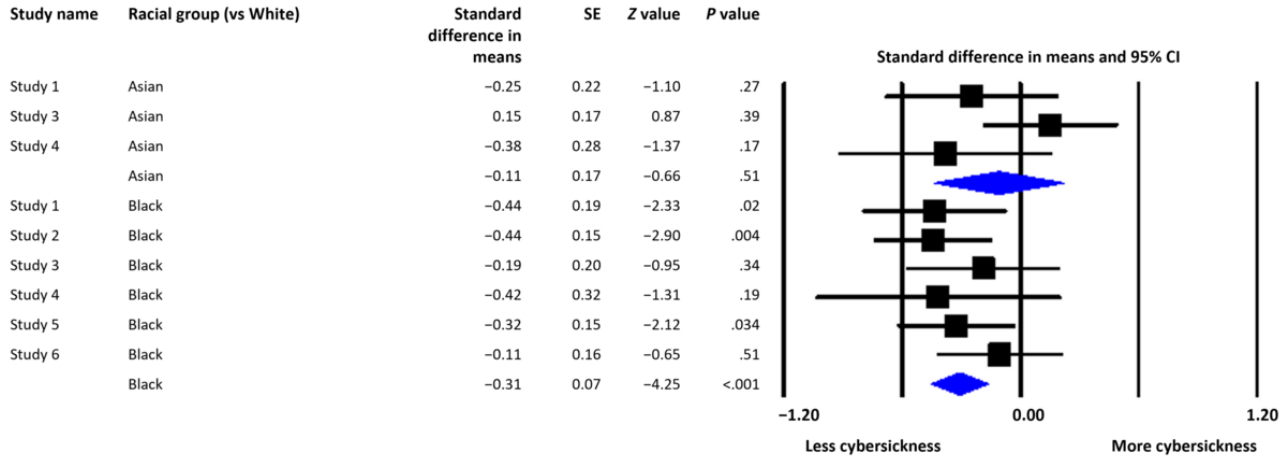
Although the results of each individual study reported above are revealing, determining whether there are racial differences in cybersickness, based on whether individual studies report a statistically significant difference, is inherently flawed. By conducting a mini meta-analysis, we were able to determine the size of these effects. In addition, by combining data in a meta-analysis, we reduced the impact of random error and increased the precision of our estimate. This increased precision allowed us to detect racial differences in cybersickness that individual studies may lack the power to detect.

In our mini meta-analysis (Figure 2), effect sizes indicated the difference in reported cybersickness between White participants and Black and Asian participants. Positive effect sizes indicate that Black and Asian participants report more cybersickness than White participants, whereas negative effects indicate that Black and Asian participants report less cybersickness than White participants. Moderator analyses were conducted using mixed-effect models.

Overall, Black participants reported significantly less cybersickness than White participants (Cohen *d*=−0.31; *P*<.001; κ=6; Figure 2). On average, Black participants reported approximately one-third of an SD less cybersickness than White participants. Asian participants did not report significantly different cybersickness levels compared with White participants (Cohen *d*=−0.11; *P*=.51; κ=3; Figure 2).

XSL•FO
RenderX

Figure 2. Forest plot depicting the standardized mean difference (Cohen *d*) in reported cybersickness for Black and Asian participants compared with White participants.



Moderator Analyses

To explore whether the racial differences in reported cybersickness between Black and White participants may be exaggerated or attenuated in certain situations, we conducted exploratory moderator analyses. Specifically, we evaluated the VR environment (buffet vs clinical), movement (seated vs standing), headset type (nVisor SX60 vs Vive), duration of experience, and year of data collection. Asian participants were excluded from these moderator analyses. This exclusion was a conservative approach to ensure that White participants (the comparison group) were included only once in the analysis. This approach prevented artificial inflation of N and the overestimation of the precision of the effect.

The moderator analyses did not reveal any variables that attenuated racial differences in cybersickness. Black participants reported less cybersickness than White participants regardless of the nature of the VR experience. Specifically, the magnitude of the racial difference was not significantly different based on whether participants engaged with the VR buffet or the clinical VR scenario (Q_1=2.155; *P*=.14). Racial differences were also unchanged regardless of whether the participants were seated or standing (Q_1=0.79; *P*=.37). Racial differences in cybersickness were also consistent regardless of the type of headset (Q_1=0.700; *P*=.40) and duration of the VR experience (B=−0.0001; 95% CI −0.002 to 0.002; *Z*=−0.14; *P*=.88). In addition, the year of study did not moderate racial differences in reported cybersickness (B=−0.0195, 95% CI −0.068 to 0.030; *Z*=−0.78; *P*=.44).

Among the studies included in this analysis, racial differences in cybersickness appear robust and consistent across various VR experiences and experimental designs. Overall, Black participants reported less cybersickness than White participants regardless of the nature of the VR experience.

Discussion

Principal Findings

This study presents the first known examination of racial differences in cybersickness. We found that, on average, Black participants reported less cybersickness than White participants, and our analyses did not reveal any moderators that attenuated this racial difference. In contrast to previous research on motion sickness [32-35,37], we found no differences in reported cybersickness between White and Asian participants. However, this comparison should be interpreted with caution, as cybersickness differs from other forms of motion sickness [38] and as we excluded potential participants who reported that they were particularly prone to motion sickness. Although our results require replication, they indicate that researchers, practitioners, and regulators may need to consider the potential for racial differences in cybersickness when evaluating VR applications for their side effects.

Potential Explanations for Racial Differences in Cybersickness

The data reported here do not allow us to determine why different racial groups reported varying levels of cybersickness. There are a multitude of factors that could influence pathways through which individuals differ in their propensity for and reporting of cybersickness. Some of these factors are discussed below. Although our data do not support any individual causal mechanism, we highlight where theories are consistent or inconsistent with our findings. Importantly, it is likely that multiple causal forces influence cybersickness experience and reporting simultaneously, working together or in opposition with one another. Further research is needed to determine the possible causal mechanisms behind racial differences in cybersickness.

One reason that people differ in their propensity for cybersickness is their previous experience with VR [54,55]. Previous research has found that people who have never used VR report more cybersickness than people who rarely use VR, and both groups report more cybersickness than those who use VR weekly [56]. Familiarity with VR is therefore associated with lower levels of cybersickness. There is no reason to believe that familiarity with VR explains the results of our studies, as most of the research was conducted before VR became a common consumer device, and we saw no evidence that racial differences in cybersickness were attenuated or augmented in more recent years. Nevertheless, familiarity with VR technology

XSL•FO
RenderX

may be a piece of this puzzle going forward. A recent survey of British internet users found that people of color are overrepresented in the VR consumer market [57]. This is supported by market research findings that Black and Hispanic consumers are more aware of and interested in VR than White consumers [58]. Therefore, familiarity-related method equivalence [59] should be monitored in future VR studies.

Another reason people differ in their propensity for cybersickness is differences in body size and proportions. Oculomotor cybersickness has been shown to decrease with higher BMI [27]. However, in these studies, racial differences largely remained when BMI was entered as a covariate, with the exception of study 1 (Multimedia Appendix 1, Table S7). Another potential factor is participants' interpupillary distance (IPD) and goodness of headset fit. Previous research has shown that people with an IPD that is poorly accommodated by standard VR headsets are more likely to experience cybersickness [14] and that the inability to accommodate diverse bodies may lead to apparent demographic differences in cybersickness. In 2 studies, researchers demonstrated that women, for whom a VR headset built for men's IPD specifications did not fit properly, reported higher cybersickness than men. However, when women's headsets were fitted to them correctly, they experienced cybersickness at a rate similar to that of men [14]. As we did not measure the IPD of our participants or the goodness of headset fit, we cannot rule out the possibility of variation within our sample and thus cannot explore whether these are a potential explanation for our results. These limitations should be addressed in future research. Previous research has suggested that in some cases, self-reported racial identity is associated with differences in average IPD measurements (refer to the study by Dodgson [60]); however, it is unknown how this might influence the fit of VR headsets or subsequent cybersickness.

Individuals may also differ in their likelihood of reporting cybersickness, just as there are individual differences in reporting other types of pain and discomfort. With some exceptions, people of color generally report higher levels of pain than White patients [61-65] (refer to the study by Plesh et al [66] for contrasting results). Systemic barriers to adequate pain management and a long history of discrimination and dehumanization likely explain these trends [67]. Various other cultural factors may also contribute to differences in reporting, including language, acculturation, learning and cultural conditioning, degree of expressiveness, heightened attention to painful stimuli, and coping styles (refer to the study by Booker [61]). Given that our results contradict much of the existing research on pain and discomfort, further research is needed to understand how sociocultural factors may specifically influence the reporting of cybersickness.

Another factor relevant to cybersickness reporting is the individual differences in how people respond to questionnaires, particularly Likert-type scales. Scales that range from 1=strongly disagree to 5=strongly agree or from 1=poor to 5=excellent are often used in health research but are troublesome because responses are influenced not only by the content of the question but also by general approaches to answering such questions. Researchers have documented that certain response styles are more common among specific racial and ethnic groups. For example, Black and Latino study participants are especially likely to use the extreme positive end of rating scales [59], whereas East Asian participants are more likely to select scale midpoints and avoid extreme responses compared with North American participants [68,69]. It has been hypothesized that these response styles reflect various cultural values on which people of color generally differ from White participants [59]. In particular, manifestations of social desirability may differ across cultures [70] in ways that result in different response styles [59]. It is difficult to know how such response styles would manifest on the cybersickness scale used in this study, which ranges from *none* to *severe*. Nevertheless, it is certainly possible that racial differences in response styles may explain the differences in cybersickness reporting between Black and White participants in our research. Future research would benefit from using more objective, passive physiological approaches to measuring cybersickness such as electroencephalography [71] to overcome difficulties with self-reporting. Nevertheless, it is important to understand potential racial differences in self-reported cybersickness, for example, when evaluating novel VR medical interventions.

Limitations

In addition to the limitations of our design that prevent us from examining why racial differences in cybersickness occur, there are other important limitations to our sample that should be considered when interpreting the results of these studies. First, we recruited participants from Washington, District of Columbia area. Participants were aware that they were volunteering for a VR experiment, and we excluded individuals who reported a high propensity for motion sickness. Therefore, the resulting samples are more likely to be interested in VR and may be less likely to experience cybersickness than the general population. In practice, exclusion because of motion sickness propensity was rare. For example, in one included sample [50], only 1.47% of potential study participants were ineligible because of this factor. Therefore, we believe that these results are a useful starting point, particularly for designing and evaluating VR medical interventions for use with populations that are not especially susceptible to cybersickness. It is worth noting that American participants represent a society that is not typical of the world's population, which limits its representativeness [72,73]; therefore, there is no reason to expect that these same racial differences would be found outside of the US context.

Second, we excluded participants who did not identify as Asian, Black, or White from this analysis. This decision was made to ensure that we had sufficient power to detect racial differences in cybersickness. However, we were unable to draw any conclusions regarding racial groups that were not well represented in our samples. Future research should attempt to oversample people of color to achieve a sufficient sample size for other racial comparisons. Another limitation is that we excluded individuals who identified as more than one race. This may have artificially created distinct racial groups that in reality are much less coherent and discrete. In addition, we did not find many of the demographic correlations with cybersickness that have been reported in previous literature (ie, age, BMI, and time

XSL•FO
RenderX

spent in VR). This is likely because of the limited range in age, BMI, and exposure time in our reported studies.

Another limitation is that we assessed cybersickness following the use of only 2 types of VR environments, neither of which is characteristic of the types of VR environments that typically elicit significant cybersickness (eg, sensory conflict and imposed motion [74]). Therefore, perhaps unsurprisingly, cybersickness ratings across all racial groups were very low. Although we anticipate that many health- and medicine-oriented VR applications will be designed to minimize cybersickness, mild cybersickness may still be of practical significance. Many medical VR experiences are designed for repeated use over time (eg, exposure therapy and pain management), and mild cybersickness may increase attrition. A recent meta-analysis of attrition in VR exposure for anxiety disorders found that attrition ranged from 2% to 41% [75]. Unfortunately, these studies rarely reported the reasons for dropout. Nevertheless, when reasons were given, cybersickness was among the top 5, accounting for 6.5% of dropouts. The most common reason given was the *inability to immerse*, which accounted for 42% of dropouts. A sense of presence is negatively related to cybersickness [19], and feeling cybersickness has been proposed as a reason for failure to immerse [76-78]. It is possible that in addition to being a direct cause of attrition itself, cybersickness may also indirectly influence attrition by reducing immersion in and enjoyment of the VR environment. Unfortunately, our study design did not allow us to investigate how cybersickness influences attrition because we used a single session of VR exposure and the discontinuation rates were very low. Research with a wider variety of VR environment types is needed, to understand how cybersickness severity relates to study attrition, intervention adherence, and the efficacy of medical VR.

Conclusions

Here, we present data demonstrating a potentially important racial difference in cybersickness that we believe should be explored in future research. The first step in continuing this investigation would be to conceptually replicate these findings with different VR experiences, headsets, and study populations. Once replicated, research should then be conducted on the potential explanations and mechanisms. We have discussed some potential causal factors in this manuscript, but we acknowledge that there are likely other important factors that we have not considered. If future research suggests that racial differences in cybersickness are primarily in reporting as opposed to experience, it would suggest the need for the development of more objective measures of cybersickness. Such a result would also constitute an important consideration for those designing and evaluating medical VR interventions to promote health equity. Notwithstanding the caveats above, the research presented here underscores the importance of testing VR applications with a diverse group of participants to move toward achieving equitable access to emerging medical VR devices.

Acknowledgments

The authors would like to acknowledge the contributions of all the original authors of the studies reported in this analysis. This research was funded by the Intramural Research Program of the National Human Genome Research Institute.

Disclaimer

The mention of commercial products, their sources, or their use in connection with the material reported herein is not to be construed as either an actual or implied endorsement of such products by the Department of Health and Human Services.

Authors' Contributions

EB and SP conceptualized the project. SP, SHT, and APD were involved in data collection. SP and AJM conducted formal analysis of the data. All authors contributed to the drafting of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary data tables.
[DOCX File , 18 KB - jmir_v24i6e36843_app1.docx]

References

1. Nesbitt K, Nalivaiko E. Cybersickness. In: Lee N, editor. Encyclopedia of Computer Graphics and Games. Cham, Switzerland: Springer; 2018.
2. Kyaw BM, Posadzki P, Paddock S, Car J, Campbell J, Tudor Car L. Effectiveness of digital education on communication skills among medical students: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Aug 27;21(8):e12967 [FREE Full text] [doi: 10.2196/12967] [Medline: 31456579]
3. Howard MC. A meta-analysis and systematic literature review of virtual reality rehabilitation programs. Comput Human Behav 2017 May;70(C):317-327. [doi: 10.1016/j.chb.2017.01.013]

XSL•FO
RenderX

4. Hammoudeh JA, Howell LK, Boutros S, Scott MA, Urata MM. Current status of surgical planning for orthognathic surgery: traditional methods versus 3D surgical planning. Plast Reconstr Surg Glob Open 2015 Feb;3(2):e307 [FREE Full text] [doi: 10.1097/GOX.0000000000000184] [Medline: 25750846]
5. Eijlers R, Utens EM, Staals LM, de Nijs PF, Berghmans JM, Wijnen RM, et al. Systematic review and meta-analysis of virtual reality in pediatrics: effects on pain and anxiety. Anesth Analg 2019 Nov;129(5):1344-1353 [FREE Full text] [doi: 10.1213/ANE.0000000000004165] [Medline: 31136330]
6. Indovina P, Barone D, Gallo L, Chirico A, De Pietro G, Giordano A. Virtual reality as a distraction intervention to relieve pain and distress during medical procedures: a comprehensive literature review. Clin J Pain 2018 Sep;34(9):858-877. [doi: 10.1097/AJP.0000000000000599] [Medline: 29485536]
7. Kenney MP, Milling LS. The effectiveness of virtual reality distraction for reducing pain: a meta-analysis. Psychol Conscious 2016;3(3):199-210. [doi: 10.1037/cns0000084]
8. Rizzo AS, Koenig ST. Is clinical virtual reality ready for primetime? Neuropsychology 2017 Nov;31(8):877-899. [doi: 10.1037/neu0000405] [Medline: 29376669]
9. Rizzo A, Thomas Koenig S, Talbot TB. Clinical results using virtual reality. J Technol Hum Serv 2019 May 17;37(1):51-74. [doi: 10.1080/15228835.2019.1604292]
10. Mishra S, Kim YS, Intarasirisawat J, Kwon YT, Lee Y, Mahmood M, et al. Soft, wireless periocular wearable electronics for real-time detection of eye vergence in a virtual reality toward mobile eye therapies. Sci Adv 2020 Mar;6(11):eaay1729 [FREE Full text] [doi: 10.1126/sciadv.aay1729] [Medline: 32201718]
11. Halbig A, Latoschik ME. A systematic review of physiological measurements, factors, methods, and applications in virtual reality. Front Virtual Real 2021 Jul 14;2:694567. [doi: 10.3389/frvir.2021.694567]
12. Hussain R, Chessa M, Solari F. Mitigating cybersickness in virtual reality systems through foveated depth-of-field blur. Sensors (Basel) 2021 Jun 10;21(12):4006 [FREE Full text] [doi: 10.3390/s21124006] [Medline: 34200616]
13. Li G, McGill M, Brewster S, Pollick F. A review of electrostimulation-based cybersickness mitigations. In: Proceedings of the 3rd IEEE International Conference on Artificial Intelligence and Virtual Reality. 2020 Presented at: AIVR '20; December 14-18, 2020; Virtual p. 151-157. [doi: 10.1109/aivr50618.2020.00034]
14. Stanney K, Fidopiastis C, Foster L. Virtual reality is sexist: but it does not have to be. Front Robot AI 2020 Jan 31;7:4 [FREE Full text] [doi: 10.3389/frobt.2020.00004] [Medline: 33501173]
15. IEEE Standard for Head-Mounted Display (HMD)-Based Virtual Reality(VR) Sickness Reduction Technology. Institute of Electrical and Electronics Engineers Standards Association. 2020. URL: https://standards.ieee.org/standard/3079-2020.html [accessed 2022-01-18]
16. Ergonomics of human-system interaction — Part 393: Structured literature review of visually induced motion sickness during watching electronic images (ISO/TR 9241-393:2020). International Organization for Standardization. 2020. URL: https://www.iso.org/standard/73225.html [accessed 2022-01-18]
17. Medical Extended Reality Program: Research on Medical Extended Reality-Based Medical Devices. U.S. Food & Drug Administration. 2021 Mar 24. URL: https://www.fda.gov/medical-devices/medical-device-regulatory-science-research-programs-conducted-osel/medical-extended-reality-program-research-medical-extended-reality-based-medical-devices [accessed 2021-10-28]
18. Petri K, Feuerstein K, Folster S, Bariszlovich F, Witte K. Effects of age, gender, familiarity with the content, and exposure time on cybersickness in immersive head-mounted display based virtual reality. Am J Biomed Sci 2020 Apr;12(2):107-121. [doi: 10.5099/aj200200107]
19. Weech S, Kenny S, Barnett-Cowan M. Presence and cybersickness in virtual reality are negatively related: a review. Front Psychol 2019 Feb 4;10:158 [FREE Full text] [doi: 10.3389/fpsyg.2019.00158] [Medline: 30778320]
20. Arns LL, Cerney MM. The relationship between age and incidence of cybersickness among immersive environment users. In: Proceedings of 2005 IEEE Annual International Symposium Virtual Reality. 2005 Presented at: VR '05; March 12-16, 2005; Bonn, Germany p. 267-268. [doi: 10.1109/VR.2005.1492788]
21. Knight MM, Arns LL. The relationship among age and other factors on incidence of cybersickness in immersive environment users. In: Proceedings of the 3rd Symposium on Applied Perception in Graphics and Visualization. 2006 Jul Presented at: APGV '06; July 28-29, 2006; Boston, MA, USA p. 162. [doi: 10.1145/1140491.1140539]
22. Park GD, Allen RW, Fiorentino D, Rosenthal TJ, Cook ML. Simulator sickness scores according to symptom susceptibility, age, and gender for an older driver assessment study. Proc Hum Factors Ergon Soc Annu Meet 2006 Oct 1;50(26):2702-2706. [doi: 10.1177/154193120605002607]
23. Saredakis D, Szpak A, Birckhead B, Keage HA, Rizzo A, Loetscher T. Factors associated with virtual reality sickness in head-mounted displays: a systematic review and meta-analysis. Front Hum Neurosci 2020 Mar 31;14:96 [FREE Full text] [doi: 10.3389/fnhum.2020.00096] [Medline: 32300295]
24. Gonçalves G, Melo M, Bessa M. Virtual reality games: a study about the level of interaction vs. narrative and the gender in presence and cybersickness. In: Proceedings of the 2018 International Conference on Graphics and Interaction. 2018 Presented at: ICGI '18; November 15-16, 2018; Lisbon, Portugal p. 1-8. [doi: 10.1109/ITCGI.2018.8602686]

XSL•FO
RenderX

25. Jun H, Miller MR, Herrera F, Reeves B, Bailenson JN. Stimulus sampling with 360-videos: examining head movements, arousal, presence, simulator sickness, and preference on a large sample of participants and videos. IEEE Trans Affective Comput 2020 Jun 24:1-1. [doi: 10.1109/taffc.2020.3004617]
26. Shafer DM, Carbonara CP, Korpi MF. Modern virtual reality technology: cybersickness, sense of presence, and gender. Media Psychol Rev 2017;11(2):1.
27. Stanney KM, Hale KS, Nahmens I, Kennedy RS. What to expect from immersive virtual environment exposure: influences of gender, body mass index, and past experience. Hum Factors 2003;45(3):504-520. [doi: 10.1518/hfes.45.3.504.27254] [Medline: 14702999]
28. Pot-Kolder R, Veling W, Counotte J, van der Gaag M. Anxiety partially mediates cybersickness symptoms in immersive virtual reality environments. Cyberpsychol Behav Soc Netw 2018 Mar;21(3):187-193. [doi: 10.1089/cyber.2017.0082] [Medline: 29356575]
29. Rebenitsch L, Owen C. Estimating cybersickness from virtual reality applications. Virtual Reality 2020 May 28;25(1):165-174. [doi: 10.1007/s10055-020-00446-6]
30. Solimini AG, Mannocci A, Di Thiene D, La Torre G. A survey of visually induced symptoms and associated factors in spectators of three dimensional stereoscopic movies. BMC Public Health 2012 Sep 13;12:779 [FREE Full text] [doi: 10.1186/1471-2458-12-779] [Medline: 22974235]
31. Howard MC, Van Zandt EC. A meta-analysis of the virtual reality problem: unequal effects of virtual reality sickness across individual differences. Virtual Reality 2021 Dec;25(4):1221-1246. [doi: 10.1007/s10055-021-00524-3]
32. Stern RM, Hu S, LeBlanc R, Koch KL. Chinese hyper-susceptibility to vection-induced motion sickness. Aviat Space Environ Med 1993 Sep;64(9 Pt 1):827-830. [Medline: 8216144]
33. Stern RM, Hu S, Uijtdehaage SH, Muth ER, Xu LH, Koch KL. Asian hypersusceptibility to motion sickness. Hum Hered 1996;46(1):7-14. [doi: 10.1159/000154318] [Medline: 8825456]
34. Stern RM, Koch KL. Motion sickness and differential susceptibility. Curr Dir Psychol Sci 1996 Aug 1;5(4):115-120. [doi: 10.1111/1467-8721.ep11452777]
35. Muth ER, Stern RM, Uijtdehaage SH, Koch KL. Effects of Asian ancestry on susceptibility to vection-induced motion sickness. In: Chen JZ, McCallum RW, editors. Electrogastrography, Principles and Applications. New York, NY, USA: Raven Press; 1994:227-233.
36. Lasch KE. Culture, pain, and culturally sensitive pain care. Pain Manag Nurs 2000 Sep;1(3 Suppl 1):16-22. [doi: 10.1053/jpmn.2000.9761] [Medline: 11710145]
37. Klosterhalfen S, Pan F, Kellermann S, Enck P. Gender and race as determinants of nausea induced by circular vection. Gend Med 2006 Sep;3(3):236-242. [doi: 10.1016/s1550-8579(06)80211-1] [Medline: 17081956]
38. Stanney KM, Kennedy RS, Drexler JM. Cybersickness is not simulator sickness. Proc Hum Factors Ergon Soc Annu Meet 1997 Oct 1;41(2):1138-1142. [doi: 10.1177/107118139704100292]
39. Araojo R. Working to advance health equity: the U.S. Food and Drug Administration Office of Minority Health and Health Equity. J Natl Med Assoc 2021 Jun;113(3):359-362 [FREE Full text] [doi: 10.1016/j.jnma.2020.04.004] [Medline: 32402441]
40. Enhance EQUITY Initiative. U.S. Food & Drug Administration. 2021 Sep 12. URL: https://www.fda.gov/consumers/minority-health-and-health-equity/enhance-equity-initiative [accessed 2021-10-28]
41. Pulse Oximeter Accuracy and Limitations: FDA Safety Communication. U.S. Food & Drug Administration. 2021 Feb 19. URL: https://www.fda.gov/medical-devices/safety-communications/pulse-oximeter-accuracy-and-limitations-fda-safety-communication [accessed 2022-01-02]
42. Persky S, Goldring MR, Turner SA, Cohen RW, Kistler WD. Validity of assessing child feeding with virtual reality. Appetite 2018 Apr 01;123:201-207 [FREE Full text] [doi: 10.1016/j.appet.2017.12.007] [Medline: 29277518]
43. Vizard. WorldViz. 2008. URL: https://www.worldviz.com/vizard-virtual-reality-software [accessed 2022-01-15]
44. Cobb SV, Nichols S, Ramsey A, Wilson JR. Virtual reality-induced symptoms and effects (VRISE). Presence (Camb) 1999 Apr;8(2):169-186. [doi: 10.1162/105474699566152]
45. McBride CM, Persky S, Wagner LK, Faith MS, Ward DS. Effects of providing personalized feedback of child's obesity risk on mothers' food choices using a virtual reality buffet. Int J Obes (Lond) 2013 Oct;37(10):1322-1327. [doi: 10.1038/ijo.2013.87] [Medline: 23736369]
46. Persky S, Eccleston CP. Medical student bias and care recommendations for an obese versus non-obese virtual patient. Int J Obes (Lond) 2011 May;35(5):728-735 [FREE Full text] [doi: 10.1038/ijo.2010.173] [Medline: 20820169]
47. Persky S, Street Jr RL. Evaluating approaches for communication about genomic influences on body weight. Ann Behav Med 2015 Oct;49(5):675-684. [doi: 10.1007/s12160-015-9701-8] [Medline: 25894275]
48. Persky S, Ferrer RA, Klein WM. Nonverbal and paraverbal behavior in (simulated) medical visits related to genomics and weight: a role for emotion and race. J Behav Med 2016 Oct;39(5):804-814 [FREE Full text] [doi: 10.1007/s10865-016-9747-5] [Medline: 27146511]
49. Persky S, Ferrer RA, Klein WM, Goldring MR, Cohen RW, Kistler WD, et al. Effects of fruit and vegetable feeding messages on mothers and fathers: interactions between emotional state and health message framing. Ann Behav Med 2019 Aug 16;53(9):789-800 [FREE Full text] [doi: 10.1093/abm/kay088] [Medline: 30395145]

XSL•FO
RenderX

50. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int J Aviat Psychol 1993 Jul;3(3):203-220. [doi: 10.1207/s15327108ijap0303_3]
51. Rivers DC, Meade AW, Lou Fuller W. Examining question and context effects in organization survey data using item response theory. Organ Res Methods 2009;12(3):529-553. [doi: 10.1177/1094428108315864]
52. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986 Sep;7(3):177-188. [doi: 10.1016/0197-2456(86)90046-2] [Medline: 3802833]
53. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Comprehensive meta-analysis. Version 3.0. National Institutes of Health. 2006. URL: https://www.meta-analysis.com/downloads/Meta-Analysis%20Manual%20V3.pdf [accessed 2022-05-17]
54. McCauley ME, Sharkey TJ. Cybersickness: perception of self-motion in virtual environments. Presence (Camb) 1992 Aug 1;1(3):311-318. [doi: 10.1162/pres.1992.1.3.311]
55. Welch RB. Perceptual Modifications: Adapting to Altered Sensory Environments. New York, NY, USA: Academic Press; 1978.
56. Shahid Anwar M, Wang J, Ahmad S, Ullah A, Khan W, Fei Z. Evaluating the factors affecting QoE of 360-degree videos and cybersickness levels predictions in virtual reality. Electronics 2020 Sep 18;9(9):1530. [doi: 10.3390/electronics9091530]
57. Allen C. Immersion Promotion. 2021 Jan. URL: https://www.immersivepromotion.com/understanding-the-vr-market-in-2021 [accessed 2022-01-18]
58. Burch A. VR and Consumer Sentiment. Touchstone Research. 2016 Jan 28. URL: https://touchstoneresearch.com/vr-and-consumer-sentiment/ [accessed 2022-01-18]
59. Landrine H, Corral I. Advancing research on racial-ethnic health disparities: improving measurement equivalence in studies with diverse samples. Front Public Health 2014 Dec 22;2:282 [FREE Full text] [doi: 10.3389/fpubh.2014.00282] [Medline: 25566524]
60. Dodgson NA. Variation and extrema of human interpupillary distance. In: Proceedings of the International Society for Optics and Photonics. 2004 Presented at: SPIE '04; January 18-22, 2004; San Jose, CA, USA p. 36-46. [doi: 10.1117/12.529999]
61. Booker SQ. African Americans' perceptions of pain and pain management: a systematic review. J Transcult Nurs 2016 Jan;27(1):73-80. [doi: 10.1177/1043659614526250] [Medline: 24841472]
62. Edwards CL, Fillingim RB, Keefe F. Race, ethnicity and pain. Pain 2001 Nov;94(2):133-137. [doi: 10.1016/S0304-3959(01)00408-0] [Medline: 11690726]
63. Kim HJ, Yang GS, Greenspan JD, Downton KD, Griffith KA, Renn CL, et al. Racial and ethnic differences in experimental pain sensitivity: systematic review and meta-analysis. Pain 2017 Feb;158(2):194-211. [doi: 10.1097/j.pain.0000000000000731] [Medline: 27682208]
64. Meints SM, Cortes A, Morais CA, Edwards RR. Racial and ethnic differences in the experience and treatment of noncancer pain. Pain Manag 2019 May;9(3):317-334 [FREE Full text] [doi: 10.2217/pmt-2018-0030] [Medline: 31140916]
65. Sheffield D, Biles PL, Orom H, Maixner W, Sheps DS. Race and sex differences in cutaneous pain perception. Psychosom Med 2000;62(4):517-523. [doi: 10.1097/00006842-200007000-00010] [Medline: 10949097]
66. Plesh O, Adams SH, Gansky SA. Racial/Ethnic and gender prevalences in reported common pains in a national sample. J Orofac Pain 2011;25(1):25-31 [FREE Full text] [Medline: 21359234]
67. Anderson KO, Green CR, Payne R. Racial and ethnic disparities in pain: causes and consequences of unequal care. J Pain 2009 Dec;10(12):1187-1204. [doi: 10.1016/j.jpain.2009.10.002] [Medline: 19944378]
68. Lee JW, Jones PS, Mineyama Y, Zhang XE. Cultural differences in responses to a Likert scale. Res Nurs Health 2002 Aug;25(4):295-306. [doi: 10.1002/nur.10041] [Medline: 12124723]
69. Chen C, Lee SY, Stevenson HW. Response style and cross-cultural comparisons of rating scales among East Asian and North American students. Psychol Sci 1995;6(3):170-175. [doi: 10.1111/j.1467-9280.1995.tb00327.x]
70. Dudley NM, McFarland LA, Goodman SA, Hunt ST, Sydell EJ. Racial differences in socially desirable responding in selection contexts: magnitude and consequences. J Pers Assess 2005 Aug;85(1):50-64. [doi: 10.1207/s15327752jpa8501_05] [Medline: 16083384]
71. Krokos E, Varshney A. Quantifying VR cybersickness using EEG. Virtual Reality 2021 May 31;26(1):77-89. [doi: 10.1007/s10055-021-00517-2]
72. Henrich J, Heine SJ, Norenzayan A. The weirdest people in the world? Behav Brain Sci 2010 Jun;33(2-3):61-135. [doi: 10.1017/S0140525X0999152X] [Medline: 20550733]
73. Rad MS, Martingano AJ, Ginges J. Toward a psychology of Homo sapiens: making psychological science more representative of the human population. Proc Natl Acad Sci U S A 2018 Nov 06;115(45):11401-11405 [FREE Full text] [doi: 10.1073/pnas.1721165115] [Medline: 30397114]
74. Stanney K, Lawson BD, Rokers B, Dennison M, Fidopiastis C, Stoffregen T, et al. Identifying causes of and solutions for cybersickness in immersive technology: reformulation of a research and development agenda. Int J Hum Comput Interact 2020 Nov 04;36(19):1783-1803. [doi: 10.1080/10447318.2020.1828535]
75. Benbow AA, Anderson PL. A meta-analytic examination of attrition in virtual reality exposure therapy for anxiety disorders. J Anxiety Disord 2019 Jan;61:18-26. [doi: 10.1016/j.janxdis.2018.06.006] [Medline: 30646997]

XSL•FO
RenderX

76. Witmer BG, Singer MJ. Measuring presence in virtual environments: a presence questionnaire. Presence (Camb) 1998 Jun;7(3):225-240. [doi: 10.1162/105474698565686]
77. Usoh M, Catena E, Arman S, Slater M. Using presence questionnaires in reality. Presence (Camb) 2000 Oct;9(5):497-503. [doi: 10.1162/105474600566989]
78. Nichols S, Haldane C, Wilson JR. Measurement of presence and its consequences in virtual environments. Int J Hum Comput Stud 2000 Mar;52(3):471-491. [doi: 10.1006/ijhc.1999.0343]

Abbreviations

IPD: interpupillary distance
SSC: Short Symptoms Checklist
VR: virtual reality

Edited by R Kukafka; submitted 27.01.22; peer-reviewed by S Weech, B Gibson; comments to author 10.02.22; revised version received 22.03.22; accepted 20.04.22; published 01.06.22.

Please cite as:
Martingano AJ, Brown E, Telaak SH, Dolwick AP, Persky S
Cybersickness Variability by Race: Findings From 6 Studies and a Mini Meta-analysis
J Med Internet Res 2022;24(6):e36843
URL: https://www.jmir.org/2022/6/e36843
doi:10.2196/36843
PMID:35648477

©Alison Jane Martingano, Ellenor Brown, Sydney H Telaak, Alexander P Dolwick, Susan Persky. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 01.06.2022. 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

Evaluation of Web-Based Health Information From the Perspective of Women With Eating Disorders: Thematic Analysis

Hana Drtilova[1], MA; Hana Machackova[1], PhD; Martina Smahelova[1], PhD

Interdisciplinary Research Team on Internet and Society, Faculty of Social Studies, Masaryk University, Brno, Czech Republic

Corresponding Author:
Hana Drtilova, MA
Interdisciplinary Research Team on Internet and Society
Faculty of Social Studies
Masaryk University
Jostova 10
Brno, 60200
Czech Republic
Phone: 420 549 49 5870
Email: drtilova@apps.fss.muni.cz

Abstract

Background: Users with experience of eating disorders use the internet as a source of information, whether for prorecovery activities (such as web-based treatment, looking for information, support, and sharing) or activities that promote eating disorder behavior as a desirable lifestyle choice (such as pro–eating disorder communities and reading and creating pro–eating disorder posts). Their assessment of web-based eating disorder–related information is crucial for understanding the context of the illness and for health professionals and their web-based interventions.

Objective: This study aimed to understand the criteria young women with the experience of eating disorders use in evaluating eating disorder–related web-based information and what eating disorder–related characteristics of these women are involved in their evaluation.

Methods: We analyzed 30 semistructured individual interviews with Czech women aged 16 to 28 years with past or present eating disorder experience using a qualitative approach. Thematic analysis was adopted as an analytical tool.

Results: The specifics of eating disorder phases (the *disorder stage* and the *treatment process*) emerged as important aspects in the process of information assessment. Other specific characteristics of respondents (eg, motivation, abilities, and resources) addressed how the respondents arrived at certain web-based information and how they evaluated it. In addition, the respondents described some content cues as features of information (eg, novelty and social information pooling). Another finding is that other users' attitudes, experiences, activities, and personal features are involved in the information evaluation of these users and the information presented by them. Finally, the respondents evaluated the websites' visual look and graphic components.

Conclusions: This study shows that web-based information evaluation reported by women with experience of eating disorders is a complex process. The assessment is influenced by current personal characteristics related to the illness (mainly the motivation for maintaining or curing the eating disorder) using cues associated with information content, other users, and website look. The study findings have important implications for health professionals, who should ask their clients questions about web-based communities and their needs to understand what information and sources they choose.

(J Med Internet Res 2022;24(6):e31148) doi:10.2196/31148

KEYWORDS

eating disorders; web-based health information; Czech women

Introduction

Background

During the past 2 decades, the internet has been integrated into our lives as an everyday tool that opens the gate to an unlimited amount of web-based information. In this content-rich environment with almost no quality control [1], the burden of information assessment shifts toward the information seeker. The internet provides user-generated content where nonspecialists offer health tips or information based on personal experiences [2], which can be highly relevant, especially in

XSL•FO
RenderX

health-related web-based searches. Wilson et al [3] revealed that among people with eating disorders (EDs) experience, 75% used the internet as a source of information, 40.8% reported visiting prorecovery sites, and 35.5% reported visiting sites that promote ED behavior. Health providers often use the internet as a tool for prevention and as an environment for ED treatment programs [4]. Consequently, for people with ED experience, web-based information can be potentially beneficial and helpful (eg, web-based treatment [5]) but also dangerous and harmful (eg, anorexia-related misinformation on YouTube [6]). For these reasons, this study aimed to deepen the knowledge of users' web-based activities with ED experience by examining their evaluation of web-based information relevant to ED topics. Specifically, the aim was to investigate the individual characteristics that shape this process and the cues used for information assessment.

EDs and Internet Use

EDs are part of the spectrum of pathological eating patterns and are perceived as either a medical illness with psychiatric features or a psychiatric illness with medical indications [7]. Regardless of the subtypes (the most known are anorexia nervosa [AN], bulimia nervosa [BN], and binge ED), EDs are burdensome in terms of significantly impaired health-related quality of patients' lives [8], not only because of the impact that the disorder can have on all body systems. The lifetime prevalence of any threshold ED among adolescents and young adults is 2.9% among women and 0.1% among men [9]. In 2020, there were 5167 patients with EDs treated in outpatient departments and hospital admissions, and the number has increased by approximately 15% in 10 years. Most patients (87%) were women and girls [10]. However, these statistics do not include people with EDs who seek help without a psychiatric context (eg, clients of nonprofit organizations) or, for instance, people who do not seek help at all. Both male and female patients with ED have high rates of psychiatric comorbidity [11]. These disorders are challenging for caregivers within the family system, as well as for health care professionals [12,13]. Moreover, EDs are egosyntonic—the person sees the disorder as part of themselves, might attribute positive valuations to ED consequences, and may perceive ED not only as an illness but also as a meaningful behavior [14]. Recovery from an ED is a long and potentially life-threatening process [15]. Despite the severity of the illness, patients with ED symptoms do not necessarily seek treatment [16].

At different levels of illness, patients have different motivations and goals for changing their ED-connected behavior. The stages of change model [17,18] proposes 6 stages of motivation for recovery. This model is particularly relevant to understanding the phases of EDs as relapses are regarded as integral parts of the change cycle, and an application to the ED context has been examined [18]. The first 2 stages are described as precontemplation (ie, nonexisting or limited intention to change behavior) and contemplation (ie, willingness to think about the change but not commit to it). These 2 stages resemble the earliest phases of ED as bounded by limited or lacking motivation to change ED-connected behavior. Following are the stages of preparation (ie, intention to change the behavior), action (ie, actively modifying the behavior), maintenance (ie, work on the prevention of relapse), and termination (ie, zero temptation to relapse). This study presumes that people with EDs often switch between stages of illness, although the distinction between these stages might be blurred or even overlapping. Moreover, this study acknowledges that motivation plays an important role in the behavior of people with ED experience, including web-based praxes, and that motivations differ with respect to the stage of illness.

Prior research has explored the role of the internet and communication technologies in the lives of people with ED experience from various perspectives. On a general level, media are studied as one of the sociocultural risk factors that contribute to the etiology of EDs through the cultural ideals of appearance and weight [19]. Other lines of research have focused on exposure to ED-promoting websites. Such exposure is associated with a reduction in the number of calories, restriction in food consumption, greater body dissatisfaction, and a greater drive for leanness and musculature, especially among vulnerable individuals [20]. In contrast, the treatment of EDs has also moved to the web-based environment, as the anonymous internet provides a safe environment for people with EDs who may experience high levels of secrecy, shame, and stigma [21]. Internet-based treatments may offer self-assessment, self-monitoring, information about EDs, and advice for systematic treatment plans. People with ED experience also use the internet and visit web-based groups to seek information, get support, and share experiences [22].

Information Evaluation in the Web-Based Environment

Although tools for the assessment of web-based health information have been developed (eg, a study by Beaunoyer et al [23]), to the best of our knowledge, little attention has been paid to how web-based information is evaluated by users with health-related problems. This study aims to fill this gap by examining the information quality assessment by people with ED experience within their specific ED-relevant web-based activities. The construct of information quality is broad and lacks conceptual clarity. For instance, Tao et al [24] distinguished 5 dimensions of information quality on health websites: completeness, understandability, relevance, depth, and accuracy. However, other indicators of the quality of web-based information have also been identified, such as perceived aesthetics, credibility, reliability, security, consistency, usefulness, and worth [25-27]. This study defines information quality assessment as an evaluation of web-based information and materials, with a focus on both web content (including web-based information and features) and its design (the representation of the content for users) [28]. Accuracy, currency, and credibility may be among the content quality criteria, and aesthetics, cultural sensitivity, or accessibility be among the design-related criteria [29]. However, the definition of *quality* can also differ interindividually depending on the specific context and purpose of the information.

Another presumption of this study is that the web-based environment is highly diverse and can be assessed differently in terms of quality. For example, an environment (presumably) controlled by expert editors, such as websites, can be perceived

XSL•FO
RenderX

as more credible than a personal blog [30]. User characteristics, such as personal traits, abilities, previous knowledge, and topic familiarity, are also considered. The assessment process also depends on interindividual differences between users and the perceived type and context of information [31-34]. For example, the dual processing model of credibility focuses on individual factors and depicts how factors independent of message quality can affect our evaluation [35,36]. According to this model, when the user has the motivation and ability (eg, the level of literacy skills) to evaluate the quality of information, they are likely to use an analytical strategy to assess credibility systematically and rigorously. The lack of motivation inhibits users from putting effort into credibility evaluations. However, if they lack the ability and yet have motivation, they will rely more on peripheral cues (eg, the appearance of the site) and heuristics to form a judgment. Such heuristics can take the form of relying on the reputation of the source or endorsement from others [37].

User Characteristics Connected to EDs

Although research evidence about the specifics of web-based information evaluations made by people with ED experience is limited, their assessment may, in some moments, contrast that of people without ED experience. For example, their motivation for evaluation may differ according to the current state of their illness (ie, affected by the egosyntonic feature of the illness, where people with ED experience might see information congruent with their pro-ED values). Thus, to understand how people with EDs process and evaluate information, it is crucial to consider their psychological traits and cognitive characteristics. For example, a review by Cassin and von Ranson [38] revealed that both AN and BN are characterized by perfectionism, obsessive compulsiveness, neuroticism, negative emotionality, harm avoidance, low self-directedness, low cooperativeness, and traits associated with avoidant psychiatric disorders.

Moreover, some of the cognitive challenges may affect how people with the experience of EDs process web-based and offline information. Current research has demonstrated an attentional bias for disorder-salient stimuli (related to food and the body), which indicates that people with ED experience have a potential overall deficit in processing conflicting information [39]. Another cognitive deficit is a weak central coherence when attention is focused on detail, resulting in global understanding deficits [40]. Furthermore, poor set shifting (ie, a lower ability to move back and forth between tasks) results in cognitive inflexibility [41,42]. This inflexibility may manifest in rigid and concrete problem solving, reliance on strict habits and rules, and difficulties with multitasking [43].

Aims of the Study

To the best of our knowledge, no study has focused on the evaluation of web-based health information in the context of ED topics among people with an ED experience. This qualitative study intends to enhance knowledge about this topic and help us understand how women with ED experience evaluate web-based information. This study uses previous knowledge related to information evaluation. Specifically, it presumes that individual characteristics, including personality, abilities, and motivation, shape the formation of judgments about information. Moreover, information assessment can be more or less thorough. Finally, this study considers that the web-based environment provides different cues that may be used in the process. On the basis of this knowledge, the following research questions were formulated: what ED-related characteristics of young women with the experience of EDs are involved in their evaluation of web-based information, and what criteria do young women with the experience of EDs use in evaluating web-based information?

Methods

Recruitment

The data were obtained from a research project that examined the role of new technologies among young people with ED experience in the Czech Republic. Respondents were recruited via leaflets handed out at universities in large Czech cities and in the waiting rooms of health care professionals working with people with EDs (mainly in hospitals and ED-focused nonprofit organizations in the Czech Republic). Owing to the seriousness of the illness, some respondents were available only via web-based means. Thus, the outreach was gradually expanded during the sampling process, with invitations to participate on websites relevant to EDs (both supporting ED behavior and the treatment of EDs). From the previous quantitative part of the research [44], we had a list of 307 Czech websites (including blogs and Facebook groups) that focused on healthy lifestyles (including fitness and nutrition) and professional help for EDs and promoted ED behavior (mainly pro-ED blogs and groups). These websites were found via search engines using keywords connected to a healthy lifestyle (ie, exercise, diet, and healthy eating), ED problematics (ie, professional help and informational websites), and ED promotion (ie, keywords identified in previous pro-ED research, such as *pro-ana*, *thinspo*, and *bonespo*). Finally, we posted an invitation on 15 websites (including Facebook groups) that focused mainly on ED information, recovery, and ED promotion and that ranked highest in website traffic. Research has shown that EDs are most prevalent among women [9], and risk factors are present in early adolescence, although anorexia and bulimia tend to emerge in late adolescence and early adulthood. However, the onset of EDs is individual [45]. Thus, the criterion for respondents was to be aged between 13 and 28 years and have experienced (now or in the past) a form of an ED.

Sample

The final sample comprised 30 Czech women aged 16 to 28 (mean 22.4, SD 3.9) years. Although EDs are increasing among men [46], and we actively recruited respondents of all genders, the recruitment of men was not successful. All participants claimed to experience or have experienced various EDs (AN 13/30, 43%; BN, 3/30, 10%; binge ED 1/30, 3%; or multiple ED diagnoses 13/30, 43%). Some of them had reached out for the help of health professionals and had an official diagnosis (27/30, 90%), whereas others did not (3/30, 10%). Respondents reported the presence of an ED in their current life (22/30, 73%) or that they were in full recovery (8/30, 27%). Experiences with the illness varied from 1 to 16 (mean 6.3, SD 4.5) years.

XSL•FO
RenderX

Procedure

A total of 30 semistructured interviews were conducted face to face (23/30, 77%) or via Skype web-based sessions (7/30, 23%), which lasted 41 to 118 (mean 61.0, SD 21.1) minutes. The interviews focused on the use of new technologies, including questions about the role of the internet in respondents' lives; for example, "What helps you orient in health-related online information?"; "From where do you retrieve the online information?"; and "What are the most common online activities?" The interviewers also asked about the criterion for information relevance and quality and on what cues respondents adopted the information and acted on it.

All participants were informed of the ethical aspects and purpose of the research, and they provided written informed consent. In the case of respondents aged <18 years, parents provided written consent. The interviewers had psychotherapy training and at least 2 years of psychotherapy practice as a condition for reducing potential stress among respondents.

Ethics Approval

This study was approved by the Ethical Committee of the Faculty of Social Studies of the Masaryk University, Brno, Czech Republic.

Data Analyses

The thematic analysis developed by Braun and Clarke [47] was used as the analytic method and was conducted by the first author of the study (HD). The inductive approach of analysis was used when the themes were content (data) driven and emerged from the interaction between the researcher and respondents, regardless of the specific questions. Therefore, researchers could capture the complexities of meaning within a text and understand the more tacit content.

During the analysis, we were inspired by the steps in the guidelines presented by Braun and Clarke [47] and Guest et al [48]. First, researchers became familiar with the data and text segmentation by rereading the transcripts and noting their initial ideas in a logbook. Passages related to information evaluation were segmented. In creating the initial codes, the authors entered the transcripts into the qualitative analytic software NVivo (version 10) and started to generate a codebook for codes and their labels. The code represents the specific, interesting, and essential elements of the text, and it has a greater level of abstraction than the themes [47,48]. The labels of these codes were mainly in the form of in vivo phrases used by respondents and a simplified description of the code content. These codes were discussed during research team meetings to create common categories for an initial category structure. Subsequently, continuing the analysis of the second half of the transcripts, the authors merged the codes into subthemes based on their similarities. Next, revisions were made to avoid the overlapping of the meanings. According to Chang et al [49], researchers included only subthemes with ≥3 respondents to prevent fragmentation. During the phase of defining and naming the themes, the authors reread the existing codes to better understand their meaning and created corresponding labels and descriptions. Saturation was reached, with the occurrence of redundancy, after 30 interviews. The last 4 interviews confirmed saturation of the themes as they did not create new categories. Subsequently, the researchers checked whether the themes and subthemes were internally coherent and consistent. As some of the subthemes still overlapped, they merged some such subthemes. In addition, the themes were renamed to better correspond to their meanings. The final list of themes and subthemes is presented in the *Results* section.

The following steps were applied to ensure the validity of the results. First, all researchers followed an interview guide to standardize the data collection. Furthermore, remarks regarding the content of the interviews were discussed. Second, the first (HD) and third author (MS) consulted on emerging themes during the entire analytical process. Third, the study's third author (MS) conducted an audit comprising reading the related parts of the data and validating the final analyses. Fourth, the second author (HM) advised on the final presentation of the results to clarify the meaning of the themes. Examples supported the transparency of the interview results. Finally, the researchers applied the verbal labels of frequencies in the *Results* section to state how many respondents mentioned a particular subtheme. Instead of using the number of respondents, a verbal label (a pronoun connoting an indeterminate quantity) was attached to enhance the qualitative methodology. Inspired by the verbal counting of Sandelowski [50], the researchers operationally define, for example, *few* as something occurring among 3 to 8 respondents (see the *Results* section).

Results

Overview

A total of 4 themes and 10 subthemes were identified, as summarized in Table 1. The theme *respondent characteristics* represents respondents' characteristics that influenced their information evaluation. Themes of *content cues*, *characteristics of other users*, and *website cues* present the respondents' cues mentioned in their evaluation. The results cover the findings relevant to the entire information assessment process. It encompasses the initial phases of information seeking, including factors that affect preferences for diverse sources and further evaluation of the found information.

The specifics of the EDs phases emerged as important aspects of respondents' characteristics. Some respondents spontaneously categorized themselves as being in the *disorder stage* or in the *treatment process* during the interviews. Respondents who described themselves as being in the *disorder stage* had statements and descriptions that fit in the precontemplation and contemplation stages of the stages of change model [17], whereas those seeing themselves as being in the *treatment process* described more processes connected to the preparation, action, maintenance, and termination stages of the model. These dimensions were emphasized for each theme and subtheme, as shown in Table 1.

XSL•FO
RenderX

Table 1. The final list of themes and subthemes and their occurrence in ED[a] phase[b].

Theme and subtheme	ED phase (disorder stage and treatment process)	Frequency label
Respondent characteristics		
Motivation	Both	Most
Abilities and resources	Both	Few
Congruence between personal experience and information	Both	Some
Content cues		
Verification	Both	Few
Novelty	Only in the disorder stage	Few
Social information pooling	More in the disorder stage	Some
Characteristics of other users		
Source expertise	Both	Most
Similarity to respondent	Both	Most
Website cues		
Reputable look	Unclear	Few
Photographs of people relevant to ED	More in the disorder stage	Some

[a]ED: eating disorder.

[b]Frequency labels in Table 1 and further in the text describe how many respondents mentioned each subtheme. *Few* indicates 3 to 8 respondents, *some* indicates 9 to 17 respondents, *most* indicates 18 to 29 respondents, and *all* indicates 30 respondents.

Respondent Characteristics

Overview

This theme captures how the characteristics of respondents, including their web-based behavior patterns, are involved in information evaluation. Specifically, it addresses how respondents arrived at certain web-based information and how they evaluated it. The following subthemes emerged: *motivation*, *abilities and resources*, and *congruence between personal experience and information.*

Motivation

Most respondents described 2 main motivations relevant to their illness: maintaining their disorder in the disorder stage or getting cured during the treatment process. The particular type of motivation affected what information the respondent chose, as one respondent revealed the following:

> *If someone, I think, has the motivation, that she can cure herself, then she is able to filter on the internet to what she wants to read, what she doesn't want to read, and what she lets influence her and what not.* [R13]

For respondents in the disorder stage, the vision of a skinny body was so strong that they looked up and accumulated as much information as possible to maintain their goals. However, they were not concerned with its evaluation. They read and eventually "tried everything" (R2). One of the respondents piled up information instead of analyzing it, and although the information was labeled as nonsense, they read and used it anyway. They also talked about their passive role in choosing information when their disorder decided what was needed. A respondent experienced the following:

> *I know what is right, what is really right, I know it. But of course, anorexia chooses what she likes, not how it's supposed to be like. So I believe more or less in almost everything.* [R29]

In contrast, respondents in the treatment process did not read or seek proana and promia information as they were afraid of being pulled back into the disorder stage. However, they described their persistent sensitivity to disordered relevant information, such as diet commercials. Nevertheless, they used different information checks, as a respondent pointed out the following:

> *When the person is in an acute stage, then she blindly follows what she wants to gather. Now I follow what I want to gather, that I want to be healthy. But in that acute phase, I don't think about if it will hurt me. And now I want to verify all information with someone responsible, with a professional who tells me "it is appropriate for you, it is not appropriate for you."* [R6]

Abilities and Resources

For a few respondents, their current situation in terms of abilities and resources affected their information selection. For instance, respondents at the disorder stage chose exercise and weight reduction tips based on their physical state. Financials played a role as well, leading to choosing diet tips and menus suited to their monetary situation. One of the respondents in the disorder stage also mentioned that the criterion for information selection was the time spent applying particular advice to her life.

Information choice was also influenced by the information-seeking strategies and skills of the respondents.

XSL•FO
RenderX

Some actively used search engines and keywords, such as *anorexia*, *bulimia*, and *eating disorders* or questions such as "How to do away with bulimia" (R2) or "How to throw up" (R4). They then mainly clicked on the first link. Active searches also included clicking on links on blogs and forums that led to similar websites.

For others, disorder-related information had appeared unwelcomely when information was "jumping out" (R10) at them against their will. Respondents discussed how the information was "attacking them" (R27) and how it was almost impossible to avoid. For respondents in the treatment process, it was proana or promia blogs or information about eating. For respondents in the disorder stage, it was information about treatment or dieting.

Congruence Between Personal Experience and Information

The knowledge and experience gathered through their disorder helped some to assess the relevance of information about EDs. The information that corresponded to personal experiences was credible and relevant. For example, respondents in the disorder stage said that they did not trust diets as they already knew what weight loss strategies were good for them:

> *I was always using my experience. So, for example, when I knew that by that [caloric] intake I had lost weight, or by eating that food I had lost weight, then I simply ate it, because I had tested it. And I did not trust anything else.* [R8]

Respondents in the treatment process viewed healthy lifestyle information as appropriate because of their experience with professionals and with new nutrition information.

Content Cues

Overview

According to the respondents' evaluation of information, content cues described the qualities (in the sense of features) of the information found on websites. The following subthemes emerged within this theme: *verification*, *novelty*, and *social information pooling*.

Verification

Information was approved by a few respondents when it was consistent across websites.

However, some respondents needed to verify the web-based information in the offline environment by comparing the information with books or, as mentioned by respondents in the treatment process, via consultation with professionals.

Novelty

Another cue for some respondents was whether the information was new. This subtheme was mentioned more by respondents in the disorder stage:

> *How should I behave towards the food, how to hide the food, what should I avoid, how to deal with various situations, what to do and what not to do. Any new information was good information for me.* [R27]

Consequently, other users' long-term sharing of new posts was considered beneficial rather than the sharing of a few posts once in a while: information posted "every day or every other day was more credible" (R25).

Social Information Pooling

The subtheme most cited by some respondents was how others' experiences helped them judge information. This subtheme was mentioned more by respondents in the disorder stage and represented sharing opinions, recommendations, and feedback on desirable topics between respondents and other users (eg, comments below articles, reactions on forums, and liking some posts via social network sites). Respondents saw others' recommendations as helpful and worthy, although they had never met them on the web or offline:

> *It is weird that I took their advice a lot. One doesn't know the other person at all, but still follows what is written there.* [R7]

However, other respondents saw recommendations as valuable if they knew the users from the web-based environment (eg, following the blog of a friend).

In this process, the number of reactions was also significant. Testimonials, positive responses, and the number of *thumbs-ups* increased the chance of seeing information as trustful, whereas their absence had the opposite effect:

> *I read the comments if [pro-ana advice] works or not. But if there was no comment, then I did not trust it. Or I wouldn't try it unless there was something written there, some opinion.* [R7]

For respondents in the disorder stage, if some information worked for others (ie, was used and acted upon), it was good:

> *But when she reads "try this and that," then she says to herself "when others do it, it must be really cool."* [R4]

For respondents in the treatment process, social pooling was essential for the assessment of the treatment procedures and practice of helping professionals.

Characteristics of Other Web-Based Users

Overview

This theme captures how other users' attitudes, experiences, activities, and personal features are involved in the information evaluation of these users and the information presented by them. Other users were mostly seen as post contributors or members of web-based communities. Their characteristics were expressed in the subthemes of *source expertise* and *similarity to the respondent*.

Source Expertise

The perception of who is an expert in an ED field served as a hint for information assessment. Respondents distinguished whether the information was provided by users who are currently experiencing or had experienced EDs or by ED specialists. The specialists were mentioned without further explanation or specified as professionals, such as psychologists, physicians,

XSL•FO
RenderX

and nutrition specialists, and were connected to institutions, such as universities, hospitals, and ED treatment centers.

The most mentioned aspect for most respondents was whether users had experience with EDs. Those who did were seen as experts in ED problems and credible as they better understood the respondents' issues and feelings. Respondents who preferred this expertise sometimes set experience with EDs as a cornerstone:

> *Certainly, there were some important basic factors. That person had to have an eating disorder, or at least not be OK with food, such as people with obesity.* [R24]

Specifically, one of the respondents in the disorder stage said that she ignored notes by professionals on a self-help website and read only the text passages written by people with EDs.

In both stages, the mentoring and labeling of respondents by professionals discouraged respondents from looking up and reading the information presented by these professionals. Professionals could not be trustworthy for a respondent in the disorder stage as they wanted her to gain weight and, therefore, did not provide complete information. In contrast, the entire community of users with ED experience was considered trustworthy. The goodwill of this community was described as being open, welcoming, and accepting, creating the feeling of an alliance.

The valuable characteristics of ED specialists on the internet were described simply as being professional, providing verified information to respondents, understanding EDs, having healthy opinions about EDs, and being selflessly helping respondents. Those who saw professionals as credible (although some questioned their expertise) perceived the ED information written by experts and treatment groups supervised by them as trustworthy.

For some respondents, the view of the expertise cue had changed over time. Those who started treatment looked more or only for information written and shared by professionals. Respondents uncertain about their willingness to be cured could have ambivalent feelings, as one respondent pointed out the following:

> *If I listened to the doctor, I would eat well. I know she gives me the right advice, that she is a professional. However, unfortunately, I adopt what suits me more. Rather from the Internet.* [R29]

Similarity to Respondent

Most respondents wanted to know many details about other users to better evaluate whether they were similar and, consequently, whether they were trustworthy, which in some respondents led to trying to find as many details as possible.

Respondents specifically mentioned similarity cues connected to their illness, including the same type of ED, a similar stage of the disorder, similar problems, and similar attitudes toward food. The cues not strictly linked to ED were described as similarities in current mood, humor, age, and writing style.

Website Cues

Overview

This theme reveals a representation of the visual look and graphic components of the website. Two subthemes were identified: *reputable look* and *photographs of people relevant to ED*.

Reputable Look

Respondents named a reputable look as a good sign for further selection. This look was not defined by the specific features of a website but rather by broad general characteristics, such as clarity, lack of bias, and good organization.

Photographs of People Relevant to ED

Pictures, videos, advertisement photos of diet products, and especially photographs of people with experience of EDs were important for respondents, mostly in the disorder stage. Before and after images helped respondents assess the actual effect of others' aims and provided tips to lose or gain weight. Moreover, the photographs acted as proof that the people in them were real and not lying. Specifically, respondents in the disorder stage believed in what was presented in the photographs:

> *I saw a picture of a woman with a gorgeous figure, so I believed that she is on that diet [from a commercial].* [R23]

A few respondents mentioned that they no longer took advertisement information for granted when they shifted to the treatment stage of ED.

Discussion

Principal Findings

The purpose of this study was to explore how young women with ED evaluate web-based information and how the specifics of their illness contribute to this evaluative process. Respondents mentioned several cues for information assessment within the websites' content, the characteristics of other users, and the website characteristics while stressing the influence of the current phase of their illness.

First, respondents' characteristics played an integral role in the evaluation, intervening in the entire process from exposure to information to final judgment. As Hargittai et al [32] suggested, obtaining information and its evaluation are more often handled as 2 separate research interests. However, our respondents holistically depicted these 2 steps. They explained how their *abilities* and *resources* affected this process and, importantly, what role *motivation* played in grounding their illness. *Motivation* turned out to be a vital part of the assessment, confirming the justified emphasis within the dual processing model of credibility by Metzger [36] and the stage of change model [17]. For respondents in the disorder stage, the motivation to lose weight was their biggest goal. They were not concerned about information evaluation and saw every piece of information as automatically good. This strategy could be a consequence of their *congruence between personal experience and information* (ie, EDs): the more experienced the respondents were, the more certain they were about the accuracy of their information

XSL•FO
RenderX

selection. Another explanation for the automatic assessment of information may be cognitive rigidity, which is more pronounced in people with ED experience (eg, see the study by Tchanturia and Hambrook [43]). As a result, when a particular type of information is already assessed as quality, additional information from a similar topic can also be considered appropriate without further evaluation. Although such quick evaluation strategies are convenient, they may generate bias. A study by Guardiola-Wanden-Berghe et al [51] suggested that the information quality of websites on dieting and EDs was poor. Viewing them as automatically believable may lead to the risk of adopting and behaving on harmful information.

The social element attached to quality cues was also prevalent throughout most themes. The *characteristics of others* was the most frequently mentioned theme, although web-based sources and other users are often masked or missing for evaluation [35]. However, reliance on social features is part of the evaluative process, as shown in other studies [37]. Similarly, *social information pooling* was a social-connected cue whereby other users' opinions and feedback were hints for information selection and evaluation. In particular, for respondents in the disorder stage, the information confirmed and approved by other users was seen as trustworthy. Weight reduction was the primary goal for respondents in the ED stage; therefore, they looked for fast and effective weight reduction information. However, this search could be limited in 2 aspects: as the subtheme *abilities and resources* revealed, they were tired and probably less inclined to put much energy into it; in addition, some of them did not perceive health professionals as credible. Thus, the opinions and reviews of others served as quick and less effortful tools for information evaluation. The more those testimonials were presented, the more chance the information was viewed as valid. These assumptions are similar to the bandwagon heuristic of Sundar [52], whereby people suppose that if many others think something is correct, then it must be. However, heuristics may lead to problems with crowd behavior, especially for young people with EDs, who are at greater risk of peer pressure [53]. Moreover, quality may be falsely connected to popularity when unpopular topics and information are discarded [2]. In addition, popular users may be perceived as falsely trustworthy. Thus, when and from whom people with EDs seek support and advice might be crucial for understanding their information evaluation and selection.

User ED expertise was another socially relevant cue in which information from others who had experience with EDs was perceived as appropriate. Boero and Pascoe [54] suggested that other users in pro-ED communities use their experiences to demonstrate their authenticity and offer advice about, for example, how to get through recovery programs without actual recovery. They also answered the questions of new members using their nutritional and medical knowledge. The respondents' positive perception of the information presented by these users may be partly understood by the Situated Identity Enactment Model of Cruwys et al [55], highlighting the role of socially bonded norms, identity, and context. Being a member of a social group may moderate conformity with the group's social norms. For example, these norms may be to follow a lifestyle presented by the group or to trust more experienced members. Being a member of an ED-experienced group is also vital in the treatment stage because of the transformation of illness identity to treatment identity [56]. If a member fails to follow the norms, they may be rejected [57]. Thus, one of the possible group norms may be to see ED-experienced group members as reliable, and respondents comforted by this norm seek to avoid group rejection. Studies suggest that communities of ED-experienced users represent a safe place for sharing experiences and a means of support and understanding [58]. The respondents in this study indeed saw their ED-related web-based communities as full of goodwill and, thus, were trustworthy.

The similar stage of an ED or the same ED type was another marker for user quality, reflecting studies showing that people seek and see information that matches what they already know [59,60]. Accordingly, respondents stated that users who were similar to respondents and supported respondents' opinions might have similar opinions and are plausible. Metzger and Flanagin [2] described a self-confirmation heuristic whereby credible information confirms pre-existing beliefs. In the case of an ED, this heuristic may enhance the egosyntonic feature of the illness, whereby respondents value their ED and see the information as good if it supports its maintenance. This confirmation of established opinions may place a burden on treatment, especially at the disorder stage. In the disorder stage, respondents ignored health professional advice and followed prodisorder tips, which enabled them to pursue an ED lifestyle. In contrast, similar experiences of recovery and treatment strategies may be supportive for those who decide to treat themselves. Consequently, health professionals should be aware of the identification needs of their clients and how their selectiveness may influence what is considered credible.

Finally, respondents also mentioned cues related to websites (specifically the reputable look and photographs associated with EDs); however, mentions of such elements were relatively scarce. This could suggest that, within the self-report context of the testimonies, respondents consciously reflected that website cues were not as significant as to be named and instead focused on social aspects in the overall evaluation.

Limitations of the Study

This study had several limitations. First, the sample was relatively homogenous as it comprised only women, although they differed in age, region, and illness experience.

Thus, the testimonies might not represent other demographics, such as men or different ethnicities. Next, the respondents reported the condition of the ED diagnosis, and not all of them had experience with professional health care. It is possible that although respondents claimed to have an ED themselves, they would not meet the diagnostic criteria of psychological diagnostic manuals, such as the Diagnostic and Statistical Manual of Mental Disorders. Similarly, experience with technology was not measured and might have varied for different illnesses and ages. In addition, the 2 stages of EDs (disorder vs treatment process) derived from the respondents were often blurred, unclear, sometimes overlapping, or even missing. Although some processes of these stages might resemble the stage of change model [17], respondents were not further asked for their motivation to change. Finally, respondents' browsing

XSL•FO
RenderX

history was self-reported, and researchers did not retrieve the browsing history data. Thus, respondents' internet use and information were unique to each and based and customized on their previous web-based activities, which can vary substantially among individuals. For future studies, we suggest using more accurate measures to use the level of change and consider different contexts and settings for respondents' information evaluation.

Conclusions and Implications

This study showed that the web-based information evaluation reported by women with the experience of EDs is a complex process. The assessment is influenced by current personal characteristics related to the illness (mainly the motivation to maintain or cure the ED) using cues associated with information content, other users, and website look. The study findings have important implications for health professionals, who should ask their clients questions about web-based communities and their needs to understand what information and sources they choose. They should support clients by consulting their judgments and uncertainties about information evaluation. Further investigation into the role of consulting web-based information with professionals in information assessment by users might benefit future therapeutic practices. Who and what is positive regarding quality and how it changes within the ED phases may help to understand the illness. Moreover, quality cues may serve as merits for designing an ideal website for health providers who use the internet for prevention and intervention. Future research might benefit from the experimental design of these websites and their evaluation by people with ED experience. For example, the personal stories of people experiencing EDs, their tips about treatment, and web-based peer groups on websites may increase the perceived quality of information and providers. However, the perceived quality of information might not be associated with the direct application and use of this information in an offline environment. Thus, the dynamism of the transfer between perceptions of information quality and acting on this information might be another research interest for the future.

Acknowledgments

This research was funded by Czech Science Foundation, grant number 19-27828X (project FUTURE).

Conflicts of Interest

None declared.

References

1. Rieh SY, Danielson DR. Credibility: a multidisciplinary framework. Annual Rev Inf Sci Technol 2008 Oct 24;41(1):307-364. [doi: 10.1002/aris.2007.1440410114]
2. Metzger MJ, Flanagin AJ. Credibility and trust of information in online environments: the use of cognitive heuristics. J Pragmatics 2013 Dec;59:210-220. [doi: 10.1016/j.pragma.2013.07.012]
3. Wilson JL, Peebles R, Hardy KK, Litt IF. Surfing for thinness: a pilot study of pro-eating disorder web site usage in adolescents with eating disorders. Pediatrics 2006 Dec;118(6):e1635-e1643. [doi: 10.1542/peds.2006-1133] [Medline: 17142493]
4. Schlegl S, Bürger C, Schmidt L, Herbst N, Voderholzer U. The potential of technology-based psychological interventions for anorexia and bulimia nervosa: a systematic review and recommendations for future research. J Med Internet Res 2015 Mar 31;17(3):e85 [FREE Full text] [doi: 10.2196/jmir.3554] [Medline: 25840591]
5. ter Huurne ED, de Haan HA, Postel MG, van der Palen J, VanDerNagel JE, DeJong CA. Web-based cognitive behavioral therapy for female patients with eating disorders: randomized controlled trial. J Med Internet Res 2015 Jun 18;17(6):e152 [FREE Full text] [doi: 10.2196/jmir.3946] [Medline: 26088580]
6. Syed-Abdul S, Fernandez-Luque L, Jian W, Li Y, Crain S, Hsu M, et al. Misleading health-related information promoted through video-based social media: anorexia on YouTube. J Med Internet Res 2013 Feb 13;15(2):e30 [FREE Full text] [doi: 10.2196/jmir.2237] [Medline: 23406655]
7. Maine M, McGilley BH, Bunnell DW. Treatment of Eating Disorders Bridging the Research - Practice Gap Book, 1st edition. Amsterdam, The Netherlands: Elsevier; 2010.
8. Ágh T, Kovács G, Supina D, Pawaskar M, Herman BK, Vokó Z, et al. A systematic review of the health-related quality of life and economic burdens of anorexia nervosa, bulimia nervosa, and binge eating disorder. Eat Weight Disord 2016 Sep;21(3):353-364 [FREE Full text] [doi: 10.1007/s40519-016-0264-x] [Medline: 26942768]
9. Nagl M, Jacobi C, Paul M, Beesdo-Baum K, Höfler M, Lieb R, et al. Prevalence, incidence, and natural course of anorexia and bulimia nervosa among adolescents and young adults. Eur Child Adolesc Psychiatry 2016 Aug;25(8):903-918. [doi: 10.1007/s00787-015-0808-z] [Medline: 26754944]
10. The number of people with eating disorders is increasing in the Czech Republic. NewsBeezer. URL: https://czechpoints.com/eating-disorders-are-on-the-rise-in-the-czech-republic/ [accessed 2022-04-20]
11. Ulfvebrand S, Birgegård A, Norring C, Högdahl L, von Hausswolff-Juhlin Y. Psychiatric comorbidity in women and men with eating disorders results from a large clinical database. Psychiatry Res 2015 Dec 15;230(2):294-299. [doi: 10.1016/j.psychres.2015.09.008] [Medline: 26416590]

XSL•FO
RenderX

12. Seah XY, Tham XC, Kamaruzaman NR, Yobas PK. Knowledge, attitudes and challenges of healthcare professionals managing people with eating disorders: a literature review. Arch Psychiatr Nurs 2017 Feb;31(1):125-136. [doi: 10.1016/j.apnu.2016.09.002] [Medline: 28104050]
13. Stefanini MC, Troiani MR, Caselli M, Dirindelli P, Lucarelli S, Caini S, et al. Living with someone with an eating disorder: factors affecting the caregivers' burden. Eat Weight Disord 2019 Dec;24(6):1209-1214. [doi: 10.1007/s40519-018-0480-7] [Medline: 29368292]
14. Clinical presentation of anorexia nervosabulimia nervosa. In: Eating Disorders and Obesity, Second Edition A Comprehensive Handbook. New York, USA: The Guilford Press; 2002.
15. Rome E, Ammerman S, Rosen D, Keller R, Lock J, Mammel K, et al. Children and adolescents with eating disorders: the state of the art. Pediatrics 2003 Jan;111(1):e98-108. [doi: 10.1542/peds.111.1.e98] [Medline: 12509603]
16. Hart LM, Granillo MT, Jorm AF, Paxton SJ. Unmet need for treatment in the eating disorders: a systematic review of eating disorder specific treatment seeking among community cases. Clin Psychol Rev 2011 Jul;31(5):727-735. [doi: 10.1016/j.cpr.2011.03.004] [Medline: 21501580]
17. DiClemente C, Prochaska J. Toward a comprehensive model of change. In: Treating Addictive Behaviors. Boston, MA, USA: Springer; 1986.
18. Hasler G, Delsignore A, Milos G, Buddeberg C, Schnyder U. Application of Prochaska's transtheoretical model of change to patients with eating disorders. J Psychosom Res 2004 Jul;57(1):67-72. [doi: 10.1016/S0022-3999(03)00562-2] [Medline: 15256297]
19. Keel PK, Forney KJ. Psychosocial risk factors for eating disorders. Int J Eat Disord 2013 Jul;46(5):433-439. [doi: 10.1002/eat.22094] [Medline: 23658086]
20. Rodgers RF, Lowy AS, Halperin DM, Franko DL. A meta-analysis examining the influence of pro-eating disorder websites on body image and eating pathology. Eur Eat Disord Rev 2016 Jan;24(1):3-8. [doi: 10.1002/erv.2390] [Medline: 26230192]
21. Murphy R, Frost S, Webster P, Schmidt U. An evaluation of web-based information. Int J Eat Disord 2004 Mar;35(2):145-154. [doi: 10.1002/eat.10245] [Medline: 14994351]
22. McCormack A. Individuals with eating disorders and the use of online support groups as a form of social support. Comput Inform Nurs 2010;28(1):12-19. [doi: 10.1097/NCN.0b013e3181c04b06] [Medline: 19940616]
23. Beaunoyer E, Arsenault M, Lomanowska AM, Guitton MJ. Understanding online health information: evaluation, tools, and strategies. Patient Educ Couns 2017 Feb;100(2):183-189. [doi: 10.1016/j.pec.2016.08.028] [Medline: 27595436]
24. Tao D, LeRouge C, Smith KJ, De Leo G. Defining information quality into health websites: a conceptual framework of health website information quality for educated young adults. JMIR Hum Factors 2017 Oct 06;4(4):e25 [FREE Full text] [doi: 10.2196/humanfactors.6455] [Medline: 28986336]
25. Kim H, Park SY, Bozeman I. Online health information search and evaluation: observations and semi-structured interviews with college students and maternal health experts. Health Info Libr J 2011 Sep;28(3):188-199 [FREE Full text] [doi: 10.1111/j.1471-1842.2011.00948.x] [Medline: 21831218]
26. 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]
27. Knight S, Burn J. Developing a framework for assessing information quality on the world wide web. Informing Sci J 2005;8:159-172. [doi: 10.28945/493]
28. Huizingh E. The content and design of web sites: an empirical study. Inf Manag 2000 Apr;37(3):123-134 [FREE Full text] [doi: 10.1016/s0378-7206(99)00044-0]
29. Zhang Y, Sun Y, Xie B. Quality of health information for consumers on the web: a systematic review of indicators, criteria, tools, and evaluation results. J Assn Inf Sci Tec 2015 Apr 29;66(10):2071-2084. [doi: 10.1002/asi.23311]
30. Hu Y, Shyam Sundar S. Effects of online health sources on credibility and behavioral intentions. Commun Res 2009 Nov 25;37(1):105-132. [doi: 10.1177/0093650209351512]
31. Fogg BJ. Prominence-interpretation theory: explaining how people assess credibility online. In: Proceedings of the Extended Abstracts on Human Factors in Computing Systems. 2003 Presented at: CHI03: Human Factors in Computing Systems; Apr 5 - 10, 2003; Ft. Lauderdale Florida USA. [doi: 10.1145/765891.765951]
32. Hargittai E, Fullerton L, Menchen-Trevino E, Thomas KY. Trust online: young adults' evaluation of web content. Int J Commun 2010;4(1):468-494 [FREE Full text]
33. Hilligoss B, Rieh SY. Developing a unifying framework of credibility assessment: construct, heuristics, and interaction in context. Inf Process Manag 2008 Jul;44(4):1467-1484. [doi: 10.1016/j.ipm.2007.10.001]
34. Rafalak M, Abramczuk K, Wierzbicki A. Incredible: is (almost) all web content trustworthy? analysis of psychological factors related to website credibility evaluation. In: Proceedings of the 23rd International Conference on World Wide Web. 2014 Presented at: WWW '14: 23rd International World Wide Web Conference; Apr 7-11, 2014; Seoul, Korea. [doi: 10.1145/2567948.2578997]
35. Digital Media, Youth, and Credibility. Cambridge, MA, USA: MIT Press; 2008.
36. Metzger MJ. Making sense of credibility on the web: models for evaluating online information and recommendations for future research. J Am Soc Inf Sci 2007 Nov;58(13):2078-2091. [doi: 10.1002/asi.20672]

XSL•FO
RenderX

37. Metzger MJ, Flanagin AJ, Medders RB. Social and heuristic approaches to credibility evaluation online. J Commun 2010 Aug 19;60(3):413-439. [doi: 10.1111/j.1460-2466.2010.01488.x]
38. Cassin SE, von Ranson KM. Personality and eating disorders: a decade in review. Clin Psychol Rev 2005 Nov;25(7):895-916. [doi: 10.1016/j.cpr.2005.04.012] [Medline: 16099563]
39. Aspen V, Darcy AM, Lock J. A review of attention biases in women with eating disorders. Cogn Emot 2013;27(5):820-838 [FREE Full text] [doi: 10.1080/02699931.2012.749777] [Medline: 23228135]
40. Lang K, Roberts M, Harrison A, Lopez C, Goddard E, Khondoker M, et al. Central coherence in eating disorders: a synthesis of studies using the Rey Osterrieth complex figure test. PLoS One 2016;11(11):e0165467 [FREE Full text] [doi: 10.1371/journal.pone.0165467] [Medline: 27806073]
41. Wu M, Brockmeyer T, Hartmann M, Skunde M, Herzog W, Friederich H. Set-shifting ability across the spectrum of eating disorders and in overweight and obesity: a systematic review and meta-analysis. Psychol Med 2014 Dec;44(16):3365-3385. [doi: 10.1017/S0033291714000294] [Medline: 25066267]
42. Kanakam N, Raoult C, Collier D, Treasure J. Set shifting and central coherence as neurocognitive endophenotypes in eating disorders: a preliminary investigation in twins. World J Biol Psychiatry 2013 Aug;14(6):464-475. [doi: 10.3109/15622975.2012.665478] [Medline: 22630167]
43. Tchanturia K, Hambrook D. Cognitive remediation therapy for anorexia nervosa. In: The Treatment of Eating Disorders: A Clinical Handbook. New York, USA: The Guilford Press; 2011.
44. Almenara CA, Machackova H, Smahel D. Sociodemographic, attitudinal, and behavioral correlates of using nutrition, weight loss, and fitness websites: an online survey. J Med Internet Res 2019 Apr 04;21(4):e10189 [FREE Full text] [doi: 10.2196/10189] [Medline: 30946018]
45. Rohde P, Stice E, Marti CN. Development and predictive effects of eating disorder risk factors during adolescence: implications for prevention efforts. Int J Eat Disord 2015 Mar 06;48(2):187-198 [FREE Full text] [doi: 10.1002/eat.22270] [Medline: 24599841]
46. Mitchison D, Hay P, Slewa-Younan S, Mond J. The changing demographic profile of eating disorder behaviors in the community. BMC Public Health 2014 Sep 11;14:943 [FREE Full text] [doi: 10.1186/1471-2458-14-943] [Medline: 25213544]
47. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
48. Guest G, MacQueen KM, Namey E. Applied Thematic Analysis. Los Angeles, CA, USA: SAGE Publications, Inc; 2014.
49. Chang Y, Voils CI, Sandelowski M, Hasselblad V, Crandell JL. Transforming verbal counts in reports of qualitative descriptive studies into numbers. West J Nurs Res 2009 Nov;31(7):837-852 [FREE Full text] [doi: 10.1177/0193945909334434] [Medline: 19448052]
50. Sandelowski M. Real qualitative researchers do not count: the use of numbers in qualitative research. Res Nurs Health 2001 Jun;24(3):230-240. [doi: 10.1002/nur.1025] [Medline: 11526621]
51. Guardiola-Wanden-Berghe R, Gil-Pérez JD, Sanz-Valero J, Wanden-Berghe C. Evaluating the quality of websites relating to diet and eating disorders. Health Info Libr J 2011 Dec;28(4):294-301 [FREE Full text] [doi: 10.1111/j.1471-1842.2011.00961.x] [Medline: 22051128]
52. Sundar SS. The MAIN model: a heuristic approach to understanding technology effects on credibility. In: Digital Media, Youth, and Credibility. Cambridge, MA, USA: The MIT Press; 2008.
53. Ferguson CJ, Muñoz ME, Garza A, Galindo M. Concurrent and prospective analyses of peer, television and social media influences on body dissatisfaction, eating disorder symptoms and life satisfaction in adolescent girls. J Youth Adolesc 2014 Jan;43(1):1-14. [doi: 10.1007/s10964-012-9898-9] [Medline: 23344652]
54. Boero N, Pascoe C. Pro-anorexia communities and online interaction: bringing the pro-ANA body online. Body Soc 2012 May 24;18(2):27-57. [doi: 10.1177/1357034X12440827]
55. Cruwys T, Platow M, Rieger E, Byrne DG, Haslam SA. The social psychology of disordered eating: the situated identity enactment model. Eur Rev Soc Psychol 2016 Oct 24;27(1):160-195. [doi: 10.1080/10463283.2016.1229891]
56. McNamara N, Parsons H. 'Everyone here wants everyone else to get better': the role of social identity in eating disorder recovery. Br J Soc Psychol 2016 Dec;55(4):662-680. [doi: 10.1111/bjso.12161] [Medline: 27667140]
57. Giles D. Constructing identities in cyberspace: the case of eating disorders. Br J Soc Psychol 2006 Sep;45(Pt 3):463-477. [doi: 10.1348/014466605X53596] [Medline: 16984715]
58. Yeshua-Katz D, Martins N. Communicating stigma: the pro-ana paradox. Health Commun 2013;28(5):499-508. [doi: 10.1080/10410236.2012.699889] [Medline: 22873763]
59. Morahan-Martin JM. How internet users find, evaluate, and use online health information: a cross-cultural review. Cyberpsychol Behav 2004 Oct;7(5):497-510. [doi: 10.1089/cpb.2004.7.497] [Medline: 15667044]
60. Senkowski V, Branscum P. How college students search the internet for weight control and weight management information: an observational study. Am J Health Educ 2015 Jul 06;46(4):231-240. [doi: 10.1080/19325037.2015.1044139]

XSL•FO
RenderX

Abbreviations

AN: anorexia nervosa
BN: bulimia nervosa
ED: eating disorder

Edited by G Eysenbach; submitted 11.06.21; peer-reviewed by G De Leo, Q Wu; comments to author 17.11.21; revised version received 11.01.22; accepted 27.04.22; published 13.06.22.

Please cite as:
Drtilova H, Machackova H, Smahelova M
Evaluation of Web-Based Health Information From the Perspective of Women With Eating Disorders: Thematic Analysis
J Med Internet Res 2022;24(6):e31148
URL: https://www.jmir.org/2022/6/e31148
doi:10.2196/31148
PMID:35699984

©Hana Drtilova, Hana Machackova, Martina Smahelova. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.06.2022. 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

Deep Phenotyping of Chinese Electronic Health Records by Recognizing Linguistic Patterns of Phenotypic Narratives With a Sequence Motif Discovery Tool: Algorithm Development and Validation

Shicheng Li[1,2*], MM; Lizong Deng[1,2*], PhD; Xu Zhang[1,2*], MM; Luming Chen[1,2,3], MD; Tao Yang[3,4], MM; Yifan Qi[1,2], MM; Taijiao Jiang[1,2,3], PhD

[1]Institute of Systems Medicine, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[2]Suzhou Institute of Systems Medicine, Suzhou, China

[3]Guangzhou Laboratory, Guangzhou, China

[4]Guangzhou Medical University, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Taijiao Jiang, PhD
Institute of Systems Medicine
Chinese Academy of Medical Sciences & Peking Union Medical College
No.100 Chongwen Road, Suzhou, Jiangsu Province, China
Beijing, 215000
China
Phone: 86 13366191184
Email: taijiaobioinfor@ism.cams.cn

Abstract

Background: Phenotype information in electronic health records (EHRs) is mainly recorded in unstructured free text, which cannot be directly used for clinical research. EHR-based deep-phenotyping methods can structure phenotype information in EHRs with high fidelity, making it the focus of medical informatics. However, developing a deep-phenotyping method for non-English EHRs (ie, Chinese EHRs) is challenging. Although numerous EHR resources exist in China, fine-grained annotation data that are suitable for developing deep-phenotyping methods are limited. It is challenging to develop a deep-phenotyping method for Chinese EHRs in such a low-resource scenario.

Objective: In this study, we aimed to develop a deep-phenotyping method with good generalization ability for Chinese EHRs based on limited fine-grained annotation data.

Methods: The core of the methodology was to identify linguistic patterns of phenotype descriptions in Chinese EHRs with a sequence motif discovery tool and perform deep phenotyping of Chinese EHRs by recognizing linguistic patterns in free text. Specifically, 1000 Chinese EHRs were manually annotated based on a fine-grained information model, PhenoSSU (Semantic Structured Unit of Phenotypes). The annotation data set was randomly divided into a training set (n=700, 70%) and a testing set (n=300, 30%). The process for mining linguistic patterns was divided into three steps. First, free text in the training set was encoded as single-letter sequences (P: phenotype, A: attribute). Second, a biological sequence analysis tool—MEME (Multiple Expectation Maximums for Motif Elicitation)—was used to identify motifs in the single-letter sequences. Finally, the identified motifs were reduced to a series of regular expressions representing linguistic patterns of PhenoSSU instances in Chinese EHRs. Based on the discovered linguistic patterns, we developed a deep-phenotyping method for Chinese EHRs, including a deep learning–based method for named entity recognition and a pattern recognition–based method for attribute prediction.

Results: In total, 51 sequence motifs with statistical significance were mined from 700 Chinese EHRs in the training set and were combined into six regular expressions. It was found that these six regular expressions could be learned from a mean of 134 (SD 9.7) annotated EHRs in the training set. The deep-phenotyping algorithm for Chinese EHRs could recognize PhenoSSU instances with an overall accuracy of 0.844 on the test set. For the subtask of entity recognition, the algorithm achieved an F1 score of 0.898 with the Bidirectional Encoder Representations from Transformers–bidirectional long short-term memory and

XSL•FO
RenderX

conditional random field model; for the subtask of attribute prediction, the algorithm achieved a weighted accuracy of 0.940 with the linguistic pattern–based method.

Conclusions: We developed a simple but effective strategy to perform deep phenotyping of Chinese EHRs with limited fine-grained annotation data. Our work will promote the second use of Chinese EHRs and give inspiration to other non–English-speaking countries.

(J Med Internet Res 2022;24(6):e37213) doi:10.2196/37213

KEYWORDS

deep phenotyping; Chinese EHRs; linguistic pattern; motif discovery; pattern recognition

Introduction

Currently, electronic health records (EHRs) are increasingly becoming an important source for clinical data mining and analysis [1]. Phenotype information that describes patients' clinical manifestations is one of the most valuable clinical information types in EHRs [2]. However, phenotype information in EHRs is mainly recorded in free text, which computers have difficulty using directly [3,4]. Therefore, it is important to develop natural language processing (NLP) technology to effectively structure phenotype information in free text. The NLP technology for structuring phenotype information in EHRs is called EHR-based phenotyping [5].

There are two key factors involved in EHR-based phenotyping [6]. The first factor is the development of an information model that can define the normalized target of phenotyping [7]. The second factor is the development of a phenotyping algorithm that can process phenotype information into a predefined information model [8]. In recent years, the focus of EHR-based phenotyping methods has shifted from the coarse-grained level to the fine-grained level [9,10]. Compared with coarse-grained phenotyping, fine-grained phenotyping can capture more phenotype details, including the phenotype concept and its associated attributes [11]. For example, in the free-text description "a sudden severe pain in the right-lower abdomen," a fine-grained deep-phenotyping method not only considers the phenotype "pain" but also its associated attributes of body location ("abdomen"), temporal pattern ("acute"), and severity ("severe"). EHR-based phenotyping that can characterize phenotype details at a fine-grained level is called EHR-based deep phenotyping [12].

Deep-phenotyping methods can characterize phenotype information in a high-fidelity way, which can potentially improve the accuracy of EHR-based applications, such as disease diagnosis and treatment [13]. Hence, deep phenotyping has become the focus of medical informatics. In recent years, a series of deep-phenotyping methods for English EHRs have been developed. For example, Peterson et al [14] used the MetaMap tool [15] to recognize phenotype concepts in EHRs, along with a neural network model to predict attribute values associated with phenotypes. They finally characterized English EHRs with the Fast Healthcare Interoperability Resources (FHIR) model [16]. Xu et al [17] developed a bidirectional long short-term memory and conditional random field (Bi-LSTM-CRF) model to recognize phenotype concepts in EHRs, together with a machine learning method to predict attribute values, and finally represented the phenotype information in English EHRs with the clinical element model (CEM) [18]. Compared to the progress of deep-phenotyping English EHRs, the method for deep-phenotyping Chinese EHRs is still in its infancy. Regarding the existence of linguistic differences, the established strategies [14,17,19,20] for deep-phenotyping English EHRs cannot be directly used for Chinese EHRs. Moreover, developing a deep-phenotyping algorithm requires fine-grained annotation data. However, it is hard to obtain a large volume of annotation data because of the high annotation cost. This means that the development of a deep-phenotyping algorithm for Chinese EHRs suffers from the challenge of low-resource scenarios [8], so it is worth considering how to develop a generalized algorithm for deep-phenotyping Chinese EHRs with limited fine-grained annotated data.

In previous work, we developed a fine-grained information model named PhenoSSU (Semantic Structured Unit of Phenotypes) [21], which can accurately characterize phenotype information from medical guidelines with 12 attributes from SNOMED CT (Systematized Nomenclature of Medicine–Clinical Terms). To explore an effective strategy for deep-phenotyping Chinese EHRs, we tried to annotate some Chinese EHRs with the PhenoSSU model. During the annotation process, some linguistic patterns of PhenoSSU instances were found to frequently occur in the free text of Chinese EHRs. For example, there is a linguistic pattern of "attribute + attribute + attribute + phenotype" in a given Chinese sentence "患者反复出现(attribute)剧烈(attribute)腹部(attribute)疼痛(phenotype)" (English translation: "patients with repeated severe abdominal pain"). If the linguistic patterns of PhenoSSU instances could be effectively learned from the corpus of Chinese EHRs, it would be possible to perform deep phenotyping of Chinese EHRs by scanning linguistic patterns of PhenoSSU instances. Therefore, how to effectively learn linguistic patterns of PhenoSSU instances from the corpus of Chinese EHRs has become an important question.

Although linguistic patterns of PhenoSSU instances can be observed and summarized manually, this is a time-consuming process that depends on experienced experts. In the field of linguistic pattern mining, the Apriori-based method is one of the most representative algorithms, which was based on the principle of frequency counts of keyword occurrences [22]. The Apriori algorithm is well suited to simple linguistic pattern mining based on word co-occurrence. For example, a recent study used the Apriori algorithm to learn linguistic patterns of cyberbullying behaviors in a social networking service [23]. When two keywords co-occur frequently, they are considered

XSL•FO
RenderX

to constitute a potential linguistic pattern, such as the co-occurrence of "foolish" and "abuse." However, the linguistic patterns of the PhenoSSU instances are more complicated. Thus, Apriori-based methods are not competent at mining linguistic patterns of PhenoSSU instances because they cannot handle the co-occurrence of a phenotype and several attribute values simultaneously. Inspired by the work of Ofer et al [24], which considered biological sequences, such as DNA sequences, as human language and used advanced NLP tools to tackle biological tasks, we aimed to model Chinese EHRs as DNA-like sequences and mine linguistic patterns with advanced bioinformatics tools. In a recent review, Castellana et al [25] surveyed 16 classic DNA motif discovery tools and evaluated their ability to discover sequence motifs nested in 29 simulated sequence data sets. The MEME (Multiple Expectation Maximums for Motif Elicitation) motif discovery tool performed best among the 16 classic DNA motif discovery tools. In this study, we characterized phenotypes as "P" and attributes as "A" to transform the free text into a single-letter sequence that could be analyzed with the MEME motif discovery tool. The sequence motifs discovered in this single-letter sequence could be viewed as linguistic patterns of PhenoSSU instances in Chinese EHRs. Based on the linguistic patterns discovered in EHRs, we could identify PhenoSSU instances by recognizing linguistic patterns in free text. To summarize, the task of deep phenotyping of Chinese EHRs could be converted into two consecutive steps of sequence motif discovery and linguistic pattern recognition.

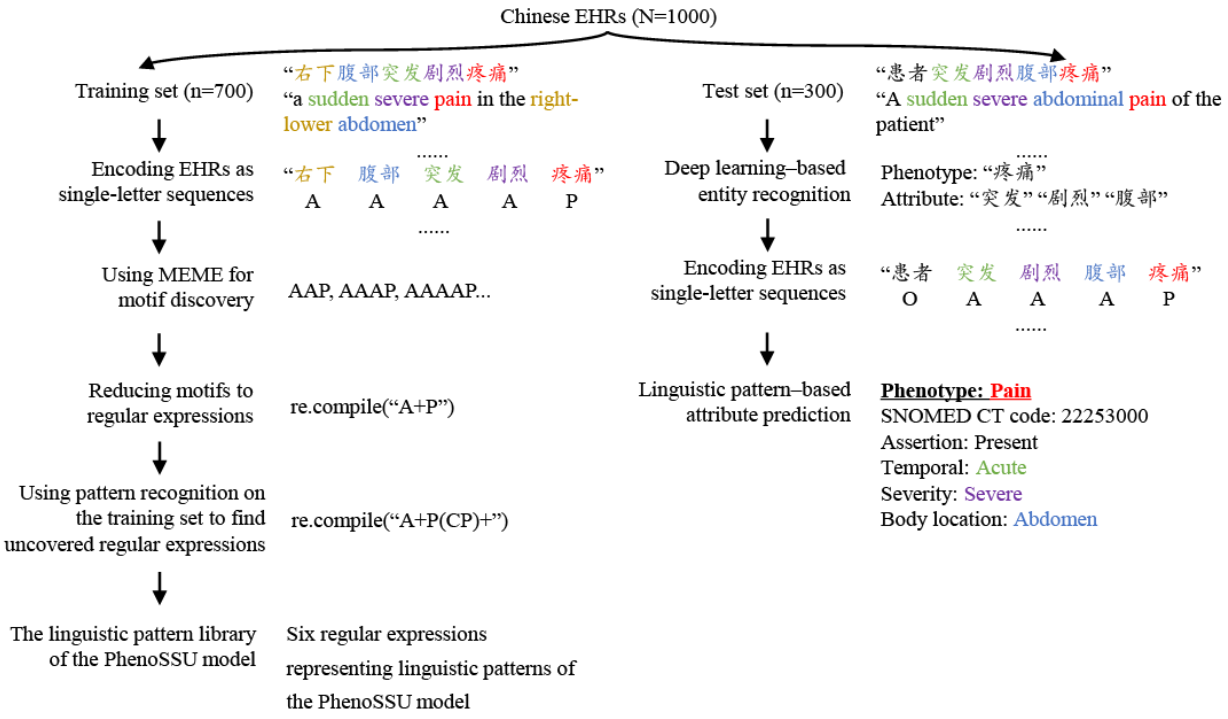
Following this idea, we aimed to identify linguistic patterns of PhenoSSU instances in Chinese EHRs with a biological sequence motif discovery tool and develop a deep-phenotyping algorithm for Chinese EHRs by scanning linguistic patterns in free text. The rest of this paper is organized as follows. The first section introduces the composition of the PhenoSSU model and its common linguistic patterns in free text. The second section introduces the method for using a biological sequence motif discovery tool to learn linguistic patterns from the corpus of Chinese EHRs. The third section introduces the method for recognizing PhenoSSU instances from Chinese EHRs based on linguistic patterns. The final section introduces a case study to illustrate the potential application of the deep-phenotyping algorithm. Although the deep-phenotyping algorithm developed in this study can only deal with Chinese EHRs, the underlying methodology can also be illuminating for other non–English-speaking countries.

Methods

Overview

In this study, a data-driven approach was proposed for learning linguistic patterns from Chinese EHRs. By using a pipeline of encoding the training set as a single-letter sequence and analyzing the sequence with the MEME motif discovery tool, we learned of six regular expressions and then introduced them into our pattern recognition–based algorithm for attribute prediction. The whole pipeline for the linguistic pattern–learning method is shown in Figure 1.

Figure 1. The pipeline for the linguistic pattern–learning method. A: attribute; C: punctuation; EHR: electronic health record; MEME: Multiple Expectation Maximums for Motif Elicitation; O: other information; P: phenotype; PhenoSSU: Semantic Structured Unit of Phenotypes; re.compile: a Python method used to compile a regular expression pattern; SNOMED CT: Systematized Nomenclature of Medicine–Clinical Terms.



XSL•FO
RenderX

The Design of the PhenoSSU Model for Representing Phenotype Information in Chinese EHRs

PhenoSSU is essentially an entity-attribute-value model consisting of phenotype terms along with standardized attributes from SNOMED CT. Compared with two commonly used information models named CEM and FHIR, the PhenoSSU model is more suitable for the task of deep phenotyping for two reasons. First, it has been shown that the PhenoSSU model is better at representing phenotype information in medical text than CEM and FHIR models [21]. Second, the PhenoSSU model puts more focus on characterizing phenotype traits with standardized attribute and value sets; as well, the attribute and value sets of the PhenoSSU model are easier to adjust according to the study-specific corpus.

To develop a fine-grained annotated corpus, 1000 Chinese EHRs of respiratory system diseases were manually annotated based on the PhenoSSU model, whose design was based on infectious diseases with a large proportion of respiratory diseases [21]. These 1000 Chinese EHRs were obtained from the EHR database of the Iiyi website [26]; all of the patients' private information in these EHRs have been masked by the Iiyi website.

During manual annotation, we optimized the attributes included in the PhenoSSU model to make them suitable for Chinese EHRs. The optimized PhenoSSU model contained 10 attributes, which could be further divided into two subtypes: (1) attributes for phrase-based phenotypes, such as "heavy cough" or "fever," including assertion, severity, temporal pattern, laterality, spatial pattern, quadrant pattern, and body location, and (2) attributes for logic-based phenotypes, such as "WBC [white blood cell] 12.5×10^9/L," including specimen, analyte, and abnormality. The composition of the PhenoSSU model is shown in Figure S1 and Table S1 in Multimedia Appendix 1, as well as the definitions, typical values, and SNOMED CT codes of attributes included in the model.

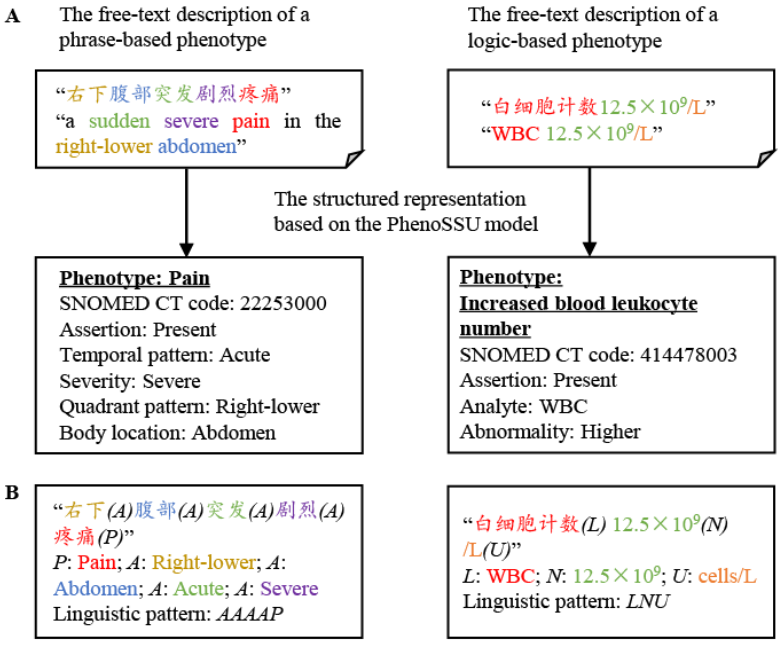
The phenotype information in free text could be structurally represented by the PhenoSSU model. For example, the description "a sudden severe pain in the right-lower abdomen" could be represented as a PhenoSSU instance consisting of the phenotype concept "pain," the assertion attribute "present," the temporal pattern attribute "acute," the severity attribute "severe," the quadrant pattern attribute "right-lower," and the body location attribute "abdomen." Meanwhile, logic-based phenotypes (ie, qualitative and quantitative test results) were also included in the PhenoSSU model. For example, "WBC 12.5×10^9/L" could be represented as a PhenoSSU instance consisting of the analyte "WBC" and the abnormality attribute "abnormality: higher," which was combined and normalized as a concept of the "increased blood leukocyte number (414478003)" in SNOMED CT (Figure 2, A). The relevant knowledge came from our previous study, LATTE (transforming lab test results) [27], which was integrated into this work, including sample sources, analyte names, and reference ranges for 1098 laboratory tests. Detailed information about the knowledge base is shown in Figures S2 and S3 in Multimedia Appendix 1.

Based on the annotation guideline of the PhenoSSU model in our previous work, two Chinese authors with medical backgrounds (LC and SL) manually annotated these medical records independently. Annotations were made on the brat rapid annotation tool platform [28]. The initial annotating agreement measured with the Cohen κ statistic was 0.851. All inconsistent annotations were decided by the project supervisor (TJ).

During annotation, we found some linguistic patterns of PhenoSSU instances in the EHR text. For example, the description of a phrase-based phenotype, "右下腹部突发剧烈疼痛" (English translation: "a sudden severe pain in the right-lower abdomen"), could be summarized as "attribute (right-lower) + attribute (abdomen) + attribute (acute) + attribute (severe) + phenotype (pain)." Similarly, the description of logic-based phenotypes had common linguistic patterns in free text, such as "analyte (WBC) + number (12.5×10^9) + unit (cells/ L)" (Figure 2, B). If we can mine linguistic patterns of PhenoSSU instances from Chinese EHRs, it would be possible to develop pattern recognition–based deep phenotyping.

XSL•FO
RenderX

Figure 2. Free-text phenotype descriptions and linguistic patterns. A. Examples of structuring free text by the PhenoSSU model. B. Examples of linguistic patterns in free text. A: attribute; L: analyte; N: number; P: pain; PhenoSSU: Semantic Structured Unit of Phenotypes; U: unit; WBC: white blood cell.

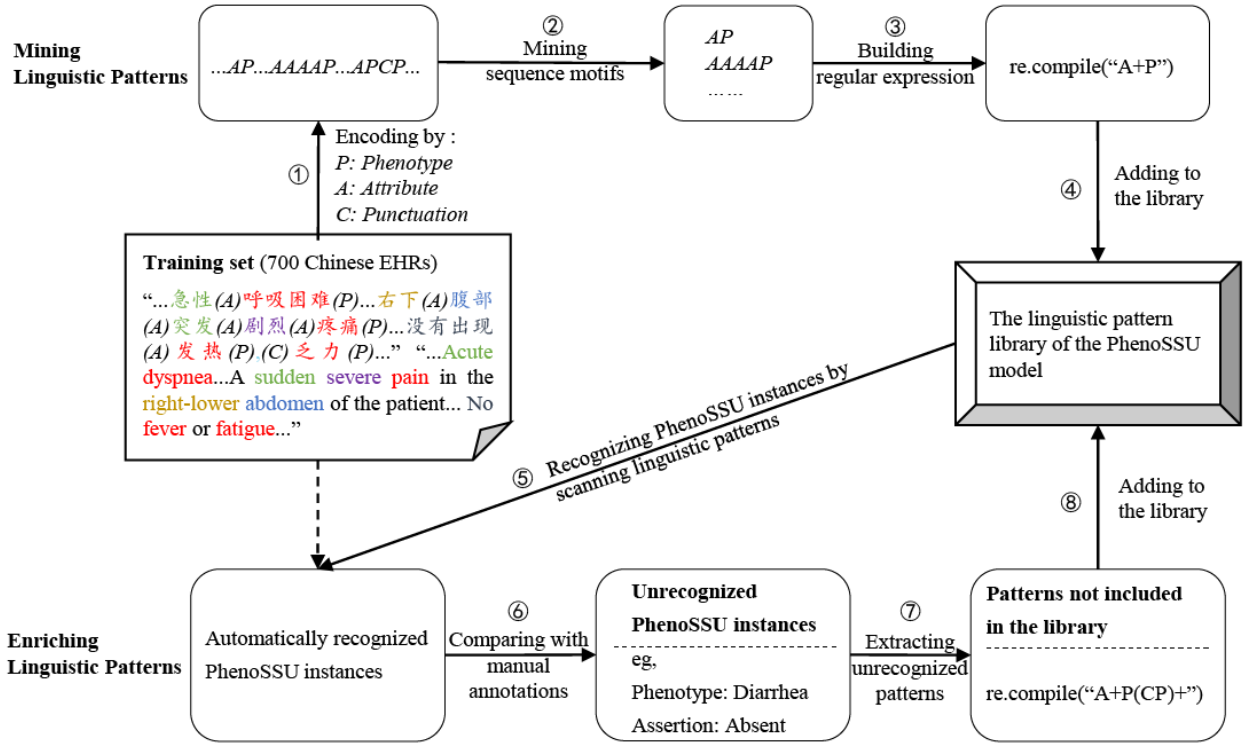


Learning Linguistic Patterns of PhenoSSU Instances From Chinese EHRs Using MEME: Workflow

Overview

In order to learn linguistic patterns of PhenoSSU instances from the Chinese EHR corpus, 1000 annotated Chinese EHRs in the study were divided into a training set (n=700, 70%) and test set (n=300, 30%). The workflow of linguistic pattern mining is shown in Figure 3, which includes two stages: pattern discovery and pattern enrichment. In the stage of linguistic pattern discovery, we used the MEME motif discovery tool, which solves the problem of motif mining with a maximum likelihood method [29] to obtain seed linguistic patterns of PhenoSSU instances. In the stage of linguistic pattern enrichment, a semiautomatic method was developed to check and fill linguistic pattern gaps. Through pattern discovery and enrichment, we built a linguistic pattern library of PhenoSSU instances.

Figure 3. The workflow of learning linguistic patterns of the PhenoSSU model from the corpus of Chinese electronic health records (EHRs). PhenoSSU: Semantic Structured Unit of Phenotypes; re.compile: a Python method used to compile a regular expression pattern.



XSL•FO
RenderX

Stage 1: Linguistic Pattern Discovery

First, free text in the training set was encoded into single-letter sequences. To represent EHRs as the input of the MEME motif discovery tool, we encoded them as single-letter sequences with the following criteria: the phenotype (ie, "fever" and "cough") was encoded as "P" and the attribute (ie, "severe") was encoded as "A." In the description of phrase-based phenotypes, "P" and "A" could be directly recognized in the original text. However, to calculate the abnormality of a logic-based phenotype, we need to combine the specimen ("S"), analyte ("L"), number ("N"), and unit ("U"). Specifically, the source of laboratory examination (ie, "blood" and "urine") was encoded as "S," the analyte (ie, "leukocyte") was encoded as "L," the number was encoded as "N" (ie, "37"), and the unit (ie, "°C") was encoded as "U." Meanwhile, the punctuation (ie, a comma) was encoded as "C," and other information was encoded as "O." In this study, EHRs were encoded using the FlashText tool, a tool for string-based concept recognition and replacement [30]. FlashText can find and replace keywords based on the trie dictionary data structure, which is 82 times faster than regular expressions. Because of its efficiency in processing text, we chose the FlashText tool for encoding text as single-letter sequences. Note that FlashText can retain the index of the strings in the original text. For example, the free-text description "患者主诉(O)急性(A)呼吸困难(P)...右下(A)腹部(A)突发(A)剧烈(A)疼痛(P)...没有出现(A)发热(P)，(C)乏力(P)" (English translation: "Patient complained of acute dyspnea...A sudden severe pain in the right-lower abdomen...No fever and fatigue") could be encoded as "AP...AAAAP...APCP." During this stage, we finally obtained single-letter sequences from whole EHRs in the training set.

Second, the MEME motif discovery tool was used to mine motifs in the single-letter sequence. The pipeline of MEME motif discovery is composed of three steps: finding starting points, maximizing the likelihood expectation, and scoring the discovered motifs.

The input was a set of unaligned sequences, and the output was a list of probable motifs. The statistical significance of the motifs in MEME was evaluated by the *E* value, which is based on the log-likelihood ratio. The settings of the MEME motif discovery tool were optimized as follows:

1. Motif discovery mode: classic mode. In classic mode, only one sequence needs to be provided. The algorithm will find the repeated sequence fragments in the sequence by likelihood ranking.
2. Select the site distribution: any number of repetitions. This option means selecting motifs that occur repeatedly.
3. How wide can motifs be: from 2 to 30. This number is the width (ie, characters in the sequence pattern) of a single motif. MEME can choose an optimal width of each motif individually by using a heuristic function. In the process, there were some motifs containing "O" (ie, other information), which was irrelevant to phenotype descriptions. Therefore, we separated out the motifs with the letter "O" to generate sequence segments that may represent linguistic patterns of PhenoSSU instances.

Third, we built regular expressions based on the discovered motifs. To make the motifs available in our algorithm, regular expressions were built. For example, we built a regular expression "A+P" based on motifs or sequence segments generated from motifs like "AP," "AAP," "AAAP," and "AAAAP."

Stage 2: Linguistic Pattern Enrichment

In this stage, a linguistic pattern recognition–based method was first used to automatically recognize PhenoSSU instances from Chinese EHRs in the training set. The workflow of linguistic pattern recognition is shown in Figure 4, which includes the following steps:

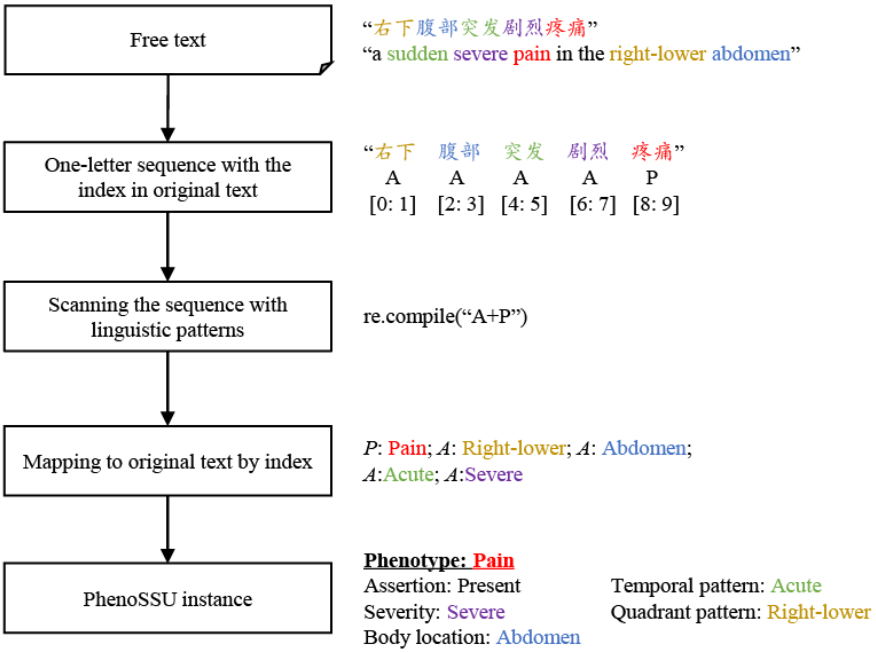
1. Encoding text as single-letter sequences. For example, the description "右下腹部突发剧烈疼痛" (English translation: "a sudden severe pain in the right-lower abdomen") was encoded as the single-letter sequence "AAAAP." The FlashText tool could record the position index of Chinese characters in every single letter, making it possible to map single letters to the original text. An example of the position index recording is shown in Figure S4 in Multimedia Appendix 1.
2. Scanning the single-letter sequence with the linguistic patterns. In this case, "AAAAP" matched the linguistic pattern "A + P" perfectly, meaning that the four attributes were associated with the phenotype.
3. Mapping these letters to the original text by index. A: right-lower; A: abdomen; A: acute; A: severe; P: pain.
4. Filling phenotypes and associated attributes in the PhenoSSU model. Finally, the description "右下腹部突发剧烈疼痛" could be transformed into a PhenoSSU instance consisting of the phenotype "pain," the assertion attribute "present," the temporal pattern attribute "acute," the severity attribute "severe," the quadrant pattern attribute "right-lower," and the body location attribute "abdomen."

Based on the above steps, we discovered the unrecognized PhenoSSU instances by comparing the automatically recognized instances with manual annotation. For example, the description "没有出现(A)发热(P),(C)乏力(P)" (English translation: "No fever or fatigue") could be encoded as "APCP," in which "AP" matched the regular expression ("A + P") in our pattern library. By mapping to the original text, "没有出现发热，乏力" was transformed into a PhenoSSU instance consisting of the phenotype "fever" and the assertion attribute "absent." However, "absent" was also the attribute of the phenotypes "diarrhea" and "weight loss," which were not recognized by the algorithm.

Finally, to check why these PhenoSSU instances were not recognized, all of them were encoded as single-letter sequences, which could be scanned with linguistic patterns. If no pattern matched, we collected such sequences to build new regular expressions and add them to the linguistic pattern library. In this example, sequences such as "APCPCP" were enriched into a regular expression "(A + P (CP) +)."

XSL•FO
RenderX

Figure 4. The workflow of recognizing PhenoSSU instances from free text via linguistic pattern recognition. The numbers within the square brackets represent the position indexes of single letters in the original text. A: attribute; P: phenotype; PhenoSSU: Semantic Structured Unit of Phenotypes; re.compile: a Python method used to compile a regular expression pattern.



Recognizing PhenoSSU Instances From Chinese EHRs: Workflow

The recognition of PhenoSSU instances could be divided into two subtasks: entity recognition and attribute prediction. To find the best strategy for the two tasks, it was essential to compare our proposed method with current state-of-the-art methods.

The first subtask was entity recognition, which aimed to recognize the text spans corresponding to phenotype and attribute entities. For the subtask of named entity recognition from Chinese EHRs, the Bidirectional Encoder Representations from Transformers (BERT)–Bi-LSTM-CRF model has proven its effectiveness in the CCKS (China Conference on Knowledge Graph and Semantic Computing) 2018 Task 1: Named Entity Recognition in Chinese Electronic Medical Records, which achieved the best F1 score of 91.43 [31]. Therefore, we compared algorithm performances of the BERT-Bi-LSTM-CRF model and the classic dictionary-based method in this study. The parameters of the BERT model were trained with the Kashgari package in Python (version 3.6.1; Python Software Foundation). In the dictionary-based method, the knowledge base of phenotypes was derived from the Chinese translations of the International Classification of Diseases, 10th Revision and 11th Revision, and the Human Phenotype Ontology (details in Table S2 in Multimedia Appendix 1). Further, the knowledge base of attribute trigger words was from the annotation of the training set. Entity recognition, combined with the other coding rules, was applied to encode free text as single-letter sequences, which would be used in the subsequent attribute prediction subtask.

The phenotype's attribute recognition was the second subtask, which aimed to predict appropriate values for the 10 attributes in the PhenoSSU model. The encoded single-letter sequences from the free text and the developed pattern recognition algorithm in the first subtask were used for attribute prediction. For the subtask of attribute prediction, we did not compare our pattern recognition algorithm with currently existing methods because the PhenoSSU model is a relatively new information model, and algorithms for deep-phenotyping Chinese EHRs based on the PhenoSSU model are very scarce. However, we have referred to state-of-the-art algorithms for deep-phenotyping English EHRs. For example, our previous work showed that the support vector machine (SVM)–based model performed best in the task of deep phenotyping of English clinical guidelines. That is why the SVM model was compared with the linguistic pattern–based method in this study. Three features were used in the SVM model: (1) the distance between phenotype and attribute words, (2) the number of pauses between phenotype and attribute words, and (3) the characteristics of attribute words (eg, some attribute words were only on the left side of phenotype words). The SVM model was built with the scikit-learn package (version 1.1.0) in Python. The parameter tuning of the SVM model was based on a hybrid search strategy. In this study, we did not use deep learning–based methods, because our previous work showed that they were not good at recognizing PhenoSSU instances, owing to the lack of training samples [21].

Evaluation of Algorithm Performance for Recognizing PhenoSSU Instances

To evaluate the algorithm's performance for recognizing PhenoSSU instances, we used the evaluation metrics in SemEval (Semantic Evaluation) 2015 Task 14: Analysis of Clinical Text [32].

In the subtask of entity recognition, the F1 score was taken as the evaluation metric. When a predicted entity word entirely coincided with a gold-standard text span, it was considered as a true positive. The precision metric was calculated as the

XSL•FO
RenderX

fraction of correctly predicted entities among all entities identified by the algorithm, and the recall metric was calculated as the fraction of correctly predicted entities among all entities identified by the annotators. The F1 score was calculated as the harmonic mean of precision and recall.

In the subtask of attribute prediction, the average accuracy and weighted average accuracy were taken as the evaluation metrics because the weighted average accuracy thoroughly considered the distribution of each attribute value in the corpus, which could better evaluate those attribute values with little distribution.

For the evaluation at the PhenoSSU-instance level, we used the combination of the F1 score for entity recognition and weighted average for attribute prediction. A PhenoSSU instance was considered correct when the phenotype and associated attribute values annotated by the algorithm were the same as the corresponding PhenoSSU instance annotated by experts.

Ethical Considerations

The 1000 Chinese EHRs of respiratory system diseases used in this study were obtained from the EHR database of the Iiyi website [26]. No ethics approval was needed because the data from downloaded EHRs, including patients' private information, were all masked by the Iiyi website.

Results

Linguistic Patterns of PhenoSSU Instances Learned From Chinese EHRs

A total of 51 sequence motifs were discovered from the Chinese EHRs in the training set (details are shown in Figure S5 in Multimedia Appendix 1). Based on the 51 motifs, we built six regular expressions (Table 1), namely linguistic patterns of the PhenoSSU instances in the Chinese EHRs. Among the regular expressions of phrase-based phenotypes, "AP +" appeared most frequently. The most common description of this regular expression was "absent" plus phenotypes, which could be used for differential diagnosis in clinical practice. The second frequent regular expression was "A + P," which usually corresponded to a detailed description of phenotypes, such as "body location + severity + phenotype." There were also complex linguistic patterns to be generalized as "A × PC × A +," for example, "严重(A)咳嗽(P)，(C)呈持续性(A)" (ie, severe cough, consistently). Among the regular expressions of logic-based phenotypes, the most typical was "S × LNU," such as the description "WBC 12 × 10^9/L." There were also linguistic patterns that directly interpreted laboratory examination results: "S × LR [results of laboratory examination]," such as "血糖升高" (ie, high blood glucose). The above results suggest that there are inherent linguistic patterns in Chinese EHRs. The detailed frequency of linguistic patterns is shown in Table S3 in Multimedia Appendix 1.

In this study, six regular expressions were learned from 700 Chinese EHRs in the training set. However, the size of the training set could be smaller than 700 in order to build the six regular expressions. To explore the potential smallest size of the training set, we conducted an experiment to explore the minimum number of EHRs that could match all six regular expressions. In the experiment, we randomly selected EHRs from the training set with stepwise increased data size, which were scanned with the six regular expressions. When all six regular expressions could be matched, that data size was recorded. This process was repeated 1000 times to calculate the mean and SD of the EHR sums that covered the six regular expressions. Results showed that in a mean of 134 (SD 9.7) EHRs, the six regular expressions could be matched. We did not use the pattern discovery method illustrated in this study because there was a semiautomatic step in the method. Repeating the pattern discovery method 1000 times would be time-consuming. A line graph was plotted to show five examples among all 1000 tests (Figure S6 in Multimedia Appendix 1).

XSL•FO
RenderX

Table 1. Six regular expressions based on linguistic patterns of the Chinese electronic health record corpus in this study.

Phenotype category and regular expressions	Example in Chinese (English translation)
Phrase-based phenotypes	
re.compile[a]("A[b]+P[c](C[d]P)+")	"无/A咳嗽/P、/C发热/P" (no cough or fever)
re.compile("AP+")	"严重/A腹痛/P腹泻/P" (severe abdominal pain and diarrhea)
re.compile("A+P")	"右下腹/A严重/A疼痛/P" (severe right-lower abdominal pain)
re.compile("A×PC×A+")	"咳嗽/P，/C呈持续性/A" (cough, consistently)
Logic-based phenotypes	
re.compile("S[e]×L[f]N[g]U[h]")	"白细胞/L 12×10^9/N /L/U" (WBC[i] 12 × 10^9/L)
re.compile("S×LR[j]")	"血/S糖/L升高/R" (high blood glucose)

[a]re.compile: a Python method used to compile a regular expression pattern.
[b]A: attribute.
[c]P: phenotype.
[d]C: punctuation.
[e]S: specimen.
[f]L: analyte.
[g]N: number.
[h]U: unit.
[i]WBC: white blood cell.
[j]R: results of laboratory examination.

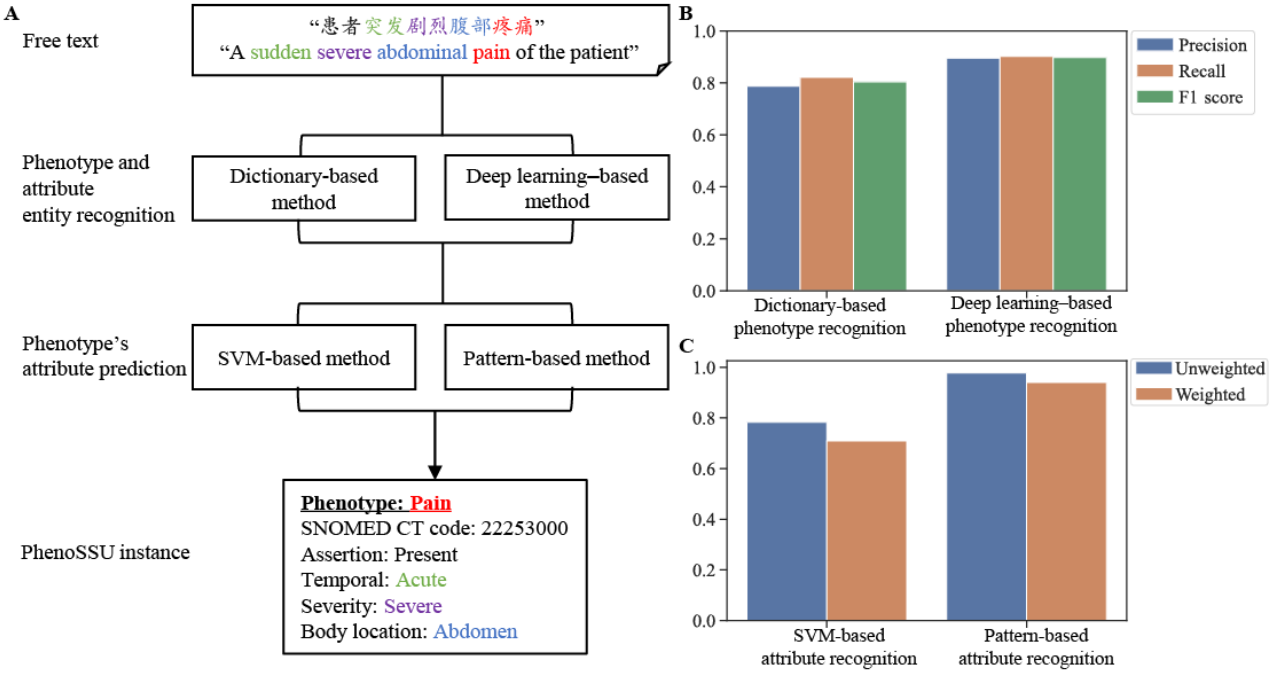
The Best Strategy for Recognizing PhenoSSU Instances

Based on the linguistic patterns of Chinese EHRs, we developed a pattern recognition–based method to identify PhenoSSU instances. To find the best strategy for recognizing PhenoSSU instances, we developed and compared different methods in the subtasks of entity recognition and attribute prediction. The results in Figure 5 show that the best strategy was to recognize entities using the deep learning–based method and then predict the attribute values using the pattern recognition–based method.

Specifically, in the entity recognition subtask, the method of deep learning (ie, BERT-Bi-LTSM-CRF) achieved the best performance, with an F1 score of 0.898 (Figure 5, B). As a comparison, the dictionary-based method achieved an F1 score of 0.804. In the subtask of attribute prediction, the pattern recognition–based method had the best performance, with an accuracy of 0.977 and a weighted average of 0.940 (Figure 5, C). The SVM-based method achieved an accuracy of 0.783 and a weighted average of 0.709. The deep-phenotyping algorithm for Chinese EHRs had an overall accuracy of 0.844 on the test set. The detailed performances of the two models for predicting attribute values are shown in Table S4 in Multimedia Appendix 1.

XSL•FO
RenderX

Figure 5. Determining the best strategy for recognizing PhenoSSU instances. A. The workflow of recognizing PhenoSSU instances from free text. B. The performance comparison between the dictionary-based method and the deep learning–based method in identifying phenotype concepts. C. The performance comparison between the SVM-based method and the pattern recognition–based method in recognizing a phenotype's attributes. PhenoSSU: Semantic Structured Unit of Phenotypes; SNOMED CT: Systematized Nomenclature of Medicine–Clinical Terms; SVM: support vector machine.

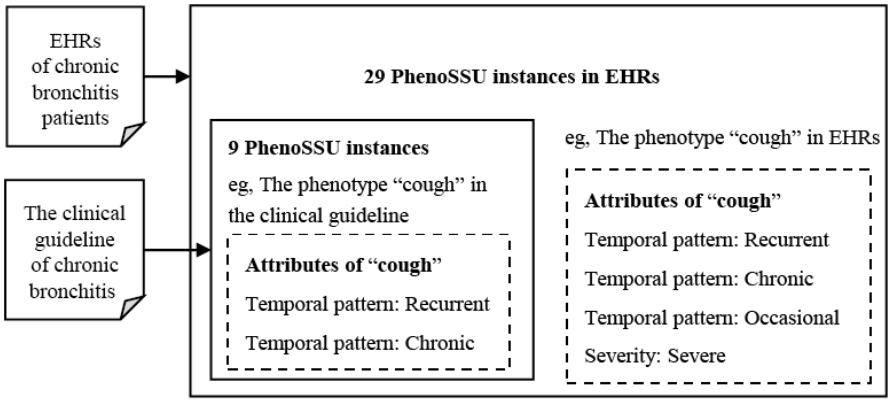


Case Study: Exploring the Real-World Evidence That Deep-Phenotyping EHRs Can Update Knowledge in Guidelines

With the pattern recognition algorithm, we could effectively structure phenotype information in Chinese EHRs. To demonstrate the potential application of deep phenotyping, a case study was conducted to update clinical guidelines by information retrieval of EHRs. In the case study, we selected the latest Chinese clinical guideline and 300 Chinese EHRs of chronic bronchitis. To recognize PhenoSSU instances from the guideline and the EHRs, we used the optimized hybrid strategy mentioned previously.

A total of 9 and 29 PhenoSSU instances were identified from the clinical guideline and the EHRs of chronic bronchitis, respectively (details are shown in Tables S5-S7 in Multimedia Appendix 1). The 9 PhenoSSU instances identified in the clinical guideline appeared in the EHRs, which meant another 20 PhenoSSU instances in the EHRs were not covered in the clinical guideline. For example, "cough: chronic" and "cough: recurrent" both appeared in the clinical guideline and the EHRs. However, the current guideline could not give suggestions to accurately diagnose patients with occasional cough or severe cough as having chronic bronchitis (Figure 6). This real-world evidence hints at the feasibility of updating knowledge in clinical guidelines through deep phenotyping of large-scale EHRs.

Figure 6. The comparison of PhenoSSU instances extracted from the clinical guidelines and electronic health records (EHRs) of chronic bronchitis. PhenoSSU: Semantic Structured Unit of Phenotypes.



XSL•FO
RenderX

Discussion

Principal Findings

In this study, we developed a simple but effective strategy to perform deep phenotyping of Chinese EHRs. The core of this strategy is learning linguistic patterns of PhenoSSU instances with a motif discovery tool from the field of bioinformatics. According to this research, biological sequence motif discovery tools could be used to effectively identify linguistic patterns of phenotype descriptions from medical texts after encoding them as DNA-like sequences. Meanwhile, the process of identifying linguistic patterns does not require too much annotation data; thus, our strategy is suitable for low-resource scenarios of deep-phenotyping Chinese EHRs.

This study was a preliminary attempt to use bioinformatics tools to tackle problems in medical informatics. By modeling natural language as single-letter sequences, it is possible that other advanced tools for analyzing biological sequences could also be used for processing natural language. For example, some researchers in the NLP field have applied a classic informatics algorithm, named the Basic Local Alignment Search Tool (BLAST), [33] to the text reuse detection task [34]. In Vesanto's work [35], the 23 most-used English letters in the data set were calculated to form a simple one-to-one mapping between English letters and arbitrary amino acids. In this way, text was encoded into single-letter sequences that BLAST could handle to calculate similarities between texts. It is believed that future communications between bioinformatics and medical informatics will become more frequent [36].

It can be concluded from this study that there exist some regular linguistic patterns for phenotype narratives in Chinese EHRs. The origin of these linguistic patterns may be the common writing habits of clinicians who try to save time by recording clinical information faithfully in as few words as possible [37]. The reason our strategy does not require large annotation samples is that it uses the inner knowledge of linguistic patterns. As we know, data-hungry strategies, such as machine learning and deep learning, require many training samples to effectively identify patterns from data. However, there are many low-resource scenarios in practice that lack sufficient annotation samples for machine learning or deep learning. This is perhaps the reason why the majority (60%) of NLP studies in the medical domain have continued to use a knowledge-based approach rather than a machine learning–based approach [4]. In recent years, researchers have become increasingly focused on integrating machine learning with human knowledge [38], which is expected to become a new paradigm to deal with low-resource scenarios in medical informatics [39].

Limitations

One limitation of this study was that linguistic patterns were learned from EHRs of respiratory diseases, which may not be applicable to other diseases. In addition, limited by the data size, the linguistic patterns in our study might be incomplete. In the future, we will continue to improve the algorithm with more Chinese EHRs from different hospital departments.

Conclusions

We developed a simple but effective strategy to perform deep phenotyping of Chinese EHRs with limited fine-grained annotation data. Our work will promote the second use of Chinese EHRs and bring inspiration to other non–English-speaking countries.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (grant 2021YFC2302000), the Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences (grants 2021-1-I2M-051 and 2021-I2M-1-001), the National Natural Science Foundation of China (grant 31671371), and the Emergency Key Program of Guangzhou Laboratory (grant EKPG21-12).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials.
[DOCX File , 902 KB - jmir_v24i6e37213_app1.docx]

References

1. Hong C, Rush E, Liu M, Zhou D, Sun J, Sonabend A, VA Million Veteran Program. Clinical knowledge extraction via sparse embedding regression (KESER) with multi-center large scale electronic health record data. NPJ Digit Med 2021 Oct 27;4(1):151 [FREE Full text] [doi: 10.1038/s41746-021-00519-z] [Medline: 34707226]
2. Katzan IL, Rudick RA. Time to integrate clinical and research informatics. Sci Transl Med 2012 Nov 28;4(162):162fs41. [doi: 10.1126/scitranslmed.3004583] [Medline: 23197569]
3. Elkin PL, Mullin S, Mardekian J, Crowner C, Sakilay S, Sinha S, et al. Using artificial intelligence with natural language processing to combine electronic health record's structured and free text data to identify nonvalvular atrial fibrillation to decrease strokes and death: Evaluation and case-control study. J Med Internet Res 2021 Nov 09;23(11):e28946 [FREE Full text] [doi: 10.2196/28946] [Medline: 34751659]

XSL•FO
RenderX

4. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: A literature review. J Biomed Inform 2018 Jan;77:34-49 [FREE Full text] [doi: 10.1016/j.jbi.2017.11.011] [Medline: 29162496]
5. Chiu P, Hripcsak G. EHR-based phenotyping: Bulk learning and evaluation. J Biomed Inform 2017 Jun;70:35-51 [FREE Full text] [doi: 10.1016/j.jbi.2017.04.009] [Medline: 28410982]
6. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE J Biomed Health Inform 2018 Sep;22(5):1589-1604 [FREE Full text] [doi: 10.1109/JBHI.2017.2767063] [Medline: 29989977]
7. Oniki TA, Zhuo N, Beebe CE, Liu H, Coyle JF, Parker CG, et al. Clinical element models in the SHARPn consortium. J Am Med Inform Assoc 2016 Mar;23(2):248-256 [FREE Full text] [doi: 10.1093/jamia/ocv134] [Medline: 26568604]
8. Shi X, Yi Y, Xiong Y, Tang B, Chen Q, Wang X, et al. Extracting entities with attributes in clinical text via joint deep learning. J Am Med Inform Assoc 2019 Dec 01;26(12):1584-1591 [FREE Full text] [doi: 10.1093/jamia/ocz158] [Medline: 31550346]
9. Hripcsak G, Albers DJ. High-fidelity phenotyping: Richness and freedom from bias. J Am Med Inform Assoc 2018 Mar 01;25(3):289-294 [FREE Full text] [doi: 10.1093/jamia/ocx110] [Medline: 29040596]
10. Zhang XA, Yates A, Vasilevsky N, Gourdine JP, Callahan TJ, Carmody LC, et al. Semantic integration of clinical laboratory tests from electronic health records for deep phenotyping and biomarker discovery. NPJ Digit Med 2019;2:32 [FREE Full text] [doi: 10.1038/s41746-019-0110-4] [Medline: 31119199]
11. Delude CM. Deep phenotyping: The details of disease. Nature 2015 Nov 05;527(7576):S14-S15. [doi: 10.1038/527S14a] [Medline: 26536218]
12. Weng C, Shah NH, Hripcsak G. Deep phenotyping: Embracing complexity and temporality-Towards scalability, portability, and interoperability. J Biomed Inform 2020 May;105:103433 [FREE Full text] [doi: 10.1016/j.jbi.2020.103433] [Medline: 32335224]
13. Robinson PN. Deep phenotyping for precision medicine. Hum Mutat 2012 May;33(5):777-780. [doi: 10.1002/humu.22080] [Medline: 22504886]
14. Peterson KJ, Jiang G, Liu H. A corpus-driven standardization framework for encoding clinical problems with HL7 FHIR. J Biomed Inform 2020 Oct;110:103541 [FREE Full text] [doi: 10.1016/j.jbi.2020.103541] [Medline: 32814201]
15. Aronson AR, Lang F. An overview of MetaMap: Historical perspective and recent advances. J Am Med Inform Assoc 2010;17(3):229-236 [FREE Full text] [doi: 10.1136/jamia.2009.002733] [Medline: 20442139]
16. Ayaz M, Pasha MF, Alzahrani MY, Budiarto R, Stiawan D. The Fast Health Interoperability Resources (FHIR) standard: Systematic literature review of implementations, applications, challenges and opportunities. JMIR Med Inform 2021 Jul 30;9(7):e21929 [FREE Full text] [doi: 10.2196/21929] [Medline: 34328424]
17. Xu J, Li Z, Wei Q, Wu Y, Xiang Y, Lee H, et al. Applying a deep learning-based sequence labeling approach to detect attributes of medical concepts in clinical text. BMC Med Inform Decis Mak 2019 Dec 05;19(Suppl 5):236 [FREE Full text] [doi: 10.1186/s12911-019-0937-2] [Medline: 31801529]
18. Tao C, Jiang G, Oniki TA, Freimuth RR, Zhu Q, Sharma D, et al. A semantic-web oriented representation of the clinical element model for secondary use of electronic health records data. J Am Med Inform Assoc 2013 May 01;20(3):554-562 [FREE Full text] [doi: 10.1136/amiajnl-2012-001326] [Medline: 23268487]
19. Lu Z, Sim J, Wang JX, Forrest CB, Krull KR, Srivastava D, et al. Natural language processing and machine learning methods to characterize unstructured patient-reported outcomes: Validation study. J Med Internet Res 2021 Nov 03;23(11):e26777 [FREE Full text] [doi: 10.2196/26777] [Medline: 34730546]
20. Yeh MC, Wang Y, Yang H, Bai K, Wang H, Li YJ. Artificial intelligence-based prediction of lung cancer risk using nonimaging electronic medical records: Deep learning approach. J Med Internet Res 2021 Aug 03;23(8):e26256 [FREE Full text] [doi: 10.2196/26256] [Medline: 34342588]
21. Deng L, Chen L, Yang T, Liu M, Li S, Jiang T. Constructing high-fidelity phenotype knowledge graphs for infectious diseases with a fine-grained semantic information model: Development and usability study. J Med Internet Res 2021 Jun 15;23(6):e26892 [FREE Full text] [doi: 10.2196/26892] [Medline: 34128811]
22. Fard MJS, Namin PA. Review of Apriori based frequent itemset mining solutions on big data. In: Proceedings of the 6th International Conference on Web Research. 2020 Presented at: The 6th International Conference on Web Research; April 22-23, 2020; Tehran, Iran p. 157-164. [doi: 10.1109/icwr49608.2020.9122295]
23. Zainol Z, Wani S, Nohuddin PNE, Noormanshah WMU, Marzukhi S. Association analysis of cyberbullying on social media using Apriori algorithm. Int J Eng Technol 2018;7(4.29):72-75 [FREE Full text] [doi: 10.14419/ijet.v7i4.29.21847]
24. Ofer D, Brandes N, Linial M. Comput Struct Biotechnol J 2021;19:1750-1758 [FREE Full text] [doi: 10.1016/j.csbj.2021.03.022] [Medline: 33897979]
25. Castellana S, Biagini T, Parca L, Petrizzelli F, Bianco SD, Vescovi AL, et al. A comparative benchmark of classic DNA motif discovery tools on synthetic data. Brief Bioinform 2021 Nov 05;22(6):bbab303. [doi: 10.1093/bib/bbab303] [Medline: 34351399]
26. Iiyi. URL: https://bingli.iiyi.com/ [accessed 2022-05-25]

XSL•FO
RenderX

27. Jiang K, Yang T, Wu C, Chen L, Mao L, Wu Y, et al. LATTE: A knowledge-based method to normalize various expressions of laboratory test results in free text of Chinese electronic health records. J Biomed Inform 2020 Feb;102:103372 [FREE Full text] [doi: 10.1016/j.jbi.2019.103372] [Medline: 31901507]
28. Stenetorp P, Pyysalo S, Topić G, Ohta T, Ananiadou S, Tsujii J. brat: A web-based tool for NLP-assisted text annotation. In: Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics. 2012 Presented at: The 13th Conference of the European Chapter of the Association for Computational Linguistics; April 23-27, 2012; Avignon, France p. 102-107.
29. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. Nucleic Acids Res 2015 Jul 01;43(W1):W39-W49 [FREE Full text] [doi: 10.1093/nar/gkv416] [Medline: 25953851]
30. Singh V. Replace or retrieve keywords in documents at scale. ArXiv. Preprint posted online on November 9, 2017 [FREE Full text]
31. Zhang J, Li J, Jiao Z, Yan J. Overview of CCKS 2018 Task 1: Named Entity Recognition in Chinese Electronic Medical Records. In: Proceedings of the 4th China Conference on Knowledge Graph and Semantic Computing. 2019 Presented at: The 4th China Conference on Knowledge Graph and Semantic Computing; August 24-27, 2019; Hangzhou, China p. 158-164. [doi: 10.1007/978-981-15-1956-7_14]
32. Elhadad N, Pradhan S, Gorman S, Manandhar S, Chapman W, Savova G. SemEval-2015 Task 14: Analysis of Clinical Text. In: Proceedings of the 9th International Workshop on Semantic Evaluation. 2015 Presented at: The 9th International Workshop on Semantic Evaluation; June 4-5, 2015; Denver, CO p. 303-310 URL: https://aclanthology.org/S15-2051.pdf [doi: 10.18653/v1/s15-2051]
33. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: A better web interface. Nucleic Acids Res 2008 Jul 01;36(Web Server issue):W5-W9 [FREE Full text] [doi: 10.1093/nar/gkn201] [Medline: 18440982]
34. Vierthaler P, Gelein M. A BLAST-based, language-agnostic text reuse algorithm with a MARKUS implementation and sequence alignment optimized for large Chinese corpora. J Cult Anal 2019;4(2):1-25 [FREE Full text] [doi: 10.22148/16.034]
35. Vesanto A. Detecting and Analyzing Text Reuse With BLAST [master's thesis]. Turku, Finland: University in Turku; 2018. URL: https://www.utupub.fi/bitstream/handle/10024/146706/Vesanto_Aleksi_opinnayte.pdf?isAllowed=y&sequence=1 [accessed 2022-05-21]
36. Ouyang Z, Sargeant J, Thomas A, Wycherley K, Ma R, Esmaeilbeigi R, et al. A scoping review of 'big data', 'informatics', and 'bioinformatics' in the animal health and veterinary medical literature. Anim Health Res Rev 2019 Dec 18;20(1):1-18. [doi: 10.1017/s1466252319000136]
37. Han H, Lopp L. Writing and reading in the electronic health record: An entirely new world. Med Educ Online 2013 Feb 05;18:1-7 [FREE Full text] [doi: 10.3402/meo.v18i0.18634] [Medline: 23394976]
38. Deng C, Ji X, Rainey C, Zhang J, Lu W. Integrating machine learning with human knowledge. iScience 2020 Nov 20;23(11):101656 [FREE Full text] [doi: 10.1016/j.isci.2020.101656] [Medline: 33134890]
39. de Jong J, Cutcutache I, Page M, Elmoufti S, Dilley C, Fröhlich H, et al. Towards realizing the vision of precision medicine: AI based prediction of clinical drug response. Brain 2021 Jul 28;144(6):1738-1750 [FREE Full text] [doi: 10.1093/brain/awab108] [Medline: 33734308]

Abbreviations

A: attribute (in the context of single-letter sequences)
BERT: Bidirectional Encoder Representations from Transformers
Bi-LSTM-CRF: bidirectional long short-term memory and conditional random field
BLAST: Basic Local Alignment Search Tool
C: punctuation (in the context of single-letter sequences)
CCKS: China Conference on Knowledge Graph and Semantic Computing
CEM: clinical element model
EHR: electronic health record
FHIR: Fast Healthcare Interoperability Resources
L: analyte (in the context of single-letter sequences)
LATTE: transforming lab test results
MEME: Multiple Expectation Maximums for Motif Elicitation
N: number (in the context of single-letter sequences)
NLP: natural language processing
O: other information (in the context of single-letter sequences)
P: phenotype (in the context of single-letter sequences)
PhenoSSU: Semantic Structured Unit of Phenotypes
R: results of laboratory examination (in the context of single-letter sequences)
S: specimen (in the context of single-letter sequences)

XSL•FO
RenderX

SemEval: Semantic Evaluation
SNOMED CT: Systematized Nomenclature of Medicine–Clinical Terms
SVM: support vector machine
U: unit (in the context of single-letter sequences)
WBC: white blood cell

Edited by A Mavragani; submitted 11.02.22; peer-reviewed by HH Wang, S Sakilay, H Liu; comments to author 09.03.22; revised version received 21.04.22; accepted 12.05.22; published 03.06.22.

Please cite as:
Li S, Deng L, Zhang X, Chen L, Yang T, Qi Y, Jiang T
Deep Phenotyping of Chinese Electronic Health Records by Recognizing Linguistic Patterns of Phenotypic Narratives With a Sequence Motif Discovery Tool: Algorithm Development and Validation
J Med Internet Res 2022;24(6):e37213
URL: https://www.jmir.org/2022/6/e37213
doi:10.2196/37213
PMID:35657661

©Shicheng Li, Lizong Deng, Xu Zhang, Luming Chen, Tao Yang, Yifan Qi, Taijiao Jiang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.06.2022. 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

Impactability Modeling for Reducing Medicare Accountable Care Organization Payments and Hospital Events in High-Need High-Cost Patients: Longitudinal Cohort Study

Maureen A Smith[1,2], MD, MPH, PhD; Menggang Yu[1], PhD; Jared D Huling[3], PhD; Xinyi Wang[1], PhD; Allie DeLonay[1], MS; Jonathan Jaffery[4,5], MD, MS

[1]Department of Population Health Sciences, University of Wisconsin–Madison, Madison, WI, United States

[2]Department of Family Medicine and Community Health, University of Wisconsin–Madison, Madison, WI, United States

[3]Division of Biostatistics, University of Minnesota, Minneapolis, MN, United States

[4]University of Wisconsin Health Office of Population Health, Madison, WI, United States

[5]Department of Medicine, University of Wisconsin–Madison, Madison, WI, United States

Corresponding Author:
Maureen A Smith, MD, MPH, PhD
Department of Population Health Sciences
University of Wisconsin–Madison
800 University Bay Dr
Suite 210
Madison, WI, 53794
United States
Phone: 1 6082624802
Email: maureensmith@wisc.edu

Abstract

Background: Impactability modeling promises to help solve the nationwide crisis in caring for high-need high-cost patients by matching specific case management programs with patients using a "benefit" or "impactability" score, but there are limitations in tailoring each model to a specific program and population.

Objective: We evaluated the impact on Medicare accountable care organization savings from developing a benefit score for patients enrolled in a historic case management program, prospectively implementing the score, and evaluating the results in a new case management program.

Methods: We conducted a longitudinal cohort study of 76,140 patients in a Medicare accountable care organization with multiple before-and-after measures of the outcome, using linked electronic health records and Medicare claims data from 2012 to 2019. There were 489 patients in the historic case management program, with 1550 matched comparison patients, and 830 patients in the new program, with 2368 matched comparison patients. The historic program targeted high-risk patients and assigned a centrally located registered nurse and social worker to each patient. The new program targeted high- and moderate-risk patients and assigned a nurse physically located in a primary care clinic. Our primary outcomes were any unplanned hospital events (admissions, observation stays, and emergency department visits), count of event-days, and Medicare payments.

Results: In the historic program, as expected, high-benefit patients enrolled in case management had fewer events, fewer event-days, and an average US $1.15 million reduction in Medicare payments per 100 patients over the subsequent year when compared with the findings in matched comparison patients. For the new program, high-benefit high-risk patients enrolled in case management had fewer events, while high-benefit moderate-risk patients enrolled in case management did not differ from matched comparison patients.

Conclusions: Although there was evidence that a benefit score could be extended to a new case management program for similar (ie, high-risk) patients, there was no evidence that it could be extended to a moderate-risk population. Extending a score to a new program and population should include evaluation of program outcomes within key subgroups. With increased attention on value-based care, policy makers and measure developers should consider ways to incorporate impactability modeling into program design and evaluation.

XSL•FO
RenderX

(J Med Internet Res 2022;24(6):e29420) doi:10.2196/29420

KEYWORDS

case management; high-risk patients; benefit score

Introduction

With a national imperative to reduce costs and improve care for high-need high-cost patients [1-3], most accountable care organizations (ACOs) [4] are implementing outpatient case management programs with a nurse or social worker coordinating care. However, there is little evidence of cost savings from these resource-intensive programs [5-7], and they vary widely in design and implementation [5,8]. This widespread implementation of unproven programs has concerned policy makers [9,10] and led to recommendations to design more effective programs [11], to improve the identification of potentially high-cost patients using predictive models [12], and even to abandon care coordination as a cost-saving strategy [13]. For example, case managers often identify patients for enrollment in a case management program using a predictive risk score to find patients at high risk of poor outcomes, such as hospital admission [14].

Rather than attempting to identify effective case management programs and standardize implementation across health systems, a fundamentally different strategy would identify patients who benefit from specific case management programs as they are implemented in practice and match patients to the most beneficial program [5]. Described by Lewis et al as "impactability modeling" [15], this pragmatic approach predicts who is likely to benefit from a particular intervention with respect to an outcome and not who is likely to have a poor outcome. Different from risk scores, these "impactability" or "benefit" scores identify patients who are likely to benefit from enrollment into case management with respect to a specific outcome (eg, preventing hospital admissions). In this way, a benefit score could allow further partitioning of a high-risk population of patients into those who are and those who are not likely to benefit from case management. For example, a high-risk patient may or may not be likely to avoid hospitalizations if enrolled into case management. This additional stratification of a high-risk population using a benefit score may help an ACO target patients for enrollment into case management. Early analyses on impactability in the Medicare population [16] (labeled "benefit score") and in the Medicaid population [17] (labeled "impactability score") successfully developed scores to identify individuals who were more likely to benefit from certain case management programs and suggested significant savings [16]. However, neither score was evaluated to determine if it could extend beyond the case management program and population on which it was developed.

The promise of impactability modeling is based on substantial evidence that patients may be more or less likely to benefit from a specific intervention depending on their personal and clinical characteristics [5], although this tailoring of program enrollment may come at a cost. Because impactability models are intrinsically linked to the specific program and population used in their development and because there is wide variation in case management program design and implementation [5,8], it is unclear whether these scores could extend to new programs or populations. To address this question, we evaluated the impact on Medicare ACO savings from a case management "benefit score" developed using a historic case management program enrolling high-risk patients (published elsewhere [16]), and compared the results to prospectively implementing the score in a new case management program enrolling both high- and moderate-risk populations. Our work extends analyses conducted under a Patient-Centered Outcomes Research Institute (PCORI)-funded health systems demonstration grant (HSD-1603-35039; "Variation in case management programs and their effectiveness in managing high-risk patients for Medicare ACOs" [18]).

Methods

Study Design and Setting

We used a longitudinal cohort study design with multiple before-and-after measures of the outcome for each case management patient and matched comparison patient [19-21]. Linked electronic health record (EHR) and Medicare enrollment and claims data were extracted from January 1, 2012, through April 30, 2019, to characterize patients during a 1-year baseline period and up to a 1-year follow-up period, as well as census data from the 2007-2011 American Community Survey. The setting was UW Health, a large health system in Wisconsin with 30 statewide academic and community-based primary care clinics and 279 primary care providers. UW Health began participating in the Medicare ACO program in 2013 and, as part of its commitment to become a learning health system [22], began developing scores to support targeted enrollment of patients into population management programs and regularly evaluating program outcomes.

Ethical Considerations

This project was deemed exempt from institutional review board oversight at the University of Wisconsin–Madison as it constitutes quality improvement or program evaluation [23]. Institutional review board review was not required because, in accordance with federal regulations, the project does not constitute research.

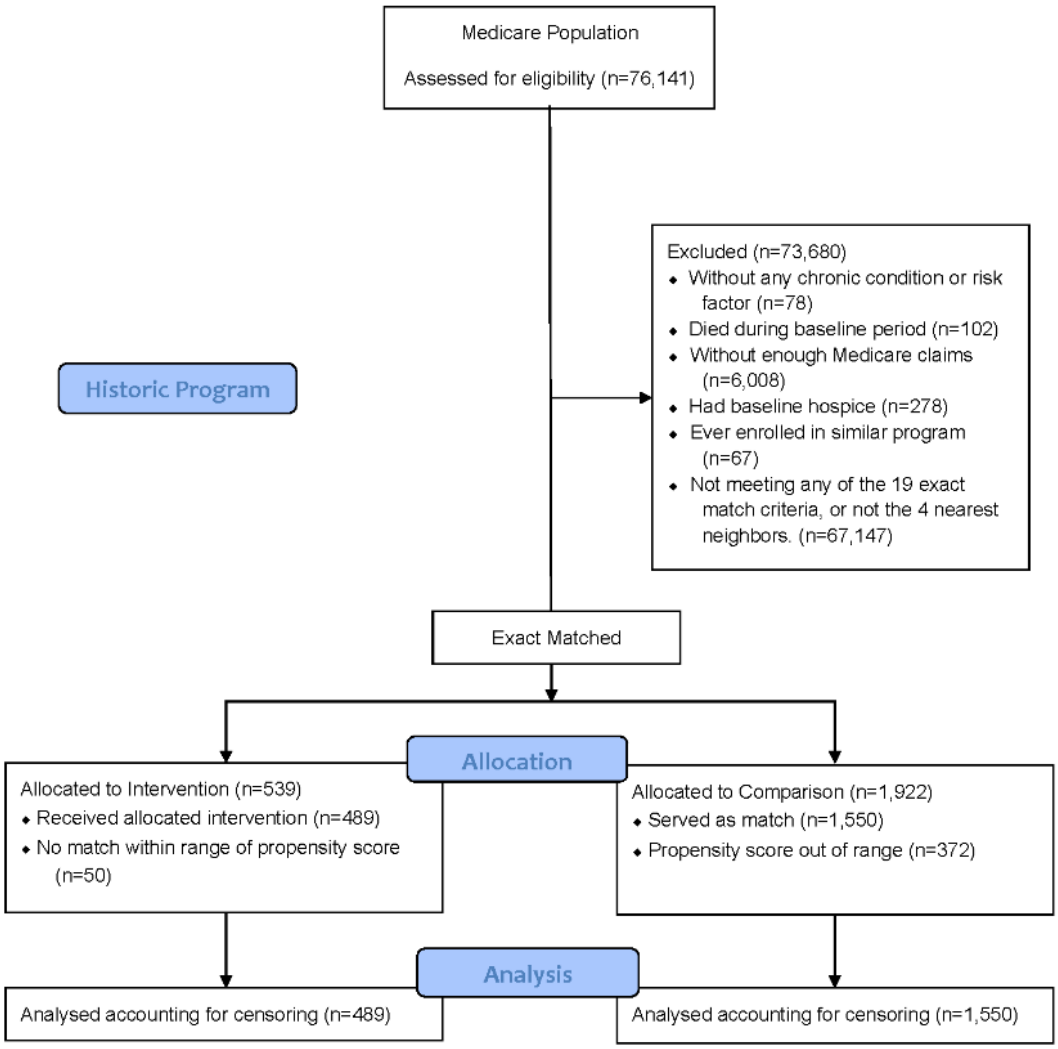
Case Management Patients

For ease of comparison, we described patients from both the historic [16] and new case management programs. We included patients aged 18 years or older enrolled for at least 1 month with (1) continuous EHR and claims data available for at least 1 year prior to enrollment in case management; (2) assignment to the ACO during baseline and follow-up periods; and (3) at least 1 month of continuous EHR and claims data during the follow-up period. Patients were excluded if they were enrolled in hospice, were on dialysis, or had end-stage renal disease during baseline. Patients were recruited for the historic case

XSL•FO
RenderX

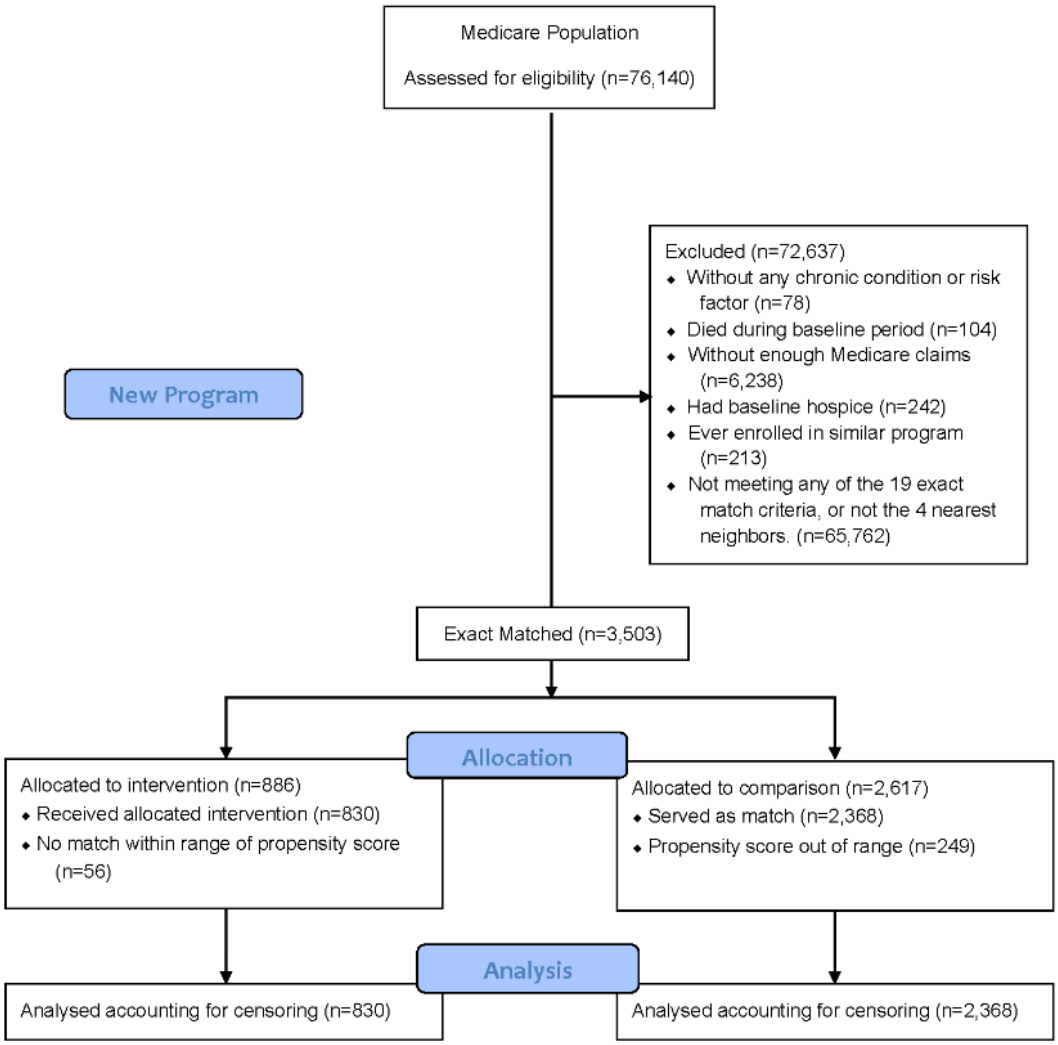
management program from June 2013 to December 2018. For the new program, we also excluded any patients previously enrolled in the historic program. Patients were recruited for the new case management program from April 2018 through March 2019. The final sample size of patients in the historic program was 489 (Figure 1) and in the new program was 830 (Figure 2).

Figure 1. Final analysis sample for the historic case management program.



XSL•FO
RenderX

Figure 2. Final analysis sample for the new case management program.



Matched Comparison Patients

We identified all possible comparison patients receiving usual care from the Medicare ACO, who had not been enrolled in the program but who had comparable patient characteristics, had data available, and met the inclusion and exclusion criteria. For each possible comparison patient, we constructed multiple baseline 1-year time periods (63,047 potential comparison patients; 732,799 potential comparison patient-episodes). We matched each case to a maximum of four of the closest eligible comparison patient-episodes. The date at which a possible comparison patient-episode had the closest match to a case with respect to baseline characteristics was the "match date" and was treated identically to the case's "enrollment date." The final sample size of matched comparison patient-episodes was 1550 for the historic program (Figure 1) and 2368 for the new program (Figure 2).

Potential Input Variables

Our strategy leveraged the high-dimensional nature of combined EHR and claims data [24]. For the baseline year for each patient, we constructed 18,406 possible input variables that encompassed sociodemographics (eg, demographics and homelessness), chronic conditions (eg, diagnoses), utilization (eg, procedures and hospitalizations), vital signs (eg, blood pressure), behaviors (eg, tobacco), laboratory values, payments, medications, patient engagement (eg, "no show" appointments), and other information (eg, advance directives). For missing information for continuous variables, we used simple mean imputation within each decile of a hierarchical condition category (HCC) score [25], and for categorical variables, a missing category was created [26]. Continuous variables were transformed into indicators representing "high" and "low" values using the median from the cases. For our core set of descriptive characteristics presented in tables, we used claims data unless otherwise specified. Baseline sociodemographic variables included age (continuous), sex (female/male), race/ethnicity (White/non-White or Hispanic), Medicaid (yes/no), disability entitlement (yes/no), residence (urban, suburban, large town, and small town/rural; categorizing ZIP code from claims) [27], and mean percentage with a high school degree in the 2007-2011 census tract (after geocoding the address from the EHR). Other variables included the HCC score (continuous) [25] and a risk score based on both claims and EHR data that predicted the risk of hospital admission or death within the next 6 months (categorized as "high" risk if the risk was >13%; otherwise, designated as "moderate") [28]. Chronic conditions included 17 medical conditions defined by Elixhauser et al using

XSL•FO
RenderX

International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnostic codes, along with an indicator variable for ≥3 of these conditions [29]. Utilization included counts of emergency department (ED) visits, unplanned hospitalizations and hospital days [30], observation stays and observation days, and total Medicare payments. ED visits that resulted in hospitalization were not counted as ED visits but were counted as part of the hospitalization.

Exact and Propensity Score Matching

To conduct matching, we constructed a high-dimensional propensity score for case management enrollment by adapting the approach from Schneeweiss et al [24]. This included (1) requiring the variables to have a prevalence between 5% and 95% among the cases and a maximum correlation of 0.8 for each covariate (14,909 variables remained); (2) prioritizing covariates using a measure of confounding bias (threshold=95% significance level; 1905 variables remained); (3) selecting covariates using logistic regression with a lasso penalty, with tuning parameters selected using a variant of the traditional stepwise selection, where the final model was chosen on the basis of the best Schwarz Bayesian criterion (37 variables remained) [31]; (4) estimating the propensity score using logistic regression and the 37 predictors, including chronic conditions, HCC scores, procedures, medication counts, telephone encounter counts, etc, for each patient-episode; (5) selecting up to four of the closest eligible comparison patient-episodes using 5 rounds of exact matching (19 exact match variables) and within exact match strata; and (6) selecting final matches using global optimal propensity score matching to minimize the overall distance between propensity scores, using a matrix of distances between all cases and potential matches [32,33]. The quality of our matching process was determined by examining standardized mean differences, which describe a variance-normalized difference in the means of confounders of the control group and the group enrolled in case management. Standardized mean differences with values around 20%-25% were considered moderately imbalanced, but with a range that was amenable to further adjustment through regression [33,34]. Of 1905 baseline variables, 4 had standardized differences between cases and comparison patients above 25%, including the count of unique prescription medication, nonthrombotic nonathlerosclerotic vascular disease or hypertensive heart disease, and professional service payment, and were included in regressions to adjust for residual confounding [33].

Outcome Measures

Our outcome measures were (1) any unplanned hospital events (admissions, observation stays, and ED visits) during a month, (2) the count of days during the month with any unplanned hospital events, and (3) total Medicare payments during a month, excluding payments for planned hospitalizations and pharmacy payments. We created a data set with 1 observation per patient-episode per month. The first month was 12 months prior to the enrollment/match date and continued for 1 to 12 months after the enrollment/match date until death or censoring due to lack of data.

Benefit Score

The benefit score [16] differs from a typical risk score in that it predicts the effectiveness or "benefit" of a treatment with respect to the outcome using patient and clinical characteristics (eg, the effectiveness of case management with respect to reducing payments), rather than predicting the outcome directly (eg, payments). This modeling approach was developed under a PCORI methodology grant (ME-1409-21219; "Developing new methods for determining treatment benefits based on individual patient traits" [35,36]). The benefit score represents the estimated reduction in Medicare payments within 1 year if the patient is enrolled in case management [16]. Important variables that determined the benefit from case management included chronic conditions (liver disease, dementia, cardiac dysrhythmias, psychiatric disease, and back disease), count of medication, count of appointment "no-shows," and use of the electronic medical record patient portal. Patients with negative savings have "no benefit" from case management, and those with positive savings have "benefit." To provide a qualitative summary of the benefit score, we divided the score into quintiles above 0 (1 to 5) and below zero (−5 to −1). As values close to 0 were ambiguous, scores from 2 to 5 were designated "high benefit" and scores from −5 to 1 were designated "no/low benefit."

Statistical Analysis

To estimate the effect of our intervention, our statistical analysis used longitudinal regression modeling of the risk-adjusted difference in outcome trajectories between the cases and comparison patients, using patient-month data. We used an intent-to-treat approach in which individuals who disenrolled from the program were treated as enrolled. We controlled for confounding using exact and propensity score matching (see "Exact and Propensity Score Matching"). After matching, our regression modeling accounted for residual confounding using inverse weighting by the propensity score and for differences in the number of matched comparison patients for each case (ranging from 1-4) by weighting using the inverse of the number of matches. We used the following link functions: logit/binomial (any events), log/zero inflated Poisson (count of event-days), and log/zero inflated gamma (payments). Models included terms for the preintervention trend, change in level, and postintervention trend in monthly events for both cases and comparison patients, and were risk-adjusted for 4 indicator variables with standardized differences above 25% (see above). We stratified our regression analyses of the new case management program by high versus moderate risk, and the final model included benefit category as an interaction term. Treatment of missing data is described in the section on input variables. Results were transformed into predicted outcomes (ie, dollar amount of Medicare payment reduction, number of event-months prevented, or number of event-days prevented) for 100 patients enrolled in case management programs for 1 year, who were similar to those included in our analyses [37]. Because the benefit score was developed on 69% of the cases in the historic program (339 patients enrolled prior to December 1, 2016) [16], intervention effect estimates for the historic program may be biased. We debiased the intervention effect estimation for the historic program using a Harrell bootstrap

XSL•FO
RenderX

bias-correction procedure [38], but found no difference after correction and thus presented uncorrected estimates; this procedure is not needed for the new program. We calculated bootstrapped 95% CIs using 400 replications for all outcome models.

Results

Characterizing Case Management Programs

The historic case management program used a team approach with a centrally located registered nurse and social worker assigned to each patient and enrolled mostly high-risk patients (Table 1). At program initiation in 2013, patients were identified for further screening using a risk score [28] (calculated monthly) that represented risk of hospital admission or death within 6 months (with "high risk" defined as >13%) or through referral by their primary care provider. After initial identification, patients were screened by nurses or social workers using an assessment tool [39]. Beginning in 2017, the benefit score was also used to identify patients for further screening (with high benefit defined as greater than US $1200 estimated reduction in Medicare payments) [16,40].

The new program relied on nurses physically located in each primary care clinic. Social workers were available only through referral and, in practice, consulted infrequently. The program enrolled both high- and moderate-risk patients. At program initiation, the health system decided to identify 80% of patients through the monthly benefit and risk scoring process developed for the historic program and 20% through the primary care provider referral process [22].

Table 1. Case management program characteristics.

Characteristic	Historic program	New program
Established (year)	2013	2018
Enrolled patients, n	550	3600
Case manager		
Profession	RN[a]+SW[b] dyad	RN
Training	CCMC[c]	In-house
Number	15	40
Case finding process		
Risk score	Yes	Yes
Referrals from providers	Yes	Yes (20% of cases)
Benefit score	Beginning 2017	Yes (80% of cases)
Intake process		
Comprehensive assessment	Within 30 days	Within 30 days
Collaborative goal-setting	Within 60 days	Within 60 days
Intervention intensity		
Contacts (#)	3 per month	1 per month
Duration, mean	160 days	200 days
Caseload	20-25 primary; 40-50 secondary	75 primary
Care integration		
Physician collaboration	Yes	Yes
Mode of contact	Telephone and in-person	Telephone and in-person
Program location	Central	Primary care clinic
Quality assurance		
Medical director attends weekly case consults	Yes	No
Medical director chart review	Yes	Yes

[a]RN: registered nurse.

[b]SW: social worker.

[c]CCMC: Commission for Case Manager Certification.

XSL•FO
RenderX

Characterizing Case Management and Comparison Patients

After matching and propensity score weighting, case management and comparison patients were similar with respect to a predetermined set of baseline sociodemographic, chronic condition [29], behavioral, and utilization variables, although cases had slightly more anxiety than comparison patients (Table 2). However, patients in the historic and new programs differed. Patients in the new program were older but less likely to live in an urban area, have Medicaid, or have disability entitlement. They also had less alcohol or drug abuse, less depression but more hypertension and diabetes with complications, lower HCC scores, and less baseline utilization, and were less likely to be high risk. Specifically, 70% of cases in the historic program were high risk compared with 58% of cases in the new program, and the median risk score for cases in the historic program was twice as high as that for cases in the new program (32% vs 16%; data not shown).

Because of these differences, we stratified cases in the new program by high risk versus moderate risk (Table 3). High-risk patients in the new program had similar or slightly higher HCC scores compared with the scores of high-risk patients in the historic program, but were older, less likely to be on Medicaid, more suburban, and more likely to have 3 or more chronic conditions. Moderate-risk patients in the new program had lower HCC scores compared with the scores of high-risk patients and were less likely to have chronic conditions but were more likely to have anxiety and depression.

XSL•FO
RenderX

Table 2. Sociodemographics, chronic conditions, and baseline utilization for historic and new case management and matched comparison patients.

Characteristic	Historic program[a]		New program[a]	
	Cases (N=489)	Comparisons (N=1550)	Cases (N=830)	Comparisons (N=2368)
Sociodemographics				
Age (years), mean (SD)	67 (20)	68 (13)	76 (11)	76 (12)
Female, n (%)	320 (65)	987 (64)	542 (65)	1564 (66)
Non-White or Hispanic, n (%)	66 (13)	121 (8)	42 (5)	128 (5)
Medicaid insurance ever, n (%)	219 (45)	643 (42)	184 (22)	594 (25)
Disability entitlement, n (%)	218 (44)	611 (39)	144 (17)	408 (17)
Rural/urban, n (%)				
Urban code	387 (79)	1031 (67)	508 (61)	1609 (68)
Suburban	56 (12)	284 (18)	180 (22)	425 (18)
Large town	35 (7)	186 (12)	128 (15)	295 (12)
Small town/rural	10 (2)	45 (3)	14 (2)	38 (2)
Percentage with a high school degree, mean (SD)	1 (0)	1 (0)	1 (0)	1 (0)
HCC[b] score, mean (SD)	3 (3)	3 (2)	2 (1)	2 (1)
High risk, n (%)	344 (70)	997 (64)	483 (58)	1341 (57)
Chronic conditions, n (%)				
3 or more chronic conditions	420 (86)	1322 (85)	757 (91)	2045 (86)
Congestive heart failure	203 (42)	650 (42)	322 (39)	862 (36)
COPD[c]/asthma	224 (46)	709 (46)	375 (45)	906 (38)
Chronic kidney disease	220 (45)	728 (47)	346 (42)	964 (41)
Alcohol or drug abuse	159 (33)	328 (21)	133 (16)	370 (16)
Anxiety	274 (56)	671 (43)	393 (47)	825 (35)
Depression	234 (48)	699 (45)	304 (37)	830 (35)
Diabetes with complications	125 (26)	396 (26)	188 (23)	479 (20)
Diabetes without complications	40 (8)	138 (9)	84 (10)	246 (10)
Hypertension	296 (60)	1044 (67)	676 (81)	1642 (69)
Liver disease	45 (9)	112 (7)	59 (7)	84 (4)
Fluid/electrolyte disorders	227 (46)	681 (44)	339 (41)	867 (37)
Metastatic cancer	19 (4)	37 (2)	38 (5)	77 (3)
Obesity	154 (31)	430 (28)	262 (32)	570 (24)
Psychosis	143 (29)	424 (27)	179 (22)	436 (18)
Peripheral vascular disease	115 (23)	396 (26)	234 (28)	667 (28)
Renal failure	129 (26)	470 (30)	209 (25)	613 (26)
Solid tumor without metastasis	34 (7)	127 (8)	87 (10)	258 (11)
Baseline utilization				
Number of ED[d] visits, mean (SD)	2.2 (4.97)	2.09 (2.89)	1.07 (1.43)	0.96 (1.44)
Number of hospitalizations, mean (SD)	1.17 (2.46)	1.11 (1.20)	0.64 (0.93)	0.54 (0.84)
Number of days in hospital, mean (SD)	5.55 (14.98)	5.30 (7.07)	2.77 (5.11)	2.33 (4.57)
Number of observation stays, mean (SD)	0.25 (0.83)	0.25 (0.49)	0.17 (0.42)	0.15 (0.41)
Number of days in observation stay, mean (SD)	0.29 (1.08)	0.33 (0.74)	0.21 (0.57)	0.18 (0.53)

XSL•FO
RenderX

Characteristic	Historic program[a]		New program[a]	
	Cases (N=489)	Comparisons (N=1550)	Cases (N=830)	Comparisons (N=2368)
Medicare payment[e], mean (SD)	29.43 (54.14)	30.57 (32.50)	21.05 (25.86)	17.37 (22.62)

[a]Numbers were adjusted for varying case control ratios.

[b]HCC: hierarchical condition category.

[c]COPD: chronic obstructive pulmonary disease.

[d]ED: emergency department.

[e]Per US $1000.

XSL•FO
RenderX

Table 3. Sociodemographics, chronic conditions, and baseline utilization for new case management and matched comparison patients, by risk.

Characteristic	New program, high risk[a]		New program, moderate risk[a]	
	Cases (N=483)	Comparisons (N=1315)	Cases (N=347)	Comparisons (N=1053)
Sociodemographics				
Age (years), mean (SD)	79 (11)	78 (11)	71 (11)	75 (12)
Female, n (%)	308 (64)	827 (63)	234 (67)	727 (69)
Non-Hispanic White, n (%)	462 (96)	1247 (95)	326 (94)	989 (94)
Non-White or Hispanic, n (%)	21 (4)	68 (5)	21 (6)	64 (6)
Medicaid insurance ever, n (%)	127 (26)	350 (27)	57 (16)	241 (23)
Disability entitlement, n (%)	79 (16)	204 (16)	65 (19)	196 (19)
Rural/urban, n (%)				
Urban code	280 (58)	909 (69)	228 (66)	699 (66)
Suburban	111 (23)	213 (16)	69 (20)	211 (20)
Large town	82 (17)	170 (13)	46 (13)	128 (12)
Small town/rural	10 (2)	22 (2)	4 (1)	15 (1)
Percentage with a high school degree, mean (SD)	1 (0)	1 (0)	1 (0)	1 (0)
HCC[b] score, mean (SD)	3 (1)	3 (1)	2 (1)	2 (1)
High risk, n (%)	483 (100)	991 (75)	0 (0)	363 (34)
Chronic conditions				
3 or more chronic conditions	472 (98)	1241 (94)	285 (82)	810 (77)
Congestive heart failure	249 (52)	651 (50)	73 (21)	219 (21)
COPD[c]/asthma	242 (50)	558 (42)	133 (38)	349 (33)
Chronic kidney disease	268 (55)	665 (51)	78 (22)	303 (29)
Alcohol or drug abuse	78 (16)	207 (16)	55 (16)	158 (15)
Anxiety	213 (44)	422 (32)	180 (52)	394 (37)
Depression	165 (34)	419 (32)	139 (40)	405 (38)
Diabetes with complications	134 (28)	337 (26)	54 (16)	146 (14)
Diabetes without complications	38 (8)	129 (10)	46 (13)	121 (11)
Hypertension	406 (84)	969 (74)	270 (78)	677 (64)
Liver disease	34 (7)	55 (4)	25 (7)	27 (3)
Fluid/electrolyte disorders	277 (57)	582 (44)	62 (18)	289 (27)
Metastatic cancer	29 (6)	60 (5)	9 (3)	17 (2)
Obesity	142 (29)	329 (25)	120 (35)	238 (23)
Psychosis	85 (18)	230 (17)	94 (27)	203 (19)
Peripheral vascular disease	184 (38)	440 (33)	50 (14)	235 (22)
Renal failure	166 (34)	431 (33)	43 (12)	188 (18)
Solid tumor without metastasis	53 (11)	160 (12)	34 (10)	100 (10)
Baseline utilization				
Number of ED[d] visits, mean (SD)	1.31 (1.60)	1.15 (1.62)	0.73 (1.07)	0.72 (1.11)
Number of hospitalizations, mean (SD)	0.93 (1.06)	0.82 (0.96)	0.24 (0.50)	0.22 (0.51)
Number of days in hospital, mean (SD)	4.18 (6.11)	3.65 (5.52)	0.80 (1.99)	0.81 (2.39)
Number of observation stays, mean (SD)	0.22 (0.46)	0.17 (0.44)	0.10 (0.33)	0.13 (0.36)

XSL•FO
RenderX

Characteristic	New program, high risk[a]		New program, moderate risk[a]	
	Cases (N=483)	Comparisons (N=1315)	Cases (N=347)	Comparisons (N=1053)
Number of days in observation stay, mean (SD)	0.28 (0.67)	0.21 (0.59)	0.10 (0.35)	0.15 (0.43)
Medicare Payment[e], mean (SD)	28.43 (28.39)	24.27 (26.33)	10.78 (17.24)	9.37 (13.26)

[a]Numbers were adjusted for varying case control ratios.

[b]HCC: hierarchical condition category.

[c]COPD: chronic obstructive pulmonary disease.

[d]ED: emergency department.

[e]Per US $1000.

Characterizing Case Management Patients by Benefit Category

Approximately one-third of the cases in the historic program were identified as high benefit, while in the new program, 43% of high-risk and 37% of moderate-risk cases were identified as high benefit (Table 4). High-benefit patients in the historic program had higher HCC scores and baseline utilization but were less likely to be high risk and had less disability entitlement when compared with the findings for no/low-benefit patients in the historic program. They were also less likely to have chronic conditions, including chronic obstructive pulmonary disease (COPD)/asthma and anxiety, but more likely to have diabetes with complications. Among high-risk patients in the new program, high-benefit patients were more likely to be female and less likely to have congestive heart failure, chronic kidney disease, alcohol or drug abuse, or valvular disease when compared with the findings for no/low-benefit patients. Among moderate-risk patients in the new program, high-benefit patients were slightly more likely to be female and were less likely to have Medicaid or disability entitlement, COPD/asthma, and alcohol or drug abuse, but more likely to have higher HCC scores and 3 or more chronic conditions, including obesity.

XSL•FO
RenderX

Table 4. Sociodemographics, chronic conditions, and baseline utilization for historic and new case management patients, by benefit category.

Characteristic	Historic program (cases only)		New program (cases only), high risk		New program (cases only), moderate risk	
	High benefit (N=170)	No/low benefit (N=319)	High benefit (N=209)	No/low benefit (N=274)	High benefit (N=129)	No/low benefit (N=218)
Sociodemographics						
Age (years), mean (SD)	64 (15)	68 (14)	79 (11)	78 (11)	73 (9)	71 (11)
Female, n (%)	107 (63)	211 (66)	147 (70)	161 (59)	91 (71)	143 (66)
Non-White or Hispanic, n (%)	23 (14)	39 (12)	10 (5)	11 (4)	10 (8)	11 (5)
Medicaid insurance ever, n (%)	76 (45)	15 (48)	54 (26)	73 (27)	14 (11)	43 (20)
Disability entitlement, n (%)	69 (40)	18 (56)	37 (18)	42 (15)	21 (16)	44 (20)
Rural/urban, n (%)						
Urban code	137 (80)	241 (75)	117 (56)	163 (59)	85 (66)	143 (66)
Suburban	19 (11)	45 (14)	47 (22)	64 (23)	27 (21)	42 (19)
Large town	11 (6)	27 (9)	43 (21)	39 (14)	17 (13)	29 (13)
Small town/rural	3 (2)	6 (2)	2 (1)	8 (3)	0 (0)	4 (2)
Percentage with a high school degree, mean (SD)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)
HCC[a] score, mean (SD)	3 (2)	3 (2)	3 (1)	3 (2)	2 (1)	1 (1)
High risk, n (%)	117 (69)	256 (80)	209 (100)	274 (100)	0 (0)	0 (0)
Chronic conditions						
3 or more chronic conditions	142 (84)	301 (94)	204 (98)	268 (98)	117 (91)	168 (77)
Congestive heart failure	74 (44)	141 (44)	102 (49)	147 (54)	29 (22)	44 (20)
COPD[b]/asthma	73 (43)	184 (58)	106 (51)	136 (50)	40 (31)	93 (43)
Chronic kidney disease	79 (47)	155 (48)	103 (49)	165 (60)	28 (22)	50 (23)
Alcohol or drug abuse	55 (33)	115 (36)	27 (13)	51 (19)	8 (6)	47 (22)
Anxiety	88 (52)	207 (65)	86 (41)	127 (46)	70 (54)	110 (50)
Depression	79 (46)	168 (53)	71 (34)	94 (34)	50 (39)	89 (41)
Diabetes with complications	43 (25)	98 (31)	51 (24)	83 (30)	20 (16)	34 (16)
Diabetes without complications	17 (10)	14 (4)	16 (8)	22 (8)	14 (11)	32 (15)
Hypertension	104 (61)	198 (62)	174 (83)	232 (85)	106 (82)	164 (75)
Liver disease	9 (6)	57 (18)	20 (10)	14 (5)	11 (9)	14 (6)
Fluid/electrolyte disorders	74 (44)	182 (57)	124 (59)	153 (56)	24 (19)	38 (17)
Metastatic cancer	8 (5)	12 (4)	12 (6)	17 (6)	4 (3)	5 (2)
Obesity	55 (33)	104 (33)	64 (31)	78 (28)	54 (42)	66 (30)
Psychosis	46 (27)	108 (34)	37 (18)	48 (18)	29 (22)	65 (30)
Peripheral vascular disease	39 (23)	88 (28)	74 (35)	110 (40)	17 (13)	33 (15)
Renal failure	48 (28)	84 (26)	59 (28)	107 (39)	16 (12)	27 (12)
Solid tumor without metastasis	13 (8)	23 (7)	22 (11)	31 (11)	12 (9)	22 (10)
Baseline utilization						
Number of ED[c] visits, mean (SD)	3.04 (4.20)	2.08 (3.30)	1.33 (1.61)	1.30 (1.60)	0.73 (1.07)	0.72 (1.07)
Number of hospitalizations, mean (SD)	1.70 (2.13)	1.10 (1.62)	0.88 (0.90)	0.97 (1.17)	0.39 (0.62)	0.15 (0.39)
Number of days in hospital, mean (SD)	8.96 (14.77)	4.79 (8.36)	4.14 (5.44)	4.21 (6.58)	1.27 (2.48)	0.53 (1.58)

XSL•FO
RenderX

Characteristic	Historic program (cases only)		New program (cases only), high risk		New program (cases only), moderate risk	
	High benefit (N=170)	No/low benefit (N=319)	High benefit (N=209)	No/low benefit (N=274)	High benefit (N=129)	No/low benefit (N=218)
Number of observation stays, mean (SD)	0.33 (0.66)	0.25 (0.60)	0.21 (0.45)	0.23 (0.47)	0.15 (0.42)	0.07 (0.26)
Number of days in observation stay, mean (SD)	0.37 (0.77)	0.29 (0.80)	0.25 (0.62)	0.30 (0.72)	0.16 (0.46)	0.07 (0.26)
Medicare payment[d], mean (SD)	44.39 (52.84)	27.20 (33.55)	27.62 (26.12)	29.05 (30.04)	15.66 (24.34)	7.89 (10.09)

[a]HCC: hierarchical condition category.

[b]COPD: chronic obstructive pulmonary disease.

[c]ED: emergency department.

[d]Per US $1000.

Relationship Between Case Management and Outcomes by Benefit Category

Across all patients, enrollment in the historic case management program was associated with 80 fewer events and 368 fewer event-days per 100 enrolled patients, although there was no difference in Medicare payments (Table 5). Among high-benefit patients, enrollment in the historic program was associated with 117 fewer events, 536 fewer event-days, and US $1,151,063 reduction in Medicare payments over the subsequent year per 100 enrolled patients when compared with the findings for comparison patients. Among no/low-benefit patients, there was no association between enrollment and outcomes.

For the new case management program, among high-risk high-benefit patients, enrollment was associated with 65 fewer events per 100 patients, with no difference in event-days or Medicare payments. Among high-risk no/low-benefit patients, there was no association between enrollment and outcomes. Among moderate-risk patients, there was no association between enrollment and outcomes for either high-benefit or no/low-benefit patients.

XSL•FO
RenderX

Table 5. Average adjusted predicted outcomes per 100 patients enrolled in case management for 1 year among historic and new case management patients, by benefit category.

Outcome	Overall		High benefit		No/low benefit	
	Value	95% CI	Value	95% CI	Value	95% CI
Historic case management program						
Reduction in the number of unplanned events	80	18 to 161	117	51 to 202	9	−65 to 97
Reduction in the number of unplanned event-days	368	32 to 820	536	214 to 962	44	−357 to 522
Medicare savings (from the outcome model) (US$)	663,742	−204,900 to 1,827,605	1,151,063	368,423 to 2,216,791	−272,124	−1,449,624 to 1,179,157
New case management program (all patients)						
Reduction in the number of unplanned events	0	−43 to 44	17	−35 to 61	−12	−57 to 41
Reduction in the number of unplanned event-days	49	−136 to 245	155	−58 to 365	−24	−248 to 203
Medicare savings (US$)	−33,840	−828,888 to 740,516	181,660	−741,324 to 1,059,818	−181,032	−964,032 to 658,798
New case management program (high-risk patients)						
Reduction in the number of unplanned events	34	−32 to 102	65	1 to 128	11	−64 to 100
Reduction in the number of unplanned event-days	132	−158 to 447	288	−29 to 596	13	−333 to 411
Medicare savings (US$)	350,407	−714,876 to 1,526,562	891,572	−291,096 to 1,976,962	−58,524	−1,193,112 to 1,263,905
New case management program (moderate risk patients)						
Reduction in the number of unplanned events	−57	−114 to 0.30	−72	−156 to 0.01	−47	−112 to 8
Reduction in the number of unplanned event-days	−190	−486 to 27	−191	−527 to 62	−189	−549 to 52
Medicare savings (US$)	−288,552	−1,447,932 to 860,098	−449,688	−2,248,944 to 1,022,261	−182,856	−1,365,516 to 827,178

Discussion

In this Medicare ACO, we found that reduction in Medicare payments and unplanned hospital events from case management participation were limited to high-risk high-benefit patients. A benefit score [16] was able to identify patients who would benefit from a new program with respect to reducing events, but only among a high-risk population with average HCC scores similar to the population on which the score was developed. The score was not able to successfully identify moderate-risk patients who might benefit.

There are several possible reasons for our findings on applying a previously developed benefit score prospectively to a new case management program and population. A possible explanation for why the score was able to successfully identify high-risk patients who might benefit from a new (and different) program is that while the historic and new case management programs differed in teams of composition and location, core elements of a program may depend more on what is done and not who does it or where it is done. When both nurses and social workers work in a practice, they tend toward different roles (social workers assess social issues and nurses coordinate hospital transitions) [41], but in solo practice, each may provide all essential elements of case management [42]. Conversely, the benefit score was developed in a high-risk population [16], not in the broader high- and moderate-risk population served by the new program. Even though targeting patients for case management using a risk score alone may be insufficient [5,43], case management programs may still be designed to optimize care for patients at a specific risk level, and enrollment of patients with different risks may mean a mismatch between program goals/activities and patient needs [44]. Our findings suggest that this score may be limited to identifying case management patients who would benefit from a new case management program only among a population similar to that on which the score was developed.

The latest projections from the Congressional Budget Office are that the Medicare trust fund will run out of money in 2024 [45]. This is the closest the fund has ever come to insolvency since Medicare was established in 1965 and demonstrates the urgent need to understand how to best provide access to high-quality care while simultaneously controlling costs. In order for impactability modeling to help solve the nationwide crisis in caring for high-need high-cost patients [15], “benefit”

XSL•FO
RenderX

or “impactability” scores will need to extend beyond the programs and populations in which they were developed. This study provides a first step toward assessing the feasibility and limits of this extension. Although we found evidence that a benefit score could extend to a new program and a similar risk population, caution is warranted as programs vary widely [5,8] and evidence of successful extension to one program does not necessarily indicate that the score could be extended to another program. Unlike risk models, impactability models are intrinsically linked to both the population and the specific program used in their development. Measurement of “similarity” (how similar is similar enough?) is an important open question [46]. More research is needed to understand the core elements of case management (to identify similar programs) and to streamline identification of similar populations.

There are several limitations to our study. First, we were limited to evaluating the impact of the pandemic in a single large health system with both academic and community clinics. This health system did participate in Medicare ACO programs, indicating that they had a strong base of primary care patients [47]. Examining different programs within a system may mitigate variability in coding and data across systems [48], but could also complicate extension to another system. Moreover, academic systems often serve a different population than community practices, but this health system had a large number of community-based primary care clinics [47]. Second, unmeasured confounding is a limitation of all observational studies. As is the case with any observational study, it is almost never possible to know the direction and magnitude of such unmeasured confounding. However, given our measurement of repeated outcomes both prepandemic and postpandemic, as well as an extensive matching process and the similarity of our matched populations, it is unlikely that any remaining small differences explain our findings. Third, we only followed outcomes for 1 year after enrollment, which may be too short to realize positive outcomes [49]. This may explain our finding in the moderate-risk population, and similar to another study [50], negative findings at 1 year might turn into positive findings at 2 years. Fourth, the study focused only on mortality and unplanned events such as hospitalizations and ED visits. Although we used validated algorithms, the definition of unplanned events likely represented some unavoidable events that may not be directly under the control of the health system. This may have slightly reduced our ability to estimate the impact of the case management programs.

The use of impactability modeling to match specific case management programs with high-need high-cost patients who might benefit is consistent with the call by Bates et al to make predictions actionable for interventions [12,51]. This approach does not rely on identifying effective case management programs and attempting to standardize their implementation nationwide, a daunting undertaking given the wide variation in programs [5,8], resistance to change within health systems [52], and practical challenges in implementing evidence-based interventions [53]. Yet, impactability modeling brings its own challenges, most importantly the limitation of tailoring each model to a specific case management program and population. Enthusiasm for this approach should be tempered until additional research provides robust strategies to identify case management programs and populations that are sufficiently similar to warrant a score’s application. In the interim, extending a score developed for a specific program and population to a different program and population should be accompanied by ongoing evaluation to confirm its applicability. Over time, policy makers and measure developers should consider impactability modeling when designing new programs and metrics.

Acknowledgments

This work was supported by the Health Innovation Program at the UW School of Medicine and Public Health. The sponsors had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and the decision to submit the manuscript for publication.

Authors' Contributions

MAS 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. Study concept and design: MAS, MY, and JH. Acquisition, analysis, or interpretation of data: MAS, MY, JH, XW, and AD. Drafting of the manuscript: MAS. Critical revision of the manuscript for important intellectual content: All authors. Statistical analysis: XW. Obtained funding: MAS. Administrative, technical, or material support: MAS. Study supervision: MAS.

Conflicts of Interest

None declared.

References

1. Blumenthal D, Chernof B, Fulmer T, Lumpkin J, Selberg J. Caring for high-need, high-cost patients - An urgent priority. N Engl J Med 2016 Sep 08;375(10):909-911. [doi: 10.1056/NEJMp1608511] [Medline: 27602661]
2. Long P, Abrams M, Milstein A, Anderson G, Apton K, Dahlberg M, et al. Effective Care for High-Need Patients: Opportunities for Improving Outcomes, Value, and Health. National Academy of Medicine. 2017. URL: https://nam.edu/wp-content/uploads/2017/06/Effective-Care-for-High-Need-Patients.pdf [accessed 2021-04-06]

XSL•FO
RenderX

3. Adams AJ, Clark DR, DeLander GE, Mackinnon GE, Malloy M, McGivney MS, et al. Report of the AACP task force on patient-centered medical homes and accountable care organizations. Am J Pharm Educ 2013 Sep 12;77(7):142 [FREE Full text] [doi: 10.5688/ajpe777142] [Medline: 24052645]
4. Fisher ES, Staiger DO, Bynum JPW, Gottlieb DJ. Creating accountable care organizations: the extended hospital medical staff. Health Aff (Millwood) 2007;26(1):w44-w57 [FREE Full text] [doi: 10.1377/hlthaff.26.1.w44] [Medline: 17148490]
5. Hickam DH, Weiss JW, Guise JM, Buckley D, Motu'apuaka M, Graham E, et al. Outpatient Case Management for Adults With Medical Illness and Complex Care Needs. Rockville, MD: Agency for Healthcare Research and Quality; 2013.
6. Stokes J, Panagioti M, Alam R, Checkland K, Cheraghi-Sohi S, Bower P. Effectiveness of case management for 'at risk' patients in primary care: A systematic review and meta-analysis. PLoS One 2015;10(7):e0132340 [FREE Full text] [doi: 10.1371/journal.pone.0132340] [Medline: 26186598]
7. Smith MA, Vaughan-Sarrazin MS, Yu M, Wang X, Nordby PA, Vogeli C, et al. The importance of health insurance claims data in creating learning health systems: evaluating care for high-need high-cost patients using the National Patient-Centered Clinical Research Network (PCORNet). J Am Med Inform Assoc 2019 Nov 01;26(11):1305-1313 [FREE Full text] [doi: 10.1093/jamia/ocz097] [Medline: 31233126]
8. Donelan K, Barreto EA, Michael CU, Nordby P, Smith M, Metlay JP. Variability in care management programs in Medicare ACOs: A survey of medical directors. J Gen Intern Med 2018 Dec;33(12):2043-2045 [FREE Full text] [doi: 10.1007/s11606-018-4609-1] [Medline: 30054887]
9. Nelson L. Lessons from Medicare's Demonstration Projects on Disease Management and Care Coordination. Congressional Budget Office. 2012. URL: https://www.cbo.gov/sites/default/files/cbofiles/attachments/WP2012-01_Nelson_Medicare_DMCC_Demonstrations.pdf [accessed 2021-04-06]
10. McDonald K, Sundaram V, Bravata D, Lewis R, Lin N, Kraft S, et al. Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies. Agency for Healthcare Research and Quality. 2007. URL: https://www.ncbi.nlm.nih.gov/books/NBK44015/pdf/Bookshelf_NBK44015.pdf [accessed 2021-04-06]
11. Blumenthal D, Abrams MK. Tailoring complex care management for high-need, high-cost patients. JAMA 2016 Oct 25;316(16):1657-1658. [doi: 10.1001/jama.2016.12388] [Medline: 27669168]
12. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. Health Aff (Millwood) 2014 Jul;33(7):1123-1131. [doi: 10.1377/hlthaff.2014.0041] [Medline: 25006137]
13. McWilliams JM. Cost containment and the tale of care coordination. N Engl J Med 2016 Dec 08;375(23):2218-2220 [FREE Full text] [doi: 10.1056/NEJMp1610821] [Medline: 27959672]
14. Meek JA. Affordable Care Act: predictive modeling challenges and opportunities for case management. Prof Case Manag 2012;17(1):15-21; quiz 22. [doi: 10.1097/NCM.0b013e318234e7dd] [Medline: 22146637]
15. Lewis GH. "Impactibility models": identifying the subgroup of high-risk patients most amenable to hospital-avoidance programs. Milbank Q 2010 Jun;88(2):240-255 [FREE Full text] [doi: 10.1111/j.1468-0009.2010.00597.x] [Medline: 20579284]
16. Huling JD, Yu M, Smith M. Fused comparative intervention scoring for heterogeneity of longitudinal intervention effects. Ann. Appl. Stat 2019 Jun 1;13(2):1639-1641. [doi: 10.1214/18-AOAS1216]
17. DuBard CA, Jackson CT. Active redesign of a Medicaid care management strategy for greater return on investment: Predicting impactability. Popul Health Manag 2018 Apr;21(2):102-109 [FREE Full text] [doi: 10.1089/pop.2017.0122] [Medline: 28968176]
18. Smith M, Donelan K, Vaughan-Sarrazin M, Yu M, Vogeli C, Wang X, et al. Examining the Effectiveness of Case Management Programs on Preventing Hospital Stays in Older Adults with Multiple Chronic Health Problems -- A PCORnet® Study. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/sites/default/files/Smith023-Final-Research-Report.pdf [accessed 2021-04-06]
19. Goldberg HI, Neighbor WE, Cheadle AD, Ramsey SD, Diehr P, Gore E. A controlled time-series trial of clinical reminders: using computerized firm systems to make quality improvement research a routine part of mainstream practice. Health Serv Res 2000 Mar;34(7):1519-1534 [FREE Full text] [Medline: 10737451]
20. Eccles M, Grimshaw J, Campbell M, Ramsay C. Research designs for studies evaluating the effectiveness of change and improvement strategies. Qual Saf Health Care 2003 Feb;12(1):47-52 [FREE Full text] [doi: 10.1136/qhc.12.1.47] [Medline: 12571345]
21. Lancaster GA, Dodd S, Williamson PR. Design and analysis of pilot studies: recommendations for good practice. J Eval Clin Pract 2004 May;10(2):307-312. [doi: 10.1111/j..2002.384.doc.x] [Medline: 15189396]
22. Smith MA, Nordby PA, Yu M, Jaffery J. A practical model for research with learning health systems: Building and implementing effective complex case management. Appl Ergon 2020 Apr;84:103023. [doi: 10.1016/j.apergo.2019.103023] [Medline: 31983393]
23. IRB QI Self-Certification Decision Tool. HIPxChange. URL: https://www.hipxchange.org/IRB [accessed 2022-06-03]
24. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. Epidemiology 2009 Jul;20(4):512-522 [FREE Full text] [doi: 10.1097/EDE.0b013e3181a663cc] [Medline: 19487948]

XSL•FO
RenderX

25. Pope GC, Kautter J, Ellis RP, Ash AS, Ayanian JZ, Lezzoni LI, et al. Risk adjustment of Medicare capitation payments using the CMS-HCC model. Health Care Financ Rev 2004;25(4):119-141 [FREE Full text] [Medline: 15493448]
26. Lee C, Luo Z, Ngiam K, Zhang M, Zheng K, Chen G, et al. Big Healthcare Data Analytics: Challenges and Applications. In: Khan S, Zomaya A, Abbas A, editors. Handbook of Large-Scale Distributed Computing in Smart Healthcare. Scalable Computing and Communications. Cham: Springer; 2017:11-41.
27. Rural-Urban Commuting Area Codes (RUCAs). Rural Health Research Center. URL: http://depts.washington.edu/uwruca/ruca-data.php [accessed 2021-09-08]
28. Huling JD, Yu M, Liang M, Smith M. Risk prediction for heterogeneous populations with application to hospital admission prediction. Biometrics 2018 Jun;74(2):557-565 [FREE Full text] [doi: 10.1111/biom.12769] [Medline: 29073325]
29. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care 1998 Jan;36(1):8-27. [doi: 10.1097/00005650-199801000-00004] [Medline: 9431328]
30. Hospital-Wide (All-Condition) 30 - Day Risk-Standardized Readmission Measure: DRAFT Measure Methodology Report. Centers for Medicare & Medicaid Services. 2011. URL: https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/mms/downloads/mmshospital-wideall-conditionreadmissionrate.pdf [accessed 2022-04-15]
31. Schwarz G. Estimating the dimension of a model. Ann. Statist 1978 Mar 1;6(2):461-464. [doi: 10.1214/aos/1176344136]
32. Rassen JA, Shelat AA, Myers J, Glynn RJ, Rothman KJ, Schneeweiss S. One-to-many propensity score matching in cohort studies. Pharmacoepidemiol Drug Saf 2012 May;21 Suppl 2:69-80. [doi: 10.1002/pds.3263] [Medline: 22552982]
33. Stuart EA. Matching methods for causal inference: A review and a look forward. Stat Sci 2010 Feb 01;25(1):1-21 [FREE Full text] [doi: 10.1214/09-STS313] [Medline: 20871802]
34. Rubin DB. Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation. In: Matched Sampling for Causal Effects. Cambridge, UK: Cambridge University Press; 2006:365-382.
35. Estimating Differential Treatment Effects Using the 'personalized' R Package. HIPxChange. URL: https://www.hipxchange.org/PersonalizedRPackage [accessed 2021-02-05]
36. Yu M, Huling J, Smith M, Shao J. Developing New Methods for Determining Treatment Benefits Based on Individual Patient Traits. Patient-Centered Outcomes Research Institute. 2020. URL: https://www.pcori.org/sites/default/files/Yu308-Final-Research-Report.pdf [accessed 2021-04-06]
37. Greene WH. Econometric Analysis. Upper Saddle River, NJ: Prentice Hall; 1997.
38. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996 Feb 28;15(4):361-387. [doi: 10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4] [Medline: 8668867]
39. Berry LL, Rock BL, Smith Houskamp B, Brueggeman J, Tucker L. Care coordination for patients with complex health profiles in inpatient and outpatient settings. Mayo Clin Proc 2013 Feb;88(2):184-194. [doi: 10.1016/j.mayocp.2012.10.016] [Medline: 23290738]
40. Case Management Benefit Scoring System. HIPxChange. URL: https://www.hipxchange.org/BenefitScore [accessed 2021-02-05]
41. Donelan K, Chang Y, Berrett-Abebe J, Spetz J, Auerbach DI, Norman L, et al. Care management for older adults: The roles of nurses, social workers, and physicians. Health Aff (Millwood) 2019 Jun;38(6):941-949. [doi: 10.1377/hlthaff.2019.00030] [Medline: 31158015]
42. Tahan HA, Campagna V. Case management roles and functions across various settings and professional disciplines. Prof Case Manag 2010;15(5):245-77; quiz 278. [doi: 10.1097/NCM.0b013e3181e94452] [Medline: 20616764]
43. Haime V, Hong C, Mandel L, Mohta N, Iezzoni LI, Ferris TG, et al. Clinician considerations when selecting high-risk patients for care management. Am J Manag Care 2015 Oct 01;21(10):e576-e582 [FREE Full text] [Medline: 26619059]
44. Price-Haywood EG, Petersen H, Burton J, Harden-Barrios J, Adubato M, Roberts M, et al. Outpatient complex case management: Health system-tailored risk stratification taxonomy to identify high-cost, high-need patients. J Gen Intern Med 2018 Nov;33(11):1921-1927 [FREE Full text] [doi: 10.1007/s11606-018-4616-2] [Medline: 30076572]
45. The Outlook for Major Federal Trust Funds: 2020 to 2030. Congressional Budget Office. URL: https://www.cbo.gov/system/files/2020-09/56523-Trust-Funds.pdf [accessed 2021-04-06]
46. Ippolito M. How similar is similar enough? Semantics and Pragmatics 2016;9(Article 6):1-60 [FREE Full text] [doi: 10.3765/sp.9.6]
47. Schulz J, DeCamp M, Berkowitz SA. Medicare Shared Savings Program: public reporting and shared savings distributions. Am J Manag Care 2015 Aug;21(8):546-553 [FREE Full text] [Medline: 26295354]
48. Estiri H, Chan Y, Baldwin L, Jung H, Cole A, Stephens KA. Visualizing anomalies in electronic health record data: The variability explorer tool. AMIA Jt Summits Transl Sci Proc 2015;2015:56-60 [FREE Full text] [Medline: 26306237]
49. Berry-Millett R, Bodenheimer T. Care management of patients with complex health care needs. Synth Proj Res Synth Rep 2009 Dec(19):52372. [Medline: 22052205]
50. Counsell SR, Callahan CM, Clark DO, Tu W, Buttar AB, Stump TE, et al. Geriatric care management for low-income seniors: a randomized controlled trial. JAMA 2007 Dec 12;298(22):2623-2633. [doi: 10.1001/jama.298.22.2623] [Medline: 18073358]

XSL•FO
RenderX

51. Smith MA, Adelaine S, Bednarz L, Patterson BW, Pothof J, Liao F. Predictive solutions in learning health systems: The critical need to systematize implementation of prediction to action to intervention. NEJM Catalyst 2021 Apr 21;2(5):0650. [doi: 10.1056/cat.20.0650]
52. Coiera E. Why system inertia makes health reform so difficult. BMJ 2011 Jun 23;342:d3693. [doi: 10.1136/bmj.d3693] [Medline: 21700652]
53. Dixon-Woods M, McNicol S, Martin G. Ten challenges in improving quality in healthcare: lessons from the Health Foundation's programme evaluations and relevant literature. BMJ Qual Saf 2012 Oct;21(10):876-884 [FREE Full text] [doi: 10.1136/bmjqs-2011-000760] [Medline: 22543475]

Abbreviations

ACO: accountable care organization
COPD: chronic obstructive pulmonary disease
ED: emergency department
EHR: electronic health record
HCC: hierarchical condition category
PCORI: Patient-Centered Outcomes Research Institute

Edited by T Leung; submitted 08.04.21; peer-reviewed by Y Chu, J Walsh; comments to author 09.12.21; revised version received 15.04.22; accepted 25.04.22; published 13.06.22.

Please cite as:
Smith MA, Yu M, Huling JD, Wang X, DeLonay A, Jaffery J
Impactability Modeling for Reducing Medicare Accountable Care Organization Payments and Hospital Events in High-Need High-Cost Patients: Longitudinal Cohort Study
J Med Internet Res 2022;24(6):e29420
URL: https://www.jmir.org/2022/6/e29420
doi:10.2196/29420
PMID:35699983

©Maureen A Smith, Menggang Yu, Jared D Huling, Xinyi Wang, Allie DeLonay, Jonathan Jaffery. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.06.2022. 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 Benefits of Crowdsourcing to Seed and Align an Algorithm in an mHealth Intervention for African American and Hispanic Adults: Survey Study

Neil Jay Sehgal[1,2], MPH, PhD; Shuo Huang[1], MPH; Neil Mason Johnson[2,3], BS; John Dickerson[2,3], MSc, PhD; Devlon Jackson[2,4], MPH, PhD; Cynthia Baur[2], MA, PhD

[1]Department of Health Policy and Management, School of Public Health, University of Maryland, College Park, MD, United States

[2]Horowitz Center for Health Literacy, School of Public Health, University of Maryland, College Park, MD, United States

[3]Department of Computer Science, University of Maryland, College Park, MD, United States

[4]Department of Behavioral and Community Health, School of Public Health, University of Maryland, College Park, MD, United States

Corresponding Author:
Neil Jay Sehgal, MPH, PhD
Department of Health Policy and Management
School of Public Health
University of Maryland
4200 Valley Drive
College Park, MD, 20742
United States
Phone: 1 3014052469
Email: sehgal@umd.edu

Abstract

Background: The lack of publicly available and culturally relevant data sets on African American and bilingual/Spanish-speaking Hispanic adults' disease prevention and health promotion priorities presents a major challenge for researchers and developers who want to create and test personalized tools built on and aligned with those priorities. Personalization depends on prediction and performance data. A recommender system (RecSys) could predict the most culturally and personally relevant preventative health information and serve it to African American and Hispanic users via a novel smartphone app. However, early in a user's experience, a RecSys can face the "cold start problem" of serving untailored and irrelevant content before it learns user preferences. For underserved African American and Hispanic populations, who are consistently being served health content targeted toward the White majority, the cold start problem can become an example of algorithmic bias. To avoid this, a RecSys needs population-appropriate seed data aligned with the app's purposes. Crowdsourcing provides a means to generate population-appropriate seed data.

Objective: Our objective was to identify and test a method to address the lack of culturally specific preventative personal health data and sidestep the type of algorithmic bias inherent in a RecSys not trained in the population of focus. We did this by collecting a large amount of data quickly and at low cost from members of the population of focus, thereby generating a novel data set based on prevention-focused, population-relevant health goals. We seeded our RecSys with data collected anonymously from self-identified Hispanic and self-identified non-Hispanic African American/Black adult respondents, using Amazon Mechanical Turk (MTurk).

Methods: MTurk provided the crowdsourcing platform for a web-based survey in which respondents completed a personal profile and a health information–seeking assessment, and provided data on family health history and personal health history. Respondents then selected their top 3 health goals related to preventable health conditions, and for each goal, reviewed and rated the top 3 information returns by importance, personal utility, whether the item should be added to their personal health library, and their satisfaction with the quality of the information returned. This paper reports the article ratings because our intent was to assess the benefits of crowdsourcing to seed a RecSys. The analysis of the data from health goals will be reported in future papers.

Results: The MTurk crowdsourcing approach generated 985 valid responses from 485 (49%) self-identified Hispanic and 500 (51%) self-identified non-Hispanic African American adults over the course of only 64 days at a cost of US $6.74 per respondent. Respondents rated 92 unique articles to inform the RecSys.

Conclusions: Researchers have options such as MTurk as a quick, low-cost means to avoid the cold start problem for algorithms and to sidestep bias and low relevance for an intended population of app users. Seeding a RecSys with responses from people

XSL•FO
RenderX

like the intended users allows for the development of a digital health tool that can recommend information to users based on similar demography, health goals, and health history. This approach minimizes the potential, initial gaps in algorithm performance; allows for quicker algorithm refinement in use; and may deliver a better user experience to individuals seeking preventative health information to improve health and achieve health goals.

(J Med Internet Res 2022;24(6):e30216) doi:10.2196/30216

KEYWORDS

crowdsourcing; health information; health promotion; prevention; public health informatics; African American, Black, Latino, and Hispanic populations; recommender system; RecSys; machine learning; Mechanical Turk; MTurk; mobile phone

Introduction

Algorithm Personalization

Algorithms are increasingly used to personalize recommendations of items in stored databases. In simple terms, a personalization algorithm is a computer-implemented service that recommends items to a user based on the known characteristics of that user and the historical preferences of other similar users. The process of training a personalization algorithm is a type of machine learning. The resulting personalization tool is in effect a recommender system (RecSys)—a collaborative information filtering system that attempts to predict a user's preferences for an item based on the previously recorded, similar preferences of other users. Collaborative filtering underlies many popular implementations of personalization algorithms including Amazon.com's "people who buy x also buy y" recommendations [1]. In public health, algorithms to offer targeted and personalized health advice based on personal risk profile and patterns of behavior are as yet an unrealized opportunity [2].

To avoid problems of early poor performance in a new RecSys, algorithms are frequently trained using publicly available data prior to being applied. However, algorithms may reproduce racial, ethnic, and gender disparities because of the data used to train them [3,4]. Racial bias has been detected in commercial algorithms used to guide health decisions among providers [5], as well as in algorithms for hiring [6], natural language processing [7], and sentencing and parole guidelines [8,9]. Algorithms trained on large population-level data sets may underperform when personalizing recommendations for diverse populations [3]. When recommending preventative health information, such underperformance may compound existing inequities in health. The risk of bias inherent in existing publicly available health information data sets is potentially high [3]. Previous qualitative work on barriers to African American and Hispanic adults' health information seeking has shown that commonly available health information resources can be racially or culturally insensitive or may be written implicitly for the dominant culture and not be culturally relevant for the intended population of users [10,11]. A RecSys trained on a data set with very few African American or Hispanic participants may cause these culturally inappropriate resources to be promoted rather than demoted by that RecSys [3].

The lack of publicly available data sets for Black and bilingual/Spanish-speaking Hispanic users of health websites presents a major challenge to researchers who want to develop personalized tools for the health behavior intervention space. Our searches (conducted repeatedly on all dates between November 2020 and November 2021) for "training data," "training data set," "seed data," "collaborative filtering," or "recsys," paired with "black," "african american," "latino," "hispanic," or "race" returned no relevant results or data sets for health information seeking in PubMed and Google Scholar. The time and cost required to collect sufficient new population-specific data to seed an algorithm are additional barriers, especially when the need is for 2 different population groups using 2 different languages, such as English and Spanish.

A potential common means of controlling algorithmic bias is "masking" the algorithm to race or gender in order to avoid capturing or exacerbating any social or structural inequity reflected in the training data. This process of excluding race or gender might solve the algorithmic bias problem in other domains where an algorithm is employed to assist in a decision-making process orthogonal to the demographic characteristic excluded. However, personalization in mobile health (mHealth) depends specifically on race- or gender-based predictions, as race, ethnicity, and gender are key social determinants of health [12]. "Fair" algorithms focused on health must account for the diversity of the groups of people the algorithm's performance may affect [5], and as such, algorithmic fairness in health requires a solution other than masking. Instead of using potentially biased training data or ignoring the impact of race and ethnicity on health, researchers and practitioners need to be able to generate, share, and use robust seed data gathered from people similar to the intended users who will be affected by the algorithm's outputs.

Background

The RecSys seeding discussed in this paper is part of a 4-year smartphone health app research study funded by the National Library of Medicine (Grant 5R01LM013039-02), titled "HealthyMe/MiSalud Smartphone Application: Identifying Mechanisms to Engage African Americans and Hispanics in Personal Health Libraries." A University of Maryland Center for Health Literacy research team is developing the RecSys to deliver personalized health content from MyHealthfinder website to English-speaking African Americans and Spanish-speaking Hispanic adults. The MyHealthfinder website is a free, no-copyright consumer health information collection in English and Spanish maintained by the United States Department of Health and Human Services. The team chose the MyHealthfinder website because the website applies health literacy principles and extensive consumer testing rather than limited, mechanistic reading grade formulas [13]. All articles are written in plain language consistent with the Federal Plain

XSL•FO
RenderX

Language Guidelines [14] and health literacy criteria in the Centers for Disease Control and Prevention (CDC) Clear Communication Index [15] and cover a wide range of health topics linked to evidence-based recommendations from key federal advisory committees. The MyHealthfinder website allows basic personalization of health articles and prevention recommendations when users enter their age, sex, and pregnancy status. The content is available through an application programming interface.

Our research team planned to use the RecSys as the core of a smartphone app with individualized recommendations, guidance on seeking further information, and capacity for users to build personalized libraries in the app [16]. One of the more frequent applications of data science is to build a RecSys with the principal capacity to predict what a user might do next with a high degree of accuracy and to provide a small set of recommended items that have a high likelihood of attracting the user [17]. Health information providers have lagged behind this trend [18].

Personalization in mHealth depends on prediction and performance data, and algorithms that utilize collaborative filtering either rely on existing data for training or are subject to the cold start problem. The cold start problem happens when insufficient data exist at the launch of a RecSys to ensure high-quality recommendations [19]. Consequently, an inadequately personalized algorithm limits the effectiveness of personalization and the utility of the RecSys itself [19]. Two associated problems with collaborative filtering algorithms are scalability and sparsity, particularly in large data sets [20]. The larger the data set, the more computational power is needed to calculate recommendations and the fewer the items any individual user will rate [20]. Scalability and sparsity also slow the process of algorithm learning; to overcome these challenges, developers often employ an initial seed data set for algorithm training. Seed data are necessary to mitigate the cold start problem. However, using data that are a poor match with the intended user group or that have implicit or explicit biases will undermine the user experience, as well as personalization, and thus the utility of a RecSys [3].

To develop a RecSys to predict the most relevant preventative health information and serve it to African American and Hispanic users, we needed seed data describing the users' health goals and the associated relevance of articles and topics in the MyHealthfinder website.

Crowdsourcing

Generating a seed data set is possible with crowdsourcing and the web-based platforms for crowdsourcing tasks used for web-based research [21-23]. Crowdsourcing refers to a set of potential processes through which tasks are proposed by an initiator to solve a problem and are completed by a crowd of individuals rather than a single individual or entity [24]. The components of the crowd operate outside of the initiator's direct control as represented by traditional, hierarchical, organizational structures [24]. The benefits to the initiator include completion of the tasks and solutions to the problem through the expertise of a crowd that would otherwise be cost- and time-prohibitive under traditional models for organizing labor [24].

Amazon Mechanical Turk (MTurk) has become increasingly popular as a crowdsourcing platform for conducting web-based research involving surveys, as MTurk facilitates access to a large and diverse participant population at a relatively low cost to investigators [21-23]. MTurk functions as a web-based labor market where registered workers complete web-based Human Intelligence Tasks (HITs) to be paid. HITs can include a range of tasks including responding to surveys, manually categorizing complex data, or transcribing data. During registration, all MTurk workers are required to electronically sign a participation agreement confirming that they are at least 18 years of age. Likewise, individual researchers must register as MTurk requesters to post HITs and collect data from consenting workers. MTurk provides a template for the construction of HIT surveys run directly on Amazon's developer platform [25]. Researchers post HITs on the Amazon marketplace that MTurk workers self-select and can set both inclusion criteria and task completion criteria. Since MTurk workers are preregistered and come from a large pool, using MTurk may help avoid many of the recruitment barriers that slow survey collection.

In aggregating seed data for an mHealth app, MTurk presents a similar challenge to other population-based surveys: while substantially gender balanced, the majority of the US MTurk workers are White compared with the general population [26,27]. However, researchers can account for this by setting inclusion criteria to garner responses from the population of focus, in our case, African American or Hispanic MTurk workers.

Methods

Overview

We used the following inclusion criteria to identify MTurk respondents for our study: (1) self-identify as African American/Black or Hispanic/Latino/Latina/Latine; (2) own a smartphone; and (3) are located in the United States. Using MTurk we were able to balance respondents by race or ethnicity. Tasks were completed in a single session. If a participant did not complete the full task, the data were not returned, and there was no cost to the project. Respondents could technically complete the full task by entering invalid data for certain text entry fields. To address this, we excluded from analyses any retained responses where invalid data were entered into text entry fields. The reliance on a single encounter and the monetary incentive for completing the HIT are powerful retention strategies. To characterize respondents, we collected self-reported demographics (race or ethnicity, age, self-identified sex, educational attainment) and 3 health behaviors (BMI, smoking, and alcohol consumption).

Our tasks for each MTurk worker included completing the following: (1) personal health profile; (2) family health history; (3) a series of questions about the experience and frustrations in finding and using health information based on the Health Information National Trends Survey fielded by the National Cancer Institute; (4) choosing 3 goals from a list of 24 derived from the Healthy People 2020 survey, part of the US 10-year health objectives; (5) reading 3 randomly selected, topically relevant articles from the MyHealthfinder web-based database

XSL•FO
RenderX

for each of the 3 selected goals; (6) rating each of the 9 articles on two 5-point Likert scales on the importance of the information and feasibility of using the information as well as 1 dichotomous scale on whether or not the respondent would choose to retain the article in a personal library; (7) reading 6 entirely random articles from MyHealthfinder website that may or may not be topical; (8) rating each of those 6 articles using the same 2 Likert and 1 dichotomous scales; (9) searching through the web-based database of the MyHealthfinder website for information relating to each of the 3 goals; and (10) rating each of the information returns, up to 3 from each of the 3 searches, using the same 2 Likert and 1 dichotomous scales. For each MTurk worker who completes the full task (all 10 components), the Amazon marketplace returns an MTurk ID and the data generated.

Among these tasks, article ratings were most important for training an algorithm. In particular, having responses about article relevance was helpful to secure unbiased and population-focused seed data. The outputs of the other HITs are also useful for informing app development but are less directly relevant to seeding a RecSys. Because this is a methodology paper focused on crowdsourcing data for RecSys development, the results of the other outputs are not reported in the next section.

In terms of data collection efficiencies to seed an algorithm, the ability to quickly collect data at a low cost per user is an important consideration. We recorded the time spent on data collection in days and the total cost (including MTurk fees as well as the cost for completed surveys excluded due to invalid data) and calculated the cost per usable respondent.

All analyses were done in Stata/MP software (version 16; StataCorp), SciPy (version 1.6.0; SciPy), and Google Sheets (Google LLC).

Ethical Considerations

The University of Maryland College Park institutional review board determined this project was exempt from institutional review board review and approval, as no identifiable private information was collected or retained by the research team, and so it did not meet the definition of human subject research.

Results

Our MTurk crowdsourcing approach produced sufficient data on participant characteristics and expressed the preferences needed to seed the algorithm, assess the cost effectiveness of the data collection method, and address algorithmic implicit bias. These included (1) producing an adequate sample size of populations traditionally with limited data, (2) reducing the data collection period and data collection cost, and (3) collecting specifically the data set required to seed an algorithm and minimize the cold start problem.

MTurk Benefit 1: Producing an Adequate Sample Size of Populations Traditionally With Limited Data

Our sampling approach produced 2578 respondents who selected and started the survey and a total of 1015 respondents who met the inclusion criteria and completed the full task. We collected and retained data from 1015 respondents out of which 30 respondents (3% of the retained sample) were excluded due to invalid data entered, for a final sample size of 985 (Table 1). A total of 500 (51%) respondents identified as non-Hispanic Black or African American and 485 (49%) identified as Hispanic/Latino/Latina/Latine. There was an almost even split between self-identified female and male respondents, and 3 respondents (less than 1%) of the sample did not identify with the binary gender designations. Respondents tended to be younger, with a mean age of 32 (SD 9) years, and 545 (55%) of the sample were between the ages of 18 and 30 years. Potentially reflective of the younger age and online recruitment of respondents, 830 (83%) respondents reported having at least some college education, of those 239 (24%) had completed college or a graduate degree.

XSL•FO
RenderX

Table 1. Self-reported participant demographics.

Characteristics	All participants (N=985)
Race/Ethnicity[a], n (%)	
Non-Hispanic Black	500 (50.76)
Hispanic/Latino/Latina/Latine	485 (49.24)
Age (years), mean (SD)	32.15 (8.75)
18-30, n (%)	545 (54.50)
31-40, n (%)	305 (30.50)
41-50, n (%)	105 (10.50)
51-60, n (%)	37 (3.70)
61-70, n (%)	8 (0.80)
Sex, n (%)	
Female	494 (49.45)
Male	502 (50.25)
Other	3 (0.30)
BMI[b], mean (SD)	26.64 (12.64)
Underweight, n (%)	94 (9.64)
Normal, n (%)	363 (37.23)
Overweight, n (%)	267 (27.38)
Obese, n (%)	251 (25.76)
Drink 2 × /week, mean (SD)	294 (39.30)
Currently smoker, mean (SD)	194 (19.62)
Educational level, n (%)	
High school or lower	165 (16.58)
Some college	591 (59.40)
College degree	115 (11.56)
Graduate degree	124 (12.46)

[a]Non-Hispanic Black and Latino/Latina/Latine are derived from self-reported race and Hispanic ethnicity items.

[b]BMI was calculated using height, weight, and sex, and using BMI English system on the Center for Disease Control and Prevention website. The ranges were devised by the World Health Organization.

MTurk Benefit 2: Reducing the Data Collection Period and the Data Collection Cost

It took 64 days to collect data for the training set. The total cost including MTurk fees and the cost for 30 unusable respondents was US $6635.20 or US $6.74 per usable respondent. An alternative data collection method resulting in 985 unique respondents would have likely taken considerably longer and incurred substantially greater expenses. Alternatively, seeding our algorithm with data from fewer unique respondents would not have adequately minimized the cold start problem.

MTurk Benefit 3: Collecting Specifically the Data Set Required to Seed the Algorithm and Minimize the Risk of the Cold Start Problem

Respondents rated a total of 92 unique articles. A selection of the top 5 articles that Black and Hispanic respondents rated by importance and by feasibility of using the information is presented in Table 2.

XSL•FO
RenderX

Table 2. Comparison of Black and Hispanic participants: the top 5 rated articles on the MyHealthfinder website.

Rating[a]	Article[b] name	
	Black participants	Hispanic participants
Importance		
1st	Reduce Your Risk of Stroke	Reduce Your Risk of Stroke
2nd	Prevent Infections When You Get Medical Care	Get Your Blood Pressure Checked
3rd	Manage Stress	Talk with Your Doctor about Taking Aspirin to Prevent Disease
4th	Quit Smoking	Manage Stress
5th	Get Screened	Take Care of Your Teeth and Gums
Feasibility		
1st	Reduce Your Risk of Stroke	Reduce Your Risk of Stroke
2nd	Learn First Aid	Manage Stress
3rd	Get Screened	Talk with Your Doctor about Taking Aspirin to Prevent Disease
4th	Manage Stress	Quit Smoking
5th	Prevent Infections When You Get Medical Care	Get Your Blood Pressure Checked

[a]Respondents rated importance and feasibility for each article on a 5-point Likert scale. Importance and feasibility are measured on a range of 1 to 5, derived from the Health Information National Trends Survey. A total of 92 unique articles were rated. We have displayed the top 5 articles by importance and feasibility for each demographic group.

[b]Articles were pulled from the MyHealthfinder website and were read and rated by the respondents.

Discussion

Principal Findings

Previous studies have shown that crowdsourcing is an effective means of gathering data from a large number of human participants quickly and at a low cost [21-23]. Our results show that crowdsourcing through a technology such as Amazon MTurk can leverage a large, low-cost sampling method to generate seed data for a RecSys and sidestep the cold start problem and the potential algorithmic racial bias inherent in using general population seed data [3]. Unlike traditional survey methods that are reliant on a response rate, the MTurk approach ensures that required cohort sizes are met as HITs remain open until prespecified participant thresholds are met, and the researcher receives data only on respondents who complete all data collection tasks.

Our approach also allows for the development of a digital health tool to recommend more relevant information to users based on similar demography and health history. This is particularly important for public health purposes, where both algorithmic bias and the common tactic of masking algorithms to demographic data might limit the utility of a prevention-focused mHealth tool [3-5]. Through crowdsourcing we were able to efficiently and affordably recruit a large sample of African American and Hispanic participants—our population of focus—to share their health goals and for each goal, rate article returns from a federally supported database of public health information. In addition, the results of the HITs that are not reported in this methodology paper also informed app design and developments beyond the RecSys.

Along with far greater flexibility in item content and greater timeliness, the cost per usable response was an order of magnitude below the cost per complete response (US $40 to US $102) compared with similarly detailed health questionnaires such as the Behavioral Risk Factor Surveillance System (BRFSS) survey [28]. Our data collection period of around 2 months is far more condensed than the BRFSS's year-round data collection. To our knowledge, we have collected the first such publicly available seed data set for health information seeking for non-Hispanic African American and Hispanic populations.

Limitations

The principal limitation of this study is that despite a large sample size and despite limiting data collection to African American and Hispanic respondents, MTurk participants are potentially demographically dissimilar in some ways to our app user population. On average, MTurk workers are younger and more educated than the general population and are likely more technologically literate as demonstrated by their participation as workers in a web-based marketplace. However, the majority of our respondents did not have a 4-year or graduate degree. A total of 756 (76%) respondents had only some college education or less, which was similar to our intended app user group. Studying the deployment of the HealthyMe/MiSalud RecSys trained on these seed data will allow us to quantify to what extent these demographic differences limited the applicability of preventative health information provided by the personal health app.

In our deployment, it is not imperative, however, that the seed data perfectly match the intended app user population, since the RecSys continues to "learn" iteratively as app users review and

XSL•FO
RenderX

rate articles, further refining the recommendations that the system makes. Importantly, in this way the limitation inherent in crowdsourcing with MTurk does not pose a significant impact on the development of a RecSys, and the benefits of demographically similar (though not identical) seed data in overcoming the cold start problem, scalability, and sparsity likely exceed the limitations of training the RecSys with MTurk data. Future evaluations and field tests of our RecSys will enable us to quantify the utility of a crowdsourced population-specific seeded RecSys versus a generically seeded RecSys or an unseeded RecSys in returning user-rated relevancy of personalized health content and improving user health information-seeking behaviors in these populations.

Conclusion

Researchers have crowdsourcing options such as Amazon MTurk, for quick, low-cost means to avoid the cold start problem for algorithms and sidestep bias and low relevance for an intended population of app users. Seeding a RecSys with more population-relevant responses allows for the development of a digital health tool that can recommend more relevant information to users based on similar demography, health goals, and health history. If made publicly available, the generation of such seed data sets can also enable other researchers and developers to more rapidly develop additional population-specific solutions for health and health literacy. In the long term, this approach may minimize potential initial gaps in algorithm performance, allow quicker algorithm refinement, and deliver a better user experience.

Conflicts of Interest

None declared.

References

1. Jacobi J, Benson E, Linden G. Personalized recommendations of items represented within a database. Google Patents. 2006. URL: https://patents.google.com/patent/US7113917B2/en [accessed 2021-05-04]
2. Panch T, Pearson-Stuttard J, Greaves F, Atun R. Artificial intelligence: opportunities and risks for public health. Lancet Digit Health 2019 May;1(1):e13-e14 [FREE Full text] [doi: 10.1016/S2589-7500(19)30002-0] [Medline: 33323236]
3. Stinson C. Algorithms are not neutral: bias in collaborative filtering. arXiv. Preprint posted online May 3, 2021 [FREE Full text] [doi: 10.1007/s43681-022-00136-w]
4. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019 Oct 25;366(6464):447-453. [doi: 10.1126/science.aax2342] [Medline: 31649194]
5. Panch T, Mattie H, Atun R. Artificial intelligence and algorithmic bias: implications for health systems. J Glob Health 2019 Dec;9(2):020318 [FREE Full text] [doi: 10.7189/jogh.09.020318] [Medline: 31788229]
6. Lambrecht A, Tucker C. Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of stem career ads. Manage Sci 2019 Jul;65(7):2966-2981. [doi: 10.1287/mnsc.2018.3093]
7. Caliskan A, Bryson JJ, Narayanan A. Semantics derived automatically from language corpora contain human-like biases. Science 2017 Apr 14;356(6334):183-186. [doi: 10.1126/science.aal4230] [Medline: 28408601]
8. Angwin J, Larson J, Mattu S, Kirchner L. Machine bias. ProPublica. URL: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [accessed 2021-05-04]
9. Chouldechova A. Fair prediction with disparate impact: a study of bias in recidivism prediction instruments. Big Data 2017 Jun;5(2):153-163. [doi: 10.1089/big.2016.0047] [Medline: 28632438]
10. Birru M, Steinman RA. Online health information and low-literacy African Americans. J Med Internet Res 2004 Sep 03;6(3):e26 [FREE Full text] [doi: 10.2196/jmir.6.3.e26] [Medline: 15471752]
11. Kvasny L. Health portals and menu-driven identities. In: Medical Informatics: Concepts, Methodologies, Tools, and Applications. Hershey, PA: IGI Global; 2009:1549-1557.
12. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring Fairness in Machine Learning to Advance Health Equity. Ann Intern Med 2018 Dec 18;169(12):866-872 [FREE Full text] [doi: 10.7326/M18-1990] [Medline: 30508424]
13. Baur C, Prue C. The CDC Clear Communication Index is a new evidence-based tool to prepare and review health information. Health Promot Pract 2014 Sep 20;15(5):629-637. [doi: 10.1177/1524839914538969] [Medline: 24951489]
14. Quesenberry AC. Plain language for patient education. J Consum Health Internet 2017 Jun 12;21(2):209-215. [doi: 10.1080/15398285.2017.1311611]
15. Hou S. Health literacy online: a guide to writing and designing easy-to-use health web sites. Health Promot Pract 2012 Sep 03;13(5):577-580. [doi: 10.1177/1524839912446480] [Medline: 22763891]
16. Cheng W, Yin G, Dong Y, Dong H, Zhang W. Collaborative filtering recommendation on users' interest sequences. PLoS One 2016 May 19;11(5):e0155739 [FREE Full text] [doi: 10.1371/journal.pone.0155739] [Medline: 27195787]
17. Lu J, Wu D, Mao M, Wang W, Zhang G. Recommender system application developments: a survey. Decis Support Syst 2015 Jun;74:12-32. [doi: 10.1016/j.dss.2015.03.008]
18. Agapito G, Simeoni M, Calabrese B, Caré I, Lamprinoudi T, Guzzi PH, et al. DIETOS: A dietary recommender system for chronic diseases monitoring and management. Comput Methods Programs Biomed 2018 Jan;153:93-104. [doi: 10.1016/j.cmpb.2017.10.014] [Medline: 29157465]

XSL•FO
RenderX

19. Azadjalal MM, Moradi P, Abdollahpouri A, Jalili M. A trust-aware recommendation method based on Pareto dominance and confidence concepts. Knowl Based Syst 2017 Jan;116:130-143. [doi: 10.1016/j.knosys.2016.10.025]
20. Massa P, Avesani P. Trust-aware recommender systems. In: Proceedings of the 2007 ACM conference on Recommender Systems. 2007 Presented at: RecSys '07; Oct 19-20, 2007; Minneapolis, MN p. 17-24. [doi: 10.1145/1297231.1297235]
21. Buhrmester M, Kwang T, Gosling S. Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data? In: Methodological Issues and Strategies in Clinical Research. Washington, DC: American Psychological Association; 2016:133-139.
22. Turner AM, Kirchhoff K, Capurro D. Using crowdsourcing technology for testing multilingual public health promotion materials. J Med Internet Res 2012 Jun 04;14(3):e79 [FREE Full text] [doi: 10.2196/jmir.2063] [Medline: 22664384]
23. Fei-Fei L, Deng J, Li K. ImageNet: constructing a large-scale image database. J Vis 2009 Aug 01;9(8):1037-1037. [doi: 10.1167/9.8.1037]
24. Estellés-Arolas E, González-Ladrón-de-Guevara F. Towards an integrated crowdsourcing definition. J Inf Sci 2012 Mar 09;38(2):189-200. [doi: 10.1177/0165551512437638]
25. Mason W, Suri S. Conducting behavioral research on Amazon's Mechanical Turk. Behav Res Methods 2012 Mar;44(1):1-23. [doi: 10.3758/s13428-011-0124-6] [Medline: 21717266]
26. Huff C, Tingley D. "Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents. Research and Politics 2015 Sep 10;2(3) [FREE Full text] [doi: 10.1177/2053168015604648]
27. Burnham MJ, Le YK, Piedmont RL. Who is Mturk? Personal characteristics and sample consistency of these online workers. Ment Health Relig Cult 2018 Jul 19;21(9-10):934-944. [doi: 10.1080/13674676.2018.1486394]
28. Costs to conduct BRFSS surveys. Utah Department of Health Center for Health Data and Informatics. 2018. URL: https://le.utah.gov/interim/2018/pdf/00003907.pdf [accessed 2022-01-26]

Abbreviations

BRFSS: Behavioral Risk Factor Surveillance System
CDC: Centers for Disease Control and Prevention
HIT: Human Intelligence Task
mHealth: mobile health
MTurk: Amazon Mechanical Turk
RecSys: recommender system

Edited by G Eysenbach; submitted 05.05.21; peer-reviewed by S Nagavally, C Hudak; comments to author 28.09.21; revised version received 31.01.22; accepted 07.03.22; published 21.06.22.

Please cite as:
Sehgal NJ, Huang S, Johnson NM, Dickerson J, Jackson D, Baur C
The Benefits of Crowdsourcing to Seed and Align an Algorithm in an mHealth Intervention for African American and Hispanic Adults: Survey Study
J Med Internet Res 2022;24(6):e30216
URL: https://www.jmir.org/2022/6/e30216
doi:10.2196/30216
PMID:35727616

©Neil Jay Sehgal, Shuo Huang, Neil Mason Johnson, John Dickerson, Devlon Jackson, Cynthia Baur. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.06.2022. 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

Machine Learning–Based Prediction Models for Different Clinical Risks in Different Hospitals: Evaluation of Live Performance

Hong Sun[1*], PhD; Kristof Depraetere[1*], MSc; Laurent Meesseman[1], MD; Patricia Cabanillas Silva[1], MSc; Ralph Szymanowsky[1], MSc; Janis Fliegenschmidt[2], BSc; Nikolai Hulde[2], MD, PhD; Vera von Dossow[2], MD, PhD, Prof Dr; Martijn Vanbiervliet[1], MSc; Jos De Baerdemaeker[1], MSc; Diana M Roccaro-Waldmeyer[1], PhD; Jörg Stieg[1], MSc; Manuel Domínguez Hidalgo[1], MSc; Fried-Michael Dahlweid[1], MD, PhD

[1]Dedalus Healthcare, Antwerp, Belgium

[2]Institute of Anesthesiology and Pain Therapy, Heart and Diabetes Centre North Rhine-Westphalia, University Hospital of Ruhr-University Bochum, Bad Oeynhausen, Germany

[*]these authors contributed equally

Corresponding Author:
Hong Sun, PhD
Dedalus Healthcare
Roderveldlaan 2
Antwerp, 2600
Belgium
Phone: 32 3444 8108
Email: hong.sun@dedalus.com

Abstract

Background: Machine learning algorithms are currently used in a wide array of clinical domains to produce models that can predict clinical risk events. Most models are developed and evaluated with retrospective data, very few are evaluated in a clinical workflow, and even fewer report performances in different hospitals. In this study, we provide detailed evaluations of clinical risk prediction models in live clinical workflows for three different use cases in three different hospitals.

Objective: The main objective of this study was to evaluate clinical risk prediction models in live clinical workflows and compare their performance in these setting with their performance when using retrospective data. We also aimed at generalizing the results by applying our investigation to three different use cases in three different hospitals.

Methods: We trained clinical risk prediction models for three use cases (ie, delirium, sepsis, and acute kidney injury) in three different hospitals with retrospective data. We used machine learning and, specifically, deep learning to train models that were based on the Transformer model. The models were trained using a calibration tool that is common for all hospitals and use cases. The models had a common design but were calibrated using each hospital's specific data. The models were deployed in these three hospitals and used in daily clinical practice. The predictions made by these models were logged and correlated with the diagnosis at discharge. We compared their performance with evaluations on retrospective data and conducted cross-hospital evaluations.

Results: The performance of the prediction models with data from live clinical workflows was similar to the performance with retrospective data. The average value of the area under the receiver operating characteristic curve (AUROC) decreased slightly by 0.6 percentage points (from 94.8% to 94.2% at discharge). The cross-hospital evaluations exhibited severely reduced performance: the average AUROC decreased by 8 percentage points (from 94.2% to 86.3% at discharge), which indicates the importance of model calibration with data from the deployment hospital.

Conclusions: Calibrating the prediction model with data from different deployment hospitals led to good performance in live settings. The performance degradation in the cross-hospital evaluation identified limitations in developing a generic model for different hospitals. Designing a generic process for model development to generate specialized prediction models for each hospital guarantees model performance in different hospitals.

(J Med Internet Res 2022;24(6):e34295) doi:10.2196/34295

XSL•FO
RenderX

KEYWORDS

machine learning; clinical risk prediction; prediction; model; model evaluation; scalability; risk; live clinical workflow; delirium; sepsis; acute kidney injury; kidney; EHR; electronic health record; workflow; algorithm

Introduction

Machine learning algorithms for clinical risk predictions are widely used in health care research and applications [1-5]. While much work has been done on developing distinct clinical risk prediction models, the scalability of the prediction models has been much less explored (ie, the extensibility of the risk prediction model for multiple diseases over different hospitals) [6].

Rajkomar et al [6] designed a single data structure based on the FHIR (Fast Healthcare Interoperability Resources) standard [7] and developed different clinical scenarios over two hospitals with this common data structure. That was the first study that reported the performance of prediction models of multiple use cases in different hospitals. Churpek et al [8] aggregated the electronic health record (EHR) from five hospitals to train a single model to make predictions on cardiac arrest, intensive care unit (ICU) transfers, or death on wards. The performance of the model outperforms the existing Modified Early Warning Score. The limitation is that both studies [6,8] were validated with retrospective data and have not yet been used in a live clinical workflow.

In our previous publication [9], we discussed the scalability issue in clinical risk prediction model development; we also presented a scalable approach for prediction model development that is applied to delirium, sepsis, and acute kidney injury (AKI) covering four different hospitals. However, these prediction models were only evaluated on retrospective data.

Evaluating the prediction models in live clinical settings is crucial because factors such as interoperability across different platforms or different prevalence can affect the performance of an artificial intelligence (AI) algorithm [10,11]. However, very few prediction models have been evaluated in a live clinical workflow. For example, several delirium prediction models that have been reported in recent years [9,12,13] have all been evaluated on retrospective data. Jauk et al [14] claimed their findings to be the only delirium prediction model that has been evaluated in a live clinical workflow. In their study, 5530 predictions were analyzed, and 119 predictions were compared with ratings of clinical experts during a period of 7 months. The limitation of Jauk et al [14] is that their model only evaluated in a single hospital. When a prediction model is evaluated in different hospitals, the performance may degrade due to the difference in EHRs and workflows between the training data and the target hospital. Wong et al [15] reported large performance degradation on sepsis prediction when a sepsis prediction model was applied in a different hospital.

Wu et al [16] considered it important to evaluate AI-based medical devices over different sites with live clinical settings to address the shortcomings, such as overfitting to training data and bias against underrepresented subgroups, among others. They investigated 130 US Food and Drug Administration–approved AI devices: 126 evaluations were performed as retrospective studies and 93 devices did not have multiple site evaluations.

In this paper, we evaluated clinical risk prediction models (ie, delirium, sepsis, and AKI) in live clinical workflows in three different hospitals in Germany. We compared the performance of the models with their performance on retrospective data from our previous work. By logging prediction requests in the production EHR system, we ran cross-hospital evaluations mimicking the performance of a prediction model in live clinical workflows of different target hospitals. Domain experts executed preliminary evaluations on clinical soundness and usefulness of the predictions by following the use of the prediction service in their daily practice.

To the best of our knowledge, we are the first to report the evaluation of machine learning–based clinical risk prediction models in the settings of production EHR systems, which focuses on evaluating several diseases in different hospitals at the same time. In addition, in the cross-hospital evaluation, we simulated the performance of a prediction model in live clinical workflows of different target hospitals.

Methods

Overview

We used a scalable approach, implemented in a calibration tool, to generate clinical risk prediction models for different use cases in three different German hospitals based on retrospective EHR data: Marienhospital Stuttgart (from 2004 to 2020), Herz- und Diabeteszentrum Nordrhein-Westfalen Bad Oeynhausen (from 2009 to 2020), and Medius Klinik Nürtingen (from 2009 to 2020). The evaluation in live systems was performed in the first half of 2021; details of the evaluation period are provided in Table S1 in Multimedia Appendix 1. The characteristics of the training set are provided in Table S2 in Multimedia Appendix 1, and the characteristics of the evaluation samples, in live systems, are provided in Table S3 in Multimedia Appendix 1. We refer to these three hospitals as hospital M, hospital H, and hospital N, respectively. The calibration process that generates prediction models is described in our previous work [9]. Using the calibration tool, models were trained independently on data from each hospital and deployed in the prediction service of the same hospital. Requests for predictions were generated from the EHR system in the FHIR [7] "RiskAssessment" format and were sent to the prediction service. The prediction service parsed each prediction request into an observation, which was used to generate a prediction. The predictions were returned and displayed in the EHR system. The observation, together with the corresponding risk score produced by the prediction model, were stored for further evaluation of model performance.

Model Development and Deployment

Figure 1 shows the process of model development and evaluation with retrospective data. The process to design the prediction model, prepare data, and train models was defined

XSL•FO
RenderX

following experiments performed on a development data set. The resulting dedicated process and prediction model design was implemented in an automated pipeline, named the calibration tool. The calibration tool provided a user-friendly approach to install, configure, and run the process of data preparation, model training, and evaluation on a customer-specific system. A command-line interface enabled service engineers to install the required software, files, and pretrained natural language processing (NLP) models and to execute the training and evaluation of the hospital-specific prediction models.

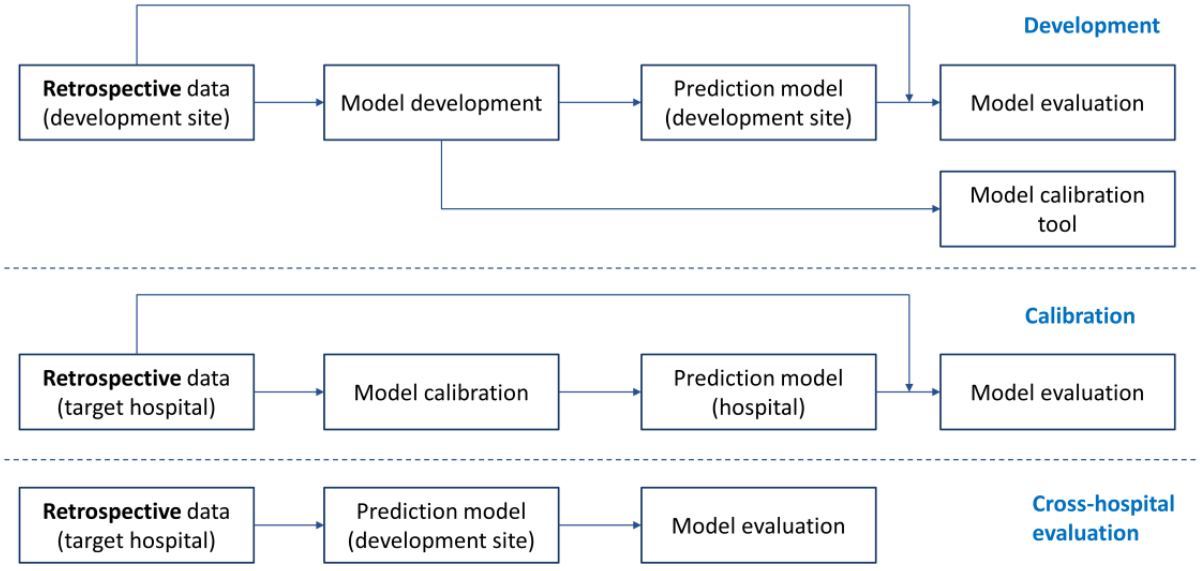
Figure 2 shows the components and interactions of the calibration tool. The lower pane defines a fixed sequence of tasks to perform in order to calibrate models for the supported use cases with data from a target hospital. The upper pane contains a set of components that execute these tasks.

We then ran the calibration tool independently in each hospital to generate clinical risk prediction models for each hospital. The models were trained based on the retrospective data that were generated as part of the clinical workflow of each target hospital. We thereby ensured that the model fit the clinical practice of the hospital where the model was to be deployed. The data checking process guaranteed that the source data were represented in the expected format. The data preparation process prepared the training and testing data. The labels of each use case were assigned by the labeler component based on the diagnosis codes assigned to each hospitalized patient at discharge. A common set of features was prepared and used by the different use cases, which included structured data, such as lab results and history of diagnosis, as well as clinical entities extracted from free-text clinical notes. Both a text search and a BERT (bidirectional encoder representations from transformers) [17] named entity recognition model were used in preparing the NLP features. The following inclusion criteria were applied during data preparation: age and gender had to be known, patients had to be 18 years or older, only inpatients could be included, and length of stay had to be limited to 90 days.

Prediction models were trained using a common model training strategy: we used the Transformer model [18] to train a binary classification model for clinical risk prediction. We concatenated the features as inputs and used the labels as targets for the model training process. The models were trained with patient data that were collected at the time of discharge with leaking features removed. In order to cope with the situation where the model was requested to make predictions when less information was available, we applied data augmentation in training sample preparation: we generated partial records in combination with the complete records to enhance the robustness of the clinical risk prediction model. More details can be found in our previous work [9]. The generated models were first examined with a model checking process, where a set of minimum requirements were assessed as unit tests. Models that passed the checks were further evaluated on their performance using common metrics, such as the area under the receiver operating characteristic curve (AUROC), sensitivity, and specificity, among others. Acceptance criteria were checked during the model evaluation process. The acceptance criteria differed among use cases and were checked for each department. Models that met the criteria could be activated in the corresponding departments to trigger alerts in the production EHR system. The acceptance criteria were complex and are explained in detail in Table S4 in Multimedia Appendix 1.

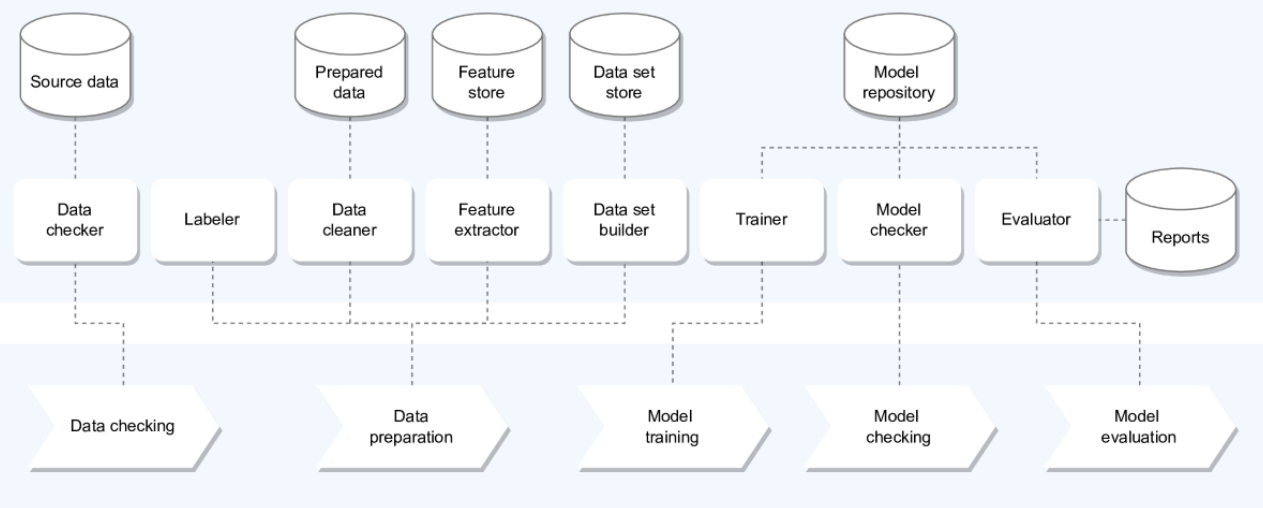
Risk prediction models evaluated in this paper were generated by our calibration tool in the three aforementioned German hospitals. The details of feature engineering and model training were presented in a former publication [9]; examples of model input features are provided in Table S5 in Multimedia Appendix 1. Preliminary cross-hospital evaluation was performed with retrospective data in our previous study and performance degradation was observed [9].

Figure 1. Model development and evaluation with retrospective data.



XSL•FO
RenderX

Figure 2. Calibration tool.

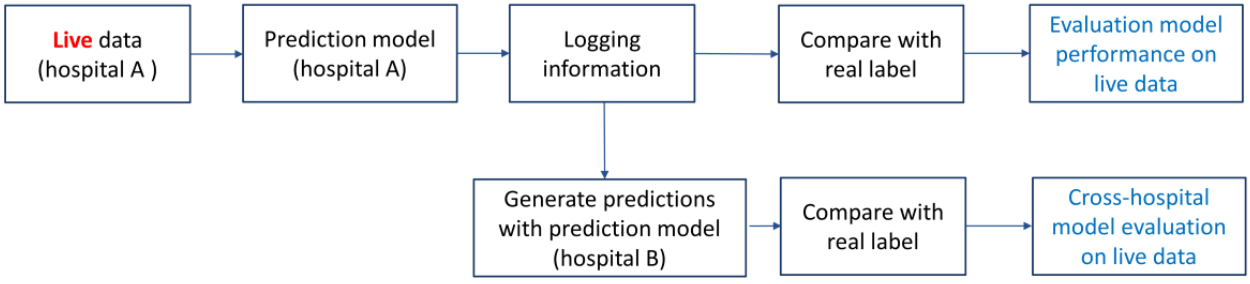


Model Evaluation With Live Data From the Clinical Workflow

Figure 3 shows the process of model evaluation with live data from the clinical workflow. Prediction services were triggered following clinical events in the EHR system (eg, when new lab results for a patient were added to the system). The EHR system sent the relevant patient record to the prediction service, where the hospitals' specialized prediction models for three different use cases, trained on the hospital data, were deployed. For each use case, the prediction model predicted the risk of developing the related disease and returned the risk score in response. Based on the defined thresholds, alerts were created in the EHR system for those that were predicted as high risk. For each prediction made by the prediction service, the corresponding request and response were stored by a logging service. By comparing the predictions made by the prediction service and the corresponding real labels, we evaluated the model performance in a live clinical workflow. Moreover, the prediction requests stored by the logging service could be used to generate predictions with a different model to simulate its performance in a live clinical workflow. This alternate model could be a model that is trained in the same hospital with a different training strategy, as well as a model that is trained on the data from a different hospital. For example, in Figure 3, the logging information stored in hospital A (ie, the hospital where risk predictions in a live EHR system are made) can be used to generate predictions with a model trained at hospital B (ie, a different hospital where a different risk prediction model is trained). By comparing those predictions with the real labels, it is possible to estimate the performance of the model of hospital B in the live clinical workflow of hospital A.

To support the evaluation presented in this paper, the JSON file logging driver (ie, the default Docker logging service) [19] was used to log the request and response of prediction services to separate JSON log files. Each prediction request log entry contained the date and time and the input features for the prediction. Each prediction response log entry contained the used input features and risk score for the prediction. An excerpt of a sample log of prediction requests is enclosed (Table S5 in Multimedia Appendix 1). In the production EHR system, the prediction service can process a patient's records and instantly make a corresponding prediction or explanation. Prediction models for delirium, sepsis, and AKI were installed in three hospitals. Prediction requests and responses were logged in these three hospitals as input for the evaluations. The response time for predictions was evaluated and provided (Figure S1 in Multimedia Appendix 1).

Figure 3. Model evaluation with live data from the clinical workflow. Hospital A refers to the hospital where risk predictions in a live electronic health record system are made. Hospital B refers to a different hospital where a different risk prediction model is trained.



Ethical Considerations

Our study to assess model performance involved the analysis of unidentifiable patients, and data use was granted to us by the pilot hospitals—hospital H, hospital M, and hospital N—for this purpose, after appropriate review. Therefore, no ethics approval by the Institutional Review Board was required. The cohort study at hospital H was approved by the Ethics

XSL•FO
RenderX

Committee of the Medical Faculty of the Ruhr-Universität Bochum (file No. Az.2021-861).

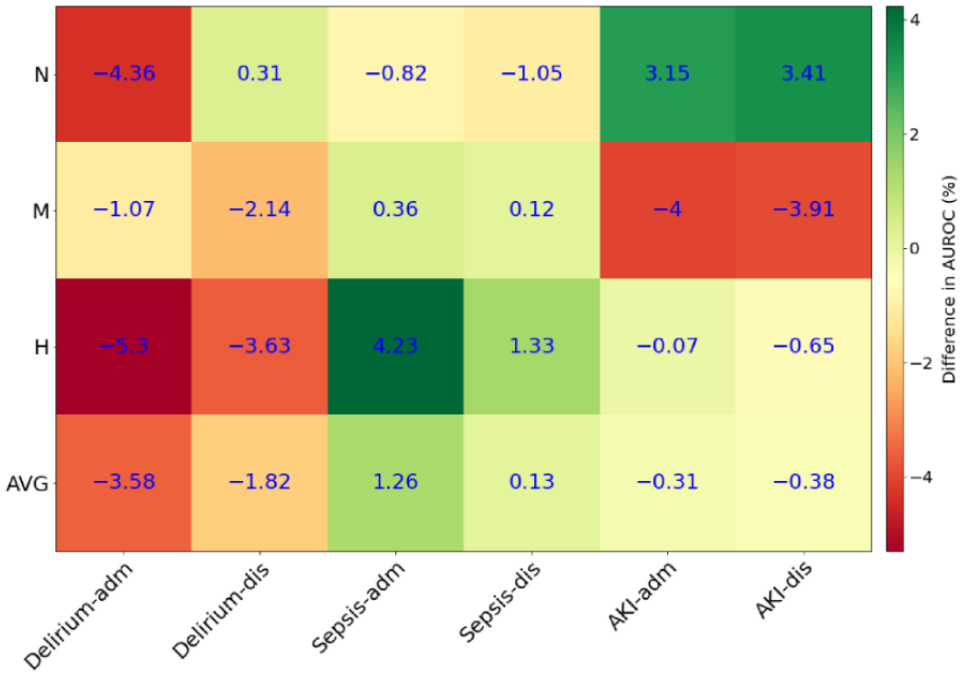
Results

Evaluation of Model Performance in Live EHR Systems

The models made predictions at different stages during a patient stay with live data; however, the performance of the clinical risk prediction models within a live clinical workflow was evaluated at the end of the day of admission, as well as on the day of discharge. The reason we checked the performance of our prediction model at these two stages was to evaluate their performance when there were limited data compared to when sufficient data were available. Leaking information, such as strong diagnostic data or textual references to the diseases to be predicted, was excluded, following the settings we applied when we evaluated the model performance on retrospective data in our previous study [9]. Taking these same strategies allowed a fair comparison between the performance achieved on live data with that obtained on retrospective data. The model performance was evaluated by the AUROC. We choose to evaluate using the AUROC because the sensitivity, specificity, and precision were dependent on the threshold (ie, defined by the point chosen on the receiver operating characteristic curve). The threshold is used by the hospitals to trigger an alert and may differ among hospitals because some hospitals favor sensitivity over specificity or vice versa. Using the AUROC allowed us to compare the outcome of three use cases at three different hospitals independently from this threshold. Sensitivity, specificity, and precision were used to decide on the threshold and are provided with the explanation of the model acceptance criteria (Table S4 in Multimedia Appendix 1).

Figure 4 evaluates the model performance as assessed by the AUROC on the live data versus the retrospective data (Table S6 in Multimedia Appendix 1). Each row in the table indicates the hospital in which the evaluation was performed. Each column indicates the use case and the point in time of the model evaluation (ie, either at the end of the admission day or at discharge). Positive values (ie, shades of green) indicate that the respective model performed better when evaluated on live data as compared to retrospective data, whereas negative values (ie, shades of red) indicate that models performed worse when evaluated on live data as compared to retrospective data. For example, the delirium model AUROC, evaluated at the end of the day of admission at hospital N, was 4.36 percentage points lower when the model was performed on the live data (AUROC=80.9%) as compared to retrospective data (AUROC=85.26%). On average, our delirium prediction model had a lower AUROC when evaluated on live data as compared to retrospective data. In contrast, our sepsis prediction model performed better on live data as compared to retrospective data, whereas the AKI prediction model performed equally well on both. At the hospital level, evaluation on live data led to higher model performance in hospital N (+0.1 percentage points) but to lower performance in hospitals M and H (–1.8 and –0.7 percentage points, respectively). When averaged across all three use cases and all three hospitals, the performance of our prediction models declined slightly when evaluated on live data (AUROC values: 83.1% at admission, 94.2% at discharge, and 88.6% on average) as compared to retrospective data (AUROC values: 83.0% at admission, 94.8% at discharge, and 89.4% on average).

Figure 4. Model performance: live data vs retrospective data. The table was generated using the AUROC values for the live and retrospective data (Table S6 in Multimedia Appendix 1). adm: admission; AKI: acute kidney injury; AVG: average; AUROC: area under the receiver operating characteristic curve; dis: discharge; H: Herz- und Diabeteszentrum Nordrhein-Westfalen Bad Oeynhausen; M: Marienhospital Stuttgart; N: Medius Klinik Nürtingen.



XSL•FO
RenderX

Cross-Hospital Evaluation

Cross-hospital evaluation was performed by extracting the observations from the prediction request in one hospital and generating predictions using a model trained on data from a different hospital. We evaluated our models in a live clinical workflow based on the logging information stored in the prediction service. The prediction requests made at different stages of a medical stay were used to generate corresponding predictions by prediction models of other hospitals. By using the prediction models of other hospitals, we simulated the performance of these models in a live clinical workflow, without the model being installed on-site.

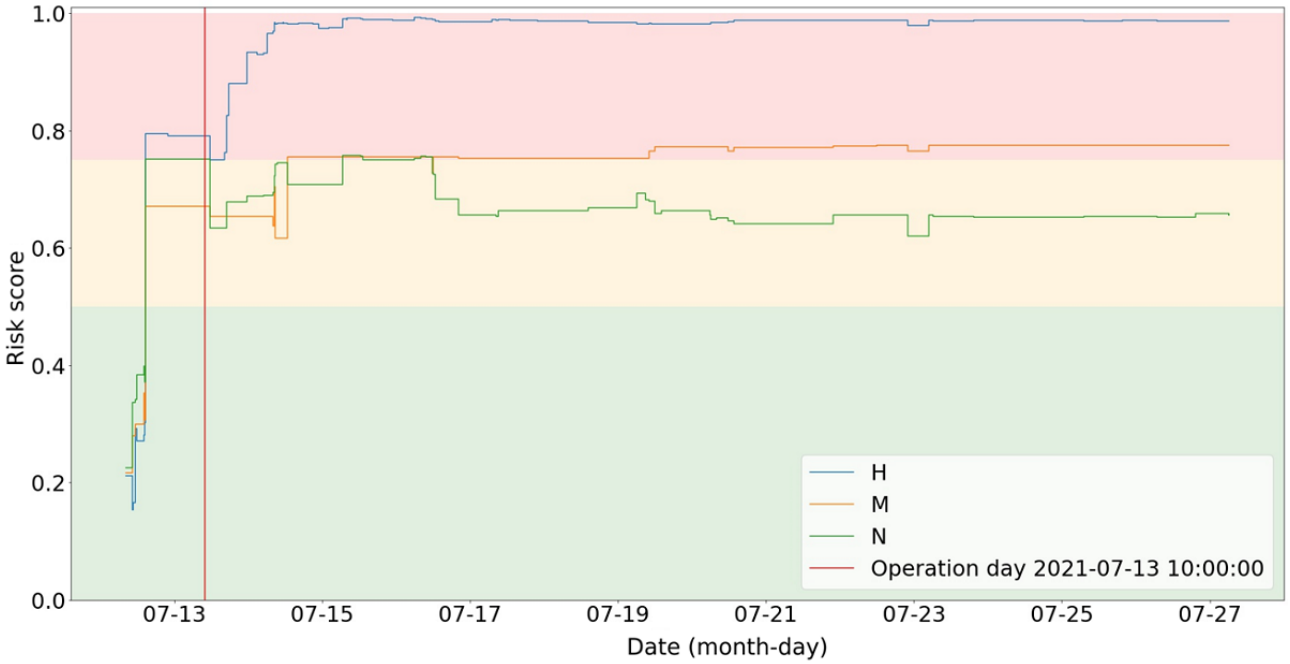
Figure 5 shows an example of simulating the performance of models trained on data from hospitals M and N, but applied to live data of the medical stay of a sample patient in hospital H. The red vertical line indicates the point in time of the patient's surgery. The three colored lines reflect the simulated model prediction over the course of the patient's medical stay in hospital H, using models trained separately on data from hospital H, M, and N.

In the presented case, postoperative delirium was confirmed by an independent evaluation—the Confusion Assessment Method for the ICU (CAM-ICU) [20]—on the first postoperative day. The CAM-ICU evaluation was not included as a feature of our training model. Of the three models, the one trained at hospital H predicted the risk of delirium before surgery and identified an increased risk after surgery. The risk after surgery increased gradually when lab results and clinical entities were added. The models trained at the other hospitals both predicted the risk of delirium before surgery, but both failed to properly identify the severity of the risk after surgery.

The detailed outcome of cross-hospital evaluation on prediction requests extracted from the live clinical workflow is provided in Table S7 in Multimedia Appendix 1. Prediction models for three different diseases (ie, delirium, sepsis, and AKI) were evaluated by comparing the AUROCs of different models at discharge. Figure 6 depicts the performance degradation of a model when trained in a certain hospital and deployed in another hospital. For each use case, AUROC values in a row are compared to the white cell in the same row, which indicates within-hospital performance. For example, when the delirium model trained on data from hospital H was deployed in hospital M (91.2%, column 2, row 1; Table S7 in Multimedia Appendix 1), the AUROC was 3.2 percentage points lower as compared to its performance in hospital H (94.4%, column 1, row 1; Table S7 in Multimedia Appendix 1). The largest performance degradation (–20.5 percentage points) was observed when the AKI model trained on data from hospital H was deployed in hospital M.

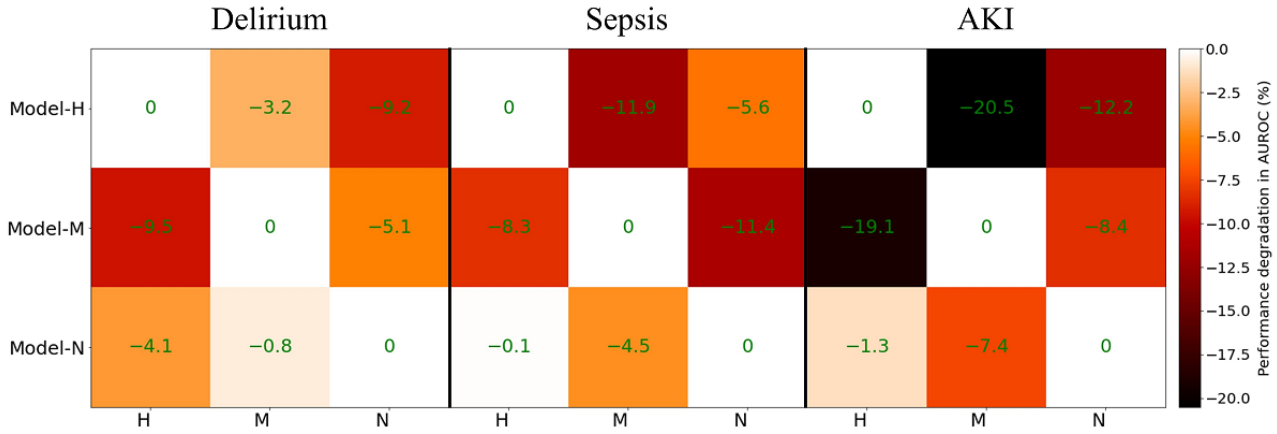
On average, the AUROC was 8 percentage points lower when a model was deployed in a hospital other than where it was trained (from 94.2% to 86.3%).

Figure 5. Delirium risk prediction of a sample patient during his medical stay based on data from the live electronic health record system. H: Herz- und Diabeteszentrum Nordrhein-Westfalen Bad Oeynhausen; M: Marienhospital Stuttgart; N: Medius Klinik Nürtingen.



XSL•FO
RenderX

Figure 6. Performance degradation of a model trained in a certain hospital (rows) and deployed in another hospital (columns). The table was generated from the AUROC values from cross-hospital evaluation on the live data (Table S6 in Multimedia Appendix 1). AKI: acute kidney injury; AUROC: area under the receiver operating characteristic curve; H: Herz- und Diabeteszentrum Nordrhein-Westfalen Bad Oeynhausen; M: Marienhospital Stuttgart; N: Medius Klinik Nürtingen.



Preliminary Evaluation of Clinical Soundness and Usefulness

The prediction models were installed in the live EHR systems of three different hospitals. These models generated predictions and triggered alerts in a live clinical workflow. Those alerts were displayed in the production EHR system and are currently under the evaluation of domain experts. A quantitative evaluation of the impact on clinical outcomes has not yet been performed. Nevertheless, the preliminary evaluation made by the domain experts assures the correctness and effectiveness of the predictions. A case study has been conducted to evaluate the performance of the delirium prediction models installed in hospital H [21]. Predictions made by the delirium risk prediction model following cardiac surgery were evaluated in the study. A cohort study investigating a larger population is also ongoing in the same hospital. The investigations identified that the prediction service could have an influence on anesthesia planning, as risk prediction is crucial for an early prevention strategy. The machine learning approach also improved postoperative care by enhanced screening efforts. In addition, the rest of this section presents our analysis of calibration and decision curves at hospital H, as well as our preliminary analysis on user feedback at hospital M.

Calibration and Decision Curve Analysis

Figure 7 shows the calibration and decision curve analysis for three use cases with the live data retrieved from hospital H. We first applied probability calibration [22,23] to generate calibration curves for each use case. The calibration curves plot the true frequency of the positive cases against its averaged predicted probability for each bin. We divided the probability into 10 bins. Predictions on the live data before and after probability calibration are shown. We used isotonic regression to perform the probability calibration. The calibration process used the first half of the live data, and the calibration curves and decision curves were generated using the second half of the live data. Due to the limited amount of available data, there are a few spikes in the calibrated curves. After the probability calibration, the decision curves [24,25] were generated to evaluate the net benefit of using the prediction models. The net benefits of the prediction models were compared with either “alert all patients” or “no alerts.” It can be observed that the prediction models were clinically useful when the threshold probability was below 90% for the AKI and sepsis use cases. For the delirium use case, the model had benefits when the threshold probability was below 70%.

Figure 8 shows the decision curves for the prediction models trained at hospitals H and M on the sepsis use case. Both curves in Figure 8 were generated using the live data from hospital H, and the predicated probabilities were both calibrated. It can be observed that the model trained at hospital H was superior compared with the model trained at hospital M.

XSL•FO
RenderX

Figure 7. Calibration and decision curve analysis. The model and data were both from hospital H (Herz- und Diabeteszentrum Nordrhein-Westfalen Bad Oeynhausen). AKI: acute kidney injury.

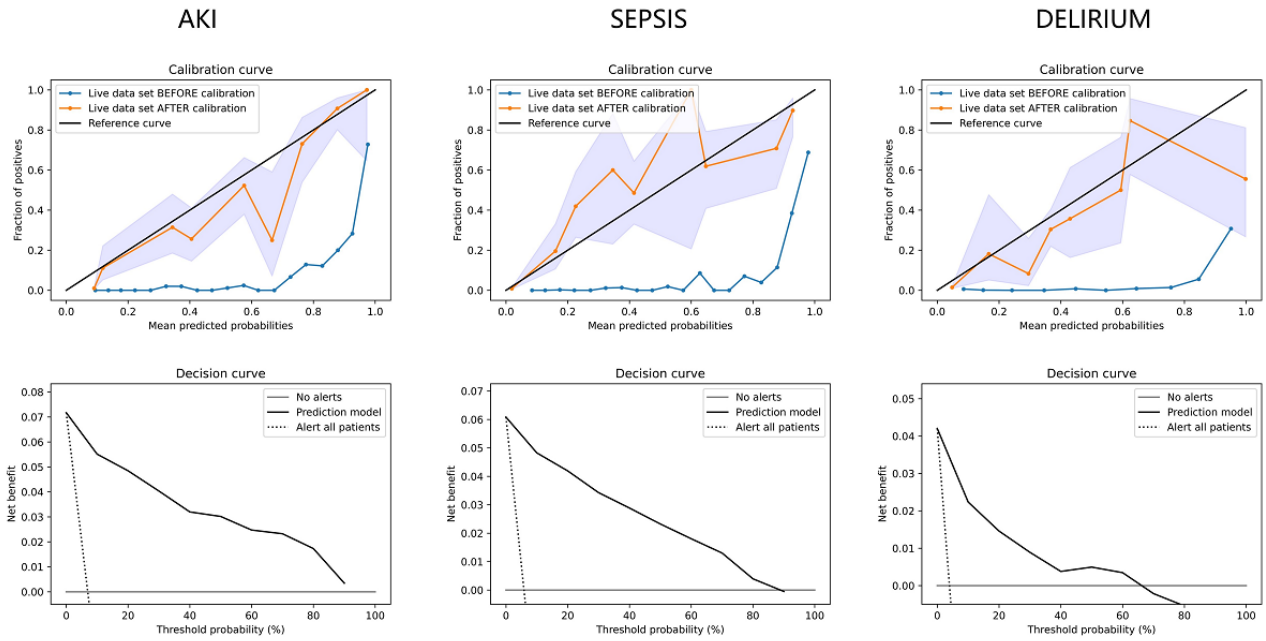
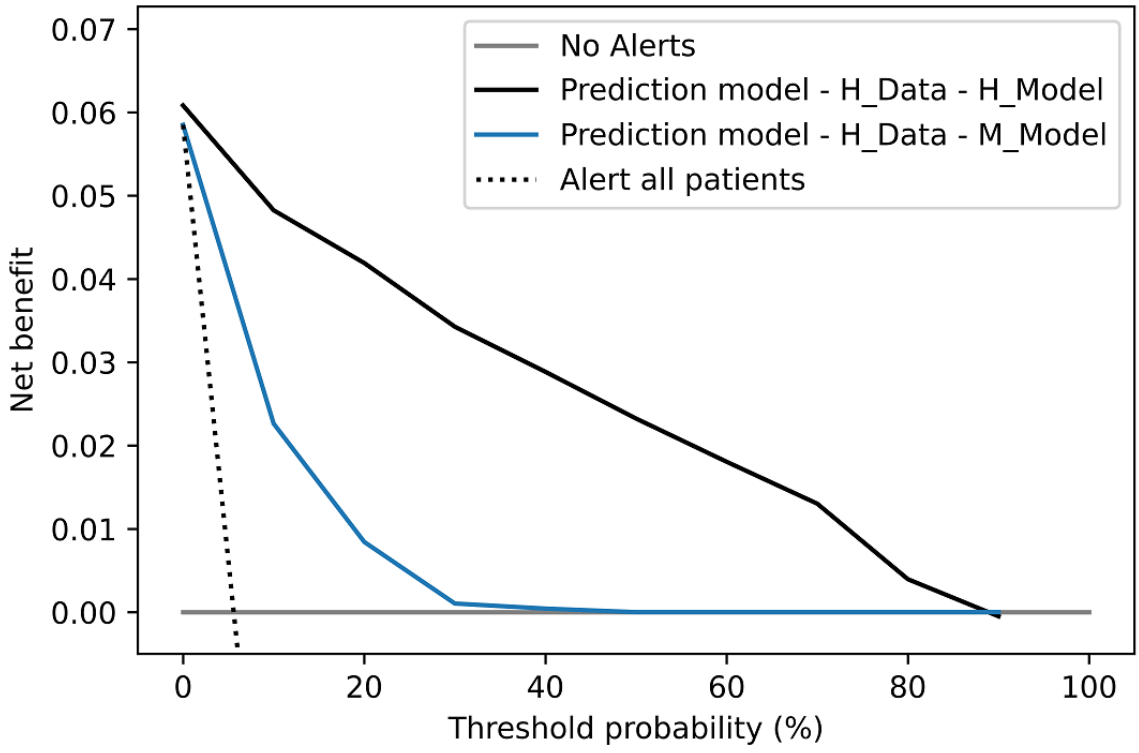


Figure 8. Decision curve analysis for the sepsis use case. Models trained at hospitals H and M, both using the live data from hospital H, are compared. H: Herz- und Diabeteszentrum Nordrhein-Westfalen Bad Oeynhausen; M: Marienhospital Stuttgart.



Preliminary Analysis of User Feedback

When the prediction models were installed in the production EHR system, the end user was able to provide their feedback when they closed an alert. There were 134 feedback entries collected for the AKI use case at hospital M. More than one-third of the feedback entries (n=46, 34.3%) indicated that the users found the predictions useful. Details of the user feedback entries can be found in Figure S2 in Multimedia Appendix 1.

A total of 27.6% (37/134) of the alerts were considered to be false positives by the end users. This is a satisfactory result, considering the low incidence of AKI (838/8861, 9.46% at hospital M). Moreover, among 37 of those evaluated as false positive cases, 20 (54%) were already discharged and coded. Of these 20 discharged cases, 4 (20%) were actually coded as

XSL•FO
RenderX

having AKI. This means that even if the physician disagrees with a prediction of high risk, there seems to still be a high risk that some patients will ultimately develop AKI, and our model can identify that risk.

In 38.1% (51/134) of the cases, the end users were already aware of the risk of AKI raised by the alert. There were two main reasons for this. Firstly, there was a clear gap between the time that the alert was created and the time that the feedback was given when the alert was closed. Secondly, alerts were only displayed in departments where the prediction service was activated; if a patient was transferred from a department where the prediction service was not activated, there would not be any alert displayed there.

Discussion

Principal Findings

The state of the art of machine learning development is to either design and train a single model and use it in different hospitals or design and train a specific model for a single hospital. We claim that defining a generic model design and training a specific instance of the model with data from a specific hospital has additional benefits for replicating the results. We observed performance degradation when a model was deployed in another hospital in our cross-hospital evaluation, a typical limitation of developing a single model for different hospitals. In the meantime, having a generic process and common model design to generate hospital-specific prediction models is a more robust solution. It resolves the intrinsic differences between different hospitals and guarantees sound performance at target hospitals. The evaluation of model performance in live clinical workflows assured the feasibility of such a generic approach, by checking the performance on three use cases at three different hospitals. In addition, by storing the logging data from live clinical workflows and having a common model design, the evaluation presented in this paper allows one to simulate the performance of a model in a live clinical workflow without it being installed on-site.

Motivations

Machine learning–based prediction models are closely tied to the data used in the training process. This dependency largely restricts the reusability of a prediction model in other hospitals. A generic model that delivers unbiased performance in different hospitals is what machine learning scientists and clinicians earnestly long for but also often fail to achieve.

The prerequisite to generate a generic model that can be used in different hospitals is to achieve semantic interoperability that guarantees a common understanding between different EHR systems [26,27]. In order to achieve semantic interoperability, clinical terminologies need to be mapped onto a standard representation. However, a recent study [28] also showed safety risks related to the use of standard terminologies, such as LOINC (Logical Observation Identifiers Names and Codes), for interoperability between organizations due to inaccurate mappings.

In addition, a disease may have very different incidence rates in different hospitals due to the type and specialties of a hospital. Such a variety also results in different clinical workflows performed in different hospitals that determine the data the hospital records. A prediction model is, therefore, considered an algorithm that captures the knowledge and practice of the physicians of a hospital, by processing hospital-specific data that are presented in their specific representation. It is challenging to overcome the vulnerability of data shifts caused by diverse clinical workflows in different hospitals. Therefore, it is hard to maintain good performance when a model runs in a different hospital than the one within which it was trained, especially if the characteristics of the EHR data and the clinical workflow differ significantly. For example, the sepsis prediction of one particular vendor achieved satisfactory results in one hospital [29], but it was substantially worse when evaluated in another hospital [15].

We also observed performance degradation when a model was deployed in other hospitals in our cross-hospital evaluation. Therefore, instead of delivering a generic prediction model to different hospitals, we designed a generic procedure for prediction model development and applied it to different hospitals. Having a generic process to generate hospital-specific prediction models is a more robust solution; it resolves the intrinsic differences between different hospitals.

Strengths

Evaluating prediction models in a live clinical workflow is crucial for validating their performances. To the best of our knowledge, we are the first to evaluate clinical prediction models on such a large scale in live clinical workflows. Such a thorough evaluation avoids overfitting to a certain disease or the settings of a particular hospital, thus allowing a fair, unbiased evaluation. The models deployed in the live clinical workflows delivered similar performances compared with those reported in our previous study [9], which were evaluated using retrospective data.

Sharing the same feature processing approach allows us to reuse the prediction requests by different prediction models. We, therefore, performed cross-hospital evaluation on three use cases in three different hospitals, mimicking the performance in a live clinical workflow rather than on retrospective data. To the best of our knowledge, this is the first study that performed cross-hospital evaluations on multiple use cases and simulated model performance in live clinical workflows.

Limitations

This study had some limitations. First, our model development and evaluation on metrics reported in this paper lacked a dynamic evaluation to predict the risk within a time window of event onset. For example, the most widely used diagnostic criterion for AKI is based on changes in serum creatinine, as defined by the Kidney Disease: Improving Global Outcomes (KDIGO) guidelines [30]. Tomašev et al [31] reported an AKI prediction model that predicts the AKI risk 48 hours before the KDIGO-defined event. In the three use cases presented in this paper, delirium was considered a mental health disease that normally does not have a precise time of onset. We have developed an AKI prediction model based on retrospective data at our development site using KDIGO events as labels. The

XSL•FO
RenderX

models were not deployed in the production system; however, their performance at two hospitals on retrospective data is enclosed (Table S8 in Multimedia Appendix 1). The AKI risk prediction curve of a sample patient during his medical stay is also provided (Figure S3 in Multimedia Appendix 1). For the sepsis prediction, we did not yet perform such a dynamic evaluation due to the lack of both scalable and accurate indicators of documented or suspected infection. Nevertheless, the AUROC for sepsis at the end of the day of admission ranged between 86.9% and 88.5% in the live system at the three different hospitals, which assures a satisfactory performance.

Second, although the metrics of the prediction models in the live clinical workflows were evaluated in different hospitals, the corresponding clinical outcomes in clinical practice are yet to be measured. Nevertheless, the preliminary clinical evaluation in hospital H affirms that there was a positive impact in the live clinical workflows, and a quantitative evaluation is scheduled as our next step of this work. The decision curve analysis and the preliminary analysis of user feedback also affirms the usefulness of our prediction models.

Third, machine learning approaches that are used to generate and validate prediction models are always data hungry [32]. Current external validation studies often suffer from small sample sizes compared with the large amount of predictor features [33]. The sample size presented in this paper was also relatively small compared with the number of predictor features used in our prediction model. Nevertheless, we also argue that for diseases with low incidence, it is difficult to obtain a large number of positive samples. The three use cases presented in this paper were running in live EHR systems for more than half a year, which we consider to be a reasonable amount of time. In addition, we ran evaluations on three different use cases at three different hospitals, which helps to justify the outcomes.

Future Directions

Our future work will focus on evaluating the detailed clinical outcomes of prediction models in clinical practice. In addition, we will also evaluate the impact of different labeling strategies, such as defining AKI events with KDIGO criteria, in live systems.

Conclusions

In this study, we found consistent performance of models when evaluated on retrospective and live data, and performance differences were observed in the cross-hospital evaluations. This ensures that designing a generic process for model development, implementing that design in a calibration tool, and generating hospital-specific prediction models with a common model design is a valid approach that guarantees model performance in different hospitals.

Acknowledgments

JF, NH, and VvD received funding from the ARGUS project (grant 100126059) from the Ruhr-Universität Bochum. Authors from Dedalus received funding from the European Union's Horizon 2020 projects GenoMed4All (Genomics and Personalized Medicine for All Though Artificial Intelligence in Haematological Diseases; grant 101017549) and PERSIST (Patients-Centered Survivorship Care Plan After Cancer Treatments Based on Big Data and Artificial Intelligence Technologies; grant 875406). The authors would like to acknowledge Dieter Vanden Abeele and Nico Lapauw for the integration of the prediction services into the ORBIS EHR system and for preparing the logs, Corry Clybouw for proofreading and improving the presentation of the paper, and Marienhospital Stuttgart, Herz- und Diabeteszentrum Nordrhein-Westfalen Bad Oeynhausen, Alexianer Krefeld, and Medius Klinik Nürtingen for assisting with the model calibration and prediction model evaluation in their production systems.

Data Availability

The patient data used in this evaluation from the three hospitals cannot be made publicly available due to patient protection. The code used to evaluate the model performance and to run the cross-hospital evaluation is available at GitHub [34].

Authors' Contributions

HS, KD, and LM conceptualized the study and designed the evaluation methods. MDH provided input to consolidate the study. HS and PCS performed the model evaluation. LM, JF, NH, and VvD provided the clinical perspective. RS coordinated the resources to perform the evaluation. KD and MDH supervised the process of model evaluation. MV, HS, KD, and LM defined the model design, and MV, JDB, HS, and KD developed the calibration tool accordingly. KD and JDB developed the solution architecture for the prediction service and its integration into the production system. HS wrote the original draft, and MDH and LM provided input as medical editors. All authors reviewed and edited the manuscript critically and approved its final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials.
[DOCX File , 208 KB - jmir_v24i6e34295_app1.docx]

XSL•FO
RenderX

References

1. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med 2019 Jan;25(1):24-29. [doi: 10.1038/s41591-018-0316-z] [Medline: 30617335]
2. Cutillo CM, Sharma KR, Foschini L, Kundu S, Mackintosh M, Mandl KD, MI in Healthcare Workshop Working Group. Machine intelligence in healthcare-perspectives on trustworthiness, explainability, usability, and transparency. NPJ Digit Med 2020;3:47 [FREE Full text] [doi: 10.1038/s41746-020-0254-2] [Medline: 32258429]
3. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019 Apr 04;380(14):1347-1358. [doi: 10.1056/NEJMra1814259] [Medline: 30943338]
4. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models with electronic health records data: A systematic review. J Am Med Inform Assoc 2017 Jan;24(1):198-208 [FREE Full text] [doi: 10.1093/jamia/ocw042] [Medline: 27189013]
5. Topol EJ. High-performance medicine: The convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
6. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med 2018;1:18 [FREE Full text] [doi: 10.1038/s41746-018-0029-1] [Medline: 31304302]
7. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: A standards-based, interoperable apps platform for electronic health records. J Am Med Inform Assoc 2016 Sep;23(5):899-908 [FREE Full text] [doi: 10.1093/jamia/ocv189] [Medline: 26911829]
8. Churpek MM, Yuen TC, Winslow C, Robicsek AA, Meltzer DO, Gibbons RD, et al. Multicenter development and validation of a risk stratification tool for ward patients. Am J Respir Crit Care Med 2014 Sep 15;190(6):649-655 [FREE Full text] [doi: 10.1164/rccm.201406-1022OC] [Medline: 25089847]
9. Sun H, Depraetere K, Meesseman L, De Roo J, Vanbiervliet M, De Baerdemaeker J, et al. A scalable approach for developing clinical risk prediction applications in different hospitals. J Biomed Inform 2021 Jun;118:103783 [FREE Full text] [doi: 10.1016/j.jbi.2021.103783] [Medline: 33887456]
10. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019 Jan;25(1):30-36 [FREE Full text] [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
11. Domalpally A, Channa R. Real-world validation of artificial intelligence algorithms for ophthalmic imaging. Lancet Digit Health 2021 Aug;3(8):e463-e464 [FREE Full text] [doi: 10.1016/S2589-7500(21)00140-0] [Medline: 34325850]
12. Kim MY, Park UJ, Kim HT, Cho WH. DELirium Prediction based on Hospital Information (Delphi) in general surgery patients. Medicine (Baltimore) 2016 Mar;95(12):e3072 [FREE Full text] [doi: 10.1097/MD.0000000000003072] [Medline: 27015177]
13. Wong A, Young AT, Liang AS, Gonzales R, Douglas VC, Hadley D. Development and validation of an electronic health record-based machine learning model to estimate delirium risk in newly hospitalized patients without known cognitive impairment. JAMA Netw Open 2018 Aug 03;1(4):e181018 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.1018] [Medline: 30646095]
14. Jauk S, Kramer D, Großauer B, Rienmüller S, Avian A, Berghold A, et al. Risk prediction of delirium in hospitalized patients using machine learning: An implementation and prospective evaluation study. J Am Med Inform Assoc 2020 Jul 01;27(9):1383-1392 [FREE Full text] [doi: 10.1093/jamia/ocaa113] [Medline: 32968811]
15. Wong A, Otles E, Donnelly JP, Krumm A, McCullough J, DeTroyer-Cooley O, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. JAMA Intern Med 2021 Aug 01;181(8):1065-1070. [doi: 10.1001/jamainternmed.2021.2626] [Medline: 34152373]
16. Wu E, Wu K, Daneshjou R, Ouyang D, Ho DE, Zou J. How medical AI devices are evaluated: Limitations and recommendations from an analysis of FDA approvals. Nat Med 2021 Apr;27(4):582-584. [doi: 10.1038/s41591-021-01312-x] [Medline: 33820998]
17. Devlin J, Chang M, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. ArXiv. Preprint posted online on May 24, 2019 2022 [FREE Full text]
18. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A, et al. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017 Presented at: The 31st International Conference on Neural Information Processing Systems; December 4-9, 2017; Long Beach, CA p. 6000-6010 URL: https://dl.acm.org/doi/pdf/10.5555/3295222.3295349
19. JSON File logging driver. Docker Docs. URL: https://docs.docker.com/config/containers/logging/json-file/ [accessed 2022-04-05]
20. Khan BA, Perkins AJ, Gao S, Hui SL, Campbell NL, Farber MO, et al. The Confusion Assessment Method for the ICU-7 delirium severity scale. Crit Care Med 2017;45(5):851-857. [doi: 10.1097/ccm.0000000000002368]
21. Fliegenschmidt J, Hulde N, Preising MG, Ruggeri S, Szymanowski R, Meesseman L, et al. Artificial intelligence predicts delirium following cardiac surgery: A case study. J Clin Anesth 2021 Dec;75:110473. [doi: 10.1016/j.jclinane.2021.110473] [Medline: 34333447]

XSL•FO
RenderX

22. Riley RD, Ensor J, Snell KIE, Debray TPA, Altman DG, Moons KGM, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: Opportunities and challenges. BMJ 2016 Jun 22;353:i3140 [FREE Full text] [doi: 10.1136/bmj.i3140] [Medline: 27334381]
23. Niculescu-Mizil A, Caruana R. Predicting good probabilities with supervised learning. In: Proceedings of the 22nd International Conference on Machine Learning. 2005 Presented at: The 22nd International Conference on Machine Learning; August 7-11, 2005; Bonn, Germany p. 625-632. [doi: 10.1145/1102351.1102430]
24. Vickers AJ, Elkin EB. Decision curve analysis: A novel method for evaluating prediction models. Med Decis Making 2006;26(6):565-574 [FREE Full text] [doi: 10.1177/0272989X06295361] [Medline: 17099194]
25. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. Diagn Progn Res 2019;3:18 [FREE Full text] [doi: 10.1186/s41512-019-0064-7] [Medline: 31592444]
26. Sun H, Depraetere K, De Roo J, Mels G, De Vloed B, Twagirumukiza M, et al. Semantic processing of EHR data for clinical research. J Biomed Inform 2015 Dec;58:247-259 [FREE Full text] [doi: 10.1016/j.jbi.2015.10.009] [Medline: 26515501]
27. Bhartiya S, Mehrotra D, Girdhar A. Issues in achieving complete interoperability while sharing electronic health records. Procedia Comput Sci 2016;78:192-198. [doi: 10.1016/j.procs.2016.02.033]
28. Carter AB, de Baca ME, Luu HS, Campbell WS, Stram MN. Use of LOINC for interoperability between organisations poses a risk to safety. Lancet Digit Health 2020 Nov;2(11):e569 [FREE Full text] [doi: 10.1016/S2589-7500(20)30244-2] [Medline: 33328084]
29. Bennett T, Russell S, King J, Schilling L, Voong C, Rogers N, et al. Accuracy of the Epic sepsis prediction model in a regional health system. ArXiv. Preprint posted online on February 19, 2019 2022 [FREE Full text]
30. Khwaja A. KDIGO clinical practice guidelines for acute kidney injury. Nephron Clin Pract 2012;120(4):c179-c184 [FREE Full text] [doi: 10.1159/000339789] [Medline: 22890468]
31. Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. Nature 2019 Aug;572(7767):116-119 [FREE Full text] [doi: 10.1038/s41586-019-1390-1] [Medline: 31367026]
32. van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: A simulation study for predicting dichotomous endpoints. BMC Med Res Methodol 2014 Dec 22;14:137 [FREE Full text] [doi: 10.1186/1471-2288-14-137] [Medline: 25532820]
33. Riley RD, Debray TPA, Collins GS, Archer L, Ensor J, van Smeden M, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. Stat Med 2021 Aug 30;40(19):4230-4251. [doi: 10.1002/sim.9025] [Medline: 34031906]
34. Evaluate ML models at different hospitals. GitHub. URL: https://github.com/patriciacs1994/Evaluate-ML-models-at-different-hospitals [accessed 2022-05-10]

Abbreviations

AI: artificial intelligence
AKI: acute kidney injury
AUROC: area under the receiver operating characteristic curve
BERT: bidirectional encoder representations from transformers
CAM-ICU: Confusion Assessment Method for the Intensive Care Unit
EHR: electronic health record
FHIR: Fast Healthcare Interoperability Resources
GenoMed4All: Genomics and Personalized Medicine for All Though Artificial Intelligence in Haematological Diseases
hospital H: Herz- und Diabeteszentrum Nordrhein-Westfalen Bad Oeynhausen
hospital M: Marienhospital Stuttgart
hospital N: Medius Klinik Nürtingen
ICU: intensive care unit
KDIGO: Kidney Disease: Improving Global Outcomes
LOINC: Logical Observation Identifiers Names and Codes
NLP: natural language processing
PERSIST: Patients-Centered Survivorship Care Plan After Cancer Treatments Based on Big Data and Artificial Intelligence Technologies

XSL•FO
RenderX

Edited by T Leung; submitted 19.10.21; peer-reviewed by JL Domínguez-Olmedo, R Bajpai, X Zhang; comments to author 07.01.22; revised version received 25.02.22; accepted 12.04.22; published 07.06.22.

Please cite as:
Sun H, Depraetere K, Meesseman L, Cabanillas Silva P, Szymanowsky R, Fliegenschmidt J, Hulde N, von Dossow V, Vanbiervliet M, De Baerdemaeker J, Roccaro-Waldmeyer DM, Stieg J, Domínguez Hidalgo M, Dahlweid FM
Machine Learning–Based Prediction Models for Different Clinical Risks in Different Hospitals: Evaluation of Live Performance
J Med Internet Res 2022;24(6):e34295
URL: https://www.jmir.org/2022/6/e34295
doi:10.2196/34295
PMID:35502887

©Hong Sun, Kristof Depraetere, Laurent Meesseman, Patricia Cabanillas Silva, Ralph Szymanowsky, Janis Fliegenschmidt, Nikolai Hulde, Vera von Dossow, Martijn Vanbiervliet, Jos De Baerdemaeker, Diana M Roccaro-Waldmeyer, Jörg Stieg, Manuel Domínguez Hidalgo, Fried-Michael Dahlweid. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.06.2022. 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

Machine Learning–Based Text Analysis to Predict Severely Injured Patients in Emergency Medical Dispatch: Model Development and Validation

Kuan-Chen Chin[1*], MD; Yu-Chia Cheng[2*], BS; Jen-Tang Sun[3], MD, MS; Chih-Yen Ou[2], BS; Chun-Hua Hu[4], EMTP; Ming-Chi Tsai[4], EMTP; Matthew Huei-Ming Ma[5,6], MD, PhD; Wen-Chu Chiang[5,6*], MD, PhD; Albert Y Chen[2*], PhD

[1]Department of Emergency Medicine, Taipei Hospital, Ministry of Health and Welfare, New Taipei City, Taiwan

[2]Department of Civil Engineering, National Taiwan University, Taipei City, Taiwan

[3]Department of Emergency Medicine, Far Eastern Memorial Hospital, New Taipei City, Taiwan

[4]Emergency Medical Service Division, Taipei City Fire Department, Taipei City, Taiwan

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

[6]Department of Emergency Medicine, National Taiwan University Hospital, Yun-Lin Branch, Yunlin County, Taiwan

[*]these authors contributed equally

Corresponding Author:
Albert Y Chen, PhD
Department of Civil Engineering
National Taiwan University
No 1, Section 4, Roosevelt Rd
Taipei City, 106
Taiwan
Phone: 886 2 3366 4255
Email: AlbertChen@ntu.edu.tw

Abstract

Background: Early recognition of severely injured patients in prehospital settings is of paramount importance for timely treatment and transportation of patients to further treatment facilities. The dispatching accuracy has seldom been addressed in previous studies.

Objective: In this study, we aimed to build a machine learning–based model through text mining of emergency calls for the automated identification of severely injured patients after a road accident.

Methods: Audio recordings of road accidents in Taipei City, Taiwan, in 2018 were obtained and randomly sampled. Data on call transfers or non-Mandarin speeches were excluded. To predict cases of severe trauma identified on-site by emergency medical technicians, all included cases were evaluated by both humans (6 dispatchers) and a machine learning model, that is, a prehospital-activated major trauma (PAMT) model. The PAMT model was developed using term frequency–inverse document frequency, rule-based classification, and a Bernoulli naïve Bayes classifier. Repeated random subsampling cross-validation was applied to evaluate the robustness of the model. The prediction performance of dispatchers and the PAMT model, in severe cases, was compared. Performance was indicated by sensitivity, specificity, positive predictive value, negative predictive value, and accuracy.

Results: Although the mean sensitivity and negative predictive value obtained by the PAMT model were higher than those of dispatchers, they obtained higher mean specificity, positive predictive value, and accuracy. The mean accuracy of the PAMT model, from certainty level 0 (lowest certainty) to level 6 (highest certainty), was higher except for levels 5 and 6. The overall performances of the dispatchers and the PAMT model were similar; however, the PAMT model had higher accuracy in cases where the dispatchers were less certain of their judgments.

Conclusions: A machine learning–based model, called the PAMT model, was developed to predict severe road accident trauma. The results of our study suggest that the accuracy of the PAMT model is not superior to that of the participating dispatchers; however, it may assist dispatchers when they lack confidence while making a judgment.

XSL•FO
RenderX

(J Med Internet Res 2022;24(6):e30210) doi:10.2196/30210

KEYWORDS

emergency medical service; emergency medical dispatch; dispatcher; trauma; machine learning; frequency–inverse document frequency; Bernoulli naïve Bayes

Introduction

Background

Trauma is a leading cause of accidental death globally. According to the World Health Organization, injuries contribute to >5 million deaths each year. Road traffic accidents accounted for most injuries and were the ninth leading cause of death in 2012 [1]. Severe trauma is a time-sensitive emergency condition. Prompt transport is beneficial for patients with neurotrauma and penetrating injuries with unstable hemodynamic features [2]. Delays in transportation are associated with poor functional outcome [3].

Prehospital triage allows severely ill patients to receive appropriate time-sensitive management. For cardiac arrest and stroke victims, dispatchers can obtain critical information on the phone, such as the patient's level of consciousness, breath patterns, or prehospital stroke scales [4,5]. However, no standardized questions have been designed for dispatchers when they encounter severe trauma. Only a few studies on helicopter emergency medical services have addressed the accuracy of dispatch for trauma victims [6]. Current trauma scales for predicting severity require either physiological or anatomical assessments [7]. Therefore, a victim's condition cannot be identified or evaluated until the first batch of emergency medical technicians (EMTs) arrives at the scene.

Motivation

Content analysis has been conducted on emergency calls to discover the factors that affect dispatch and have the potential to assist prehospital triage [8,9]. Specifically, text classification has demonstrated the effectiveness of classifying events recorded during phone calls [10]. In addition, natural language processing has been used in emergency medicine. Text mining techniques have been used to predict the triage level, length of stay, disposition, and mortality in emergency department patients [11-16]. A textual analysis–based machine learning framework was developed to assist dispatchers during the prehospital phase in out-of-hospital cardiac arrest (OHCA) recognition; this framework has been commercialized [17-20]. These techniques make it possible to stratify the risk to patients when structured questions are unavailable, similar to the assessment of trauma patients over the phone.

The classic process of text classification includes text preprocessing, feature extraction, and classifier construction. Text preprocessing aims to remove noise and effectively retrieve information through text cleaning and organization [21]. Common feature extraction approaches can be loosely divided into two domains: word frequency and semantics [22,23]. Machine and deep learning models, such as k-nearest neighbors, decision trees, support vector machines, multilayer perceptron classifiers, and naïve Bayes, are widely used as classifiers [24-28].

Aim

We hypothesized that severe trauma cases could be recognized based on the content of communication between callers and call takers during emergency calls. The main research question and objective of this study was to develop a machine learning–based model through text mining of emergency calls to automatically identify severely injured patients in road accidents. We focused on road accidents instead of all trauma cases because they are the major cause of trauma, and compared with other types of injuries, the content of emergency calls for road accidents is homogeneous. As there are no suitable previous studies for comparison, our second objective was to compare the results of the model with 6 participating dispatchers' judgment.

Methods

Study Design and Setting

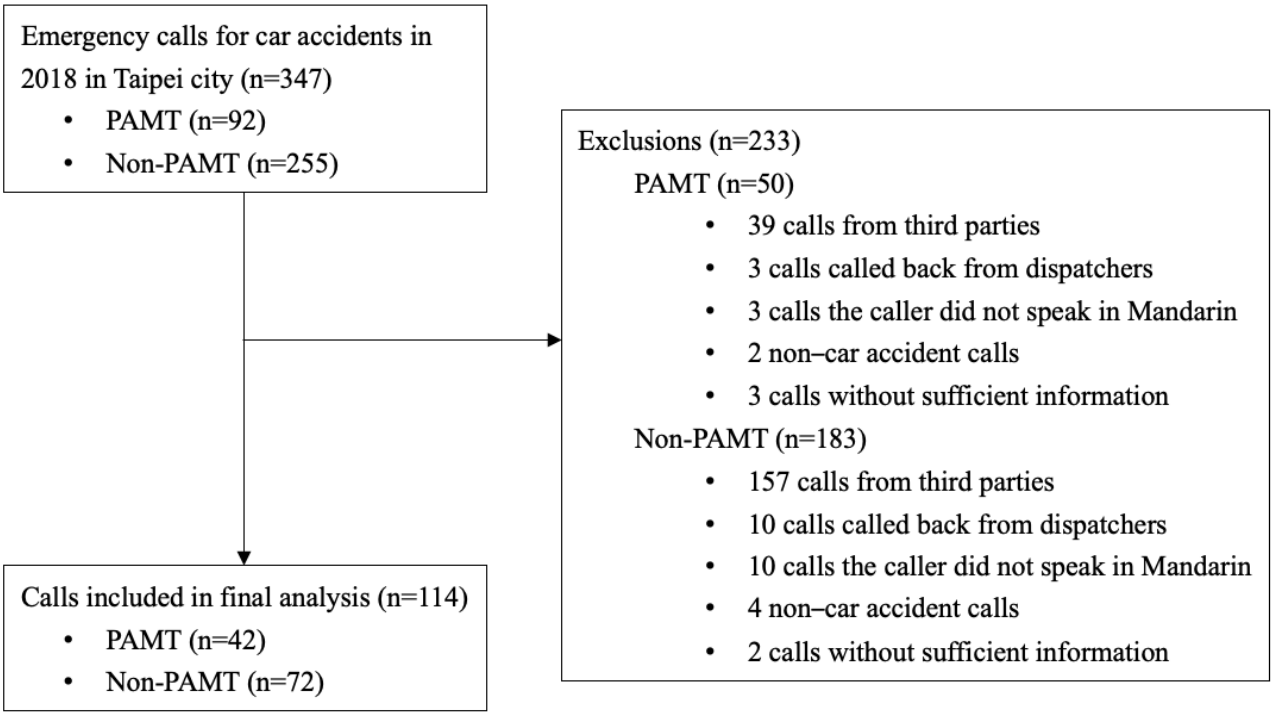
This paper describes a cross-sectional study on identifying severely injured patients in road accidents by analyzing Mandarin text of emergency calls using machine learning. The results were compared with those of human judgment. We defined severely injured patients as those who fit the major trauma criteria of the EMT trauma triage protocol, that is, prehospital-activated major trauma (PAMT).

Data Acquisition

Data were obtained from the Taipei Trauma Registry, which is a database of trauma accident information from 8 out of 18 hospitals with first aid capabilities. A random sample of one-fourth of the total cases considered as PAMT in 2018 was retrieved. After excluding cases without complete information, 92 PAMT patients (92 of 377 registered cases) were enrolled. As control cases, 3 consecutive non-PAMT road accident calls were matched with each PAMT on the same day from the dispatch system. If the number of non-PAMT cases to be matched on a given day was insufficient, only 1 or 2 calls were included. A total of 92 PAMT calls and 255 non-PAMT calls were considered in this study. The exclusion criteria were as follows: the caller was not by the side of the victim, the caller did not speak Mandarin, the accident was not vehicle-related, and the calls did not provide sufficient information. The final data for analysis included 114 cases in total, which comprised 42 PAMT and 72 non-PAMT cases (Figure 1).

XSL•FO
RenderX

Figure 1. Data acquisition and study design. PAMT: prehospital-activated major trauma.



Ethics Approval

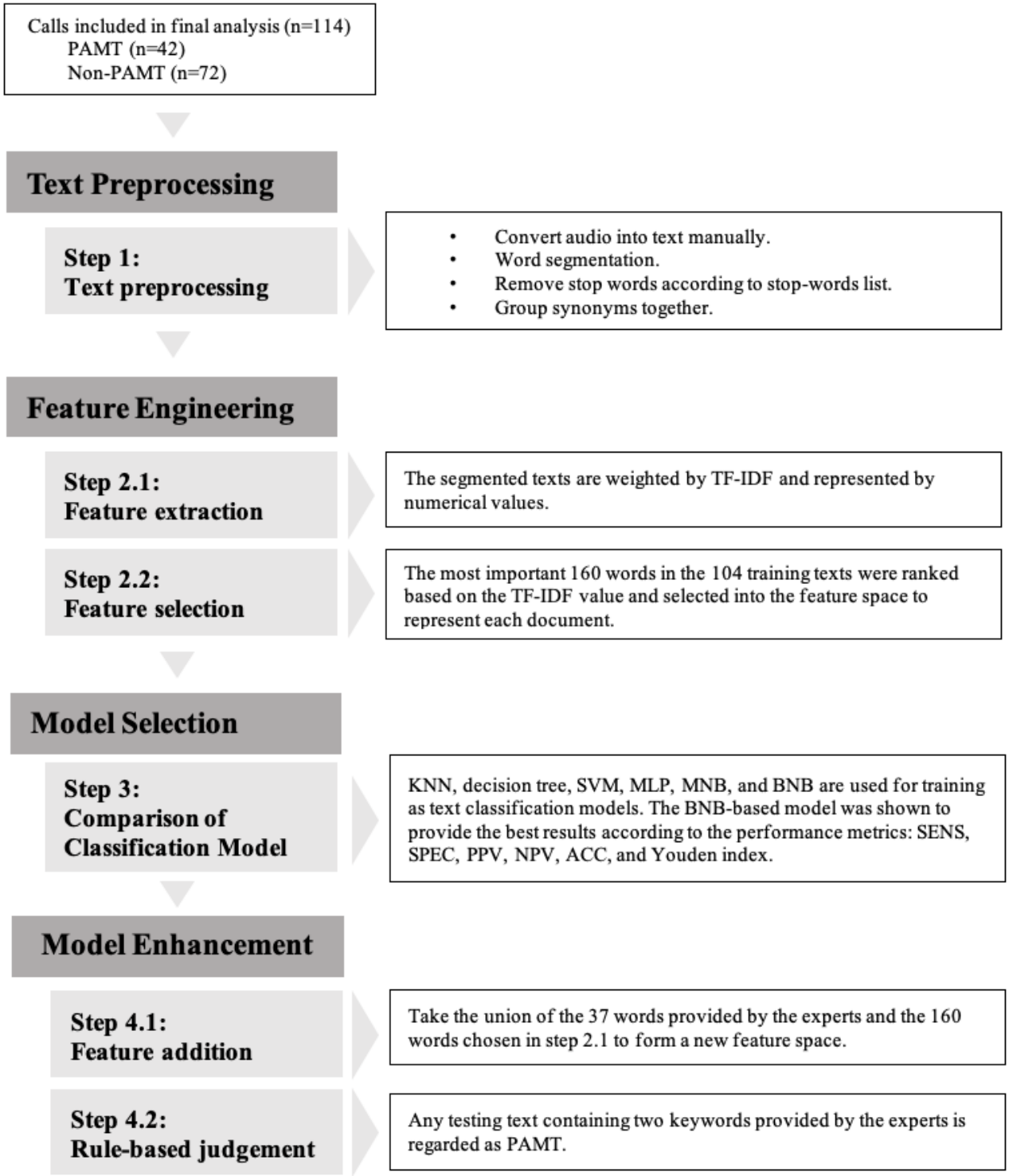
This study was approved by the institutional review board of the National Taiwan University Hospital (case number 201902043RINB).

Model Development

As shown in Figure 2, formal model development comprises four steps: (1) text preprocessing, (2) feature engineering, (3) model classification, and (4) model enhancement, which was conducted to improve model performance.

XSL•FO
RenderX

Figure 2. Model development. PAMT: prehospital activated major trauma; TF-IDF: term frequency–inverse document frequency; KNN: k-nearest neighbors; SVM: support vector machine; MNB: multinomial naïve Bayes; BNB: Bernoulli naïve Bayes; MLP: multilayer perceptron; SENS: sensitivity; SPEC: specificity; PPV: positive predictive value; NPV: negative predictive value; ACC: accuracy. Repeated random subsampling-cross validation (RRS-CV) for 100 times were performed in the step of model enhancement. All training data include 39 PAMT and 65 non-PAMT cases; testing data included 3 PAMT and 7 non-PAMT cases.



Text Preprocessing (Step 1)

The purpose of text preprocessing is to organize the data such that useful information can be retrieved. This process includes word segmentation, stop word removal, and synonym grouping (Figure 2). First, each emergency call was manually converted into a text form. The continuous text string was then segmented into words, which were the shortest units of meaning, consisting of at least one character. Segmentation was performed using the Chinese word segmentation system developed by the Institute of Information Science and the Institute of Linguistics of Academia Sinica [29,30]. To eliminate segmentation errors caused by ambiguous Chinese compound words, a dictionary of special terms with specific weights was manually constructed based on experience and trial and error. The segmentation

XSL•FO
RenderX

system refers to the weight required to force certain words to merge or separate. Subsequently, stop words were removed to remove insignificant words, such as conjunctions, pronouns, and articles. Then, synonyms were grouped and regarded as the same word, potentially reducing the model overfitting to specific words, thus providing a means for bias-variance control. From >27,000 characters in the original 114 texts, approximately 7000 different word meanings were identified.

Feature Engineering (Step 2)

In feature engineering, the segmented words were transformed into a machine-readable format by feature extraction (step 2.1, Figure 2). As emergency calls are often short, and conversations are urgent, important words are frequently mentioned (Multimedia Appendix 1). Thus, we used term frequency–inverse document frequency (TF-IDF) to weigh each word. The TF-IDF calculation consists of two sections: TF and IDF (Multimedia Appendix 2). TF illustrates the word frequency, whereas IDF explains the rarity of words appearing in the entire document. A higher frequency of occurrence of a word in one specific text indicates its importance. In contrast, a higher frequency of occurrence of the word in the entire body of texts lowers its importance. By considering these 2 frequencies simultaneously, we ranked all words by importance to conduct feature selection (step 2.2, Figure 2). The most important 160 out of 7000 words were chosen based on the experiments. The selected features were placed in a feature space to reduce the number of dimensions and to make the results more explanatory. The feature space included the selected features used to develop the model.

Model Selection (Step 3)

For model selection, we evaluated several commonly used machine learning models for text classification, including k-nearest neighbors, decision tree, support vector machine, multilayer perceptron, multinomial naïve Bayes, and Bernoulli naïve Bayes (BNB). Repeated random subsampling cross-validation (RRS-CV) was conducted 100 times to avoid overfitting and to obtain more stable and reliable classification results. RRS-CV splits samples in a randomized and repeated manner without replacement. The performance of the different models used for comparison was the average of 100 RRS-CV scores. According to Table 1, among these, the BNB-based model achieved the best results. The BNB classifier, which is a supervised learning model, is based on Bayes' theorem. It assumes that each input variable is independent of the other variables. According to the BNB equation in Multimedia Appendix 2, the calculation concentrates on binary information of whether the word appears in a document. The Boolean expression of the selected features forms the feature vector for each document. The category estimation of a document depends on the maximum a posteriori of each class *k*, which consists of the likelihood of the document being given by class *k* and its prior probability. The category with the highest maximum a posteriori labeled the classified documents. To avoid a zero-probability situation, Laplace smoothing was used to set the additive smoothing parameter to one. Consequently, no hyperparameter tuning was required for BNB. Compared with other text classification models, the BNB model has the advantages of simplicity, efficient computational speed, and ability to achieve a high level of accuracy without hyperparameter tuning. Furthermore, this model is suitable for processing small-scale data and short texts [31,32]. The results and hyperparameter tuning of other models are presented in Multimedia Appendix 3.

Table 1. Comparison of machine learning models.

Model	SENS[a] (%)	SPEC[b] (%)	PPV[c] (%)	NPV[d] (%)	ACC[e] (%)	Youden index
KNN[f]	18.7	89.0	32.6	72.1	67.9	0.077
Decision tree	32.7	76.0	35.9	72.9	63.0	0.087
SVM[g]	55.7	74.0	49.3	80.3	68.5	0.297
MNB[h]	19.0	96.1	42.2	73.8	73.0	0.151
BNB[i]	*53.0*[j]	*86.7*[j]	*67.0*[j]	*81.6*[j]	*76.6*[j]	*0.397*[j]
MLP[k]	53.7	79.0	55.6	80.6	71.4	0.327

[a]SENS: sensitivity.

[b]SPEC: specificity.

[c]PPV: positive predictive value.

[d]NPV: negative predictive value.

[e]ACC: accuracy.

[f]KNN: k-nearest neighbors.

[g]SVM: support vector machine.

[h]MNB: multinomial naïve Bayes.

[i]BNB: Bernoulli naïve Bayes.

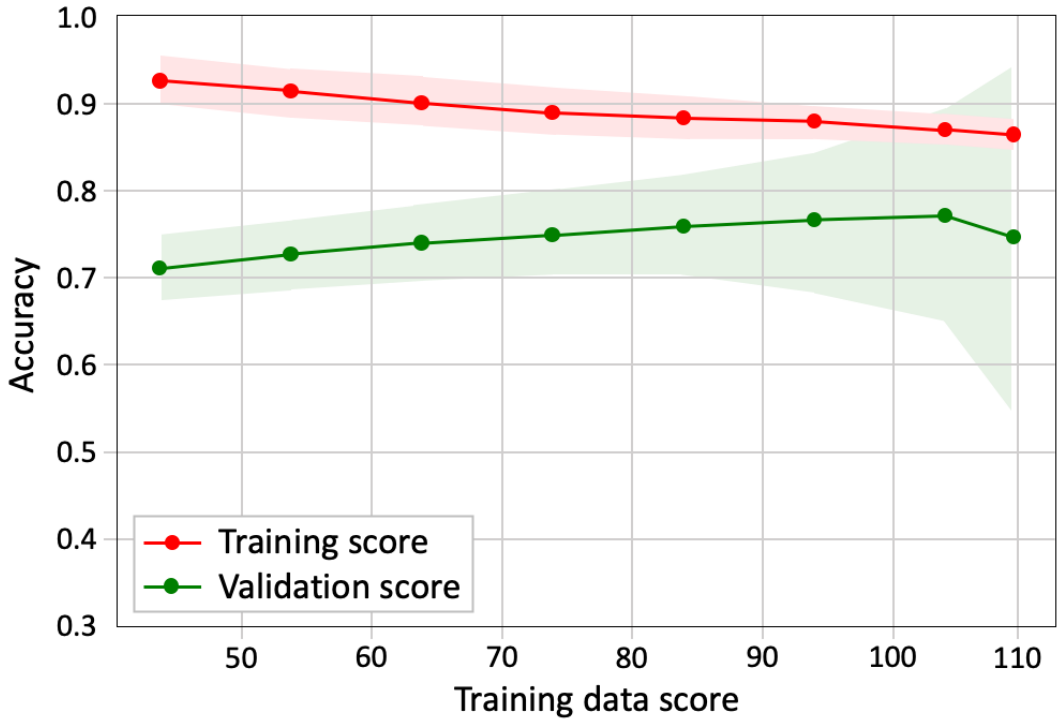
[j]BNB-based model achieved the best ACC and Youden index.

[k]MLP: multilayer perceptron.

XSL•FO
RenderX

For the split of training and validation data, we set a fixed ratio of PAMT to non-PAMT cases in the validation data. As shown in Figure 3, when the amount of training data becomes larger than that of the validation data, the training score gradually decreases and the validation score increases. The 2 lines were closest when the training and validation data sizes were 104 and 10, respectively. The convergence illustrates that, at this number of training samples, adding more training data does not significantly improve the classification performance. Therefore, for all text classification models, 104 texts were randomly selected as training data and the remaining 10 texts were used as validation data (Figure 2). The training data included 39 PAMT and 65 non-PAMT cases, and the validation data included 3 PAMT and 7 non-PAMT cases. The ground truth of model classification is the on-scene judgment of the EMT, which is presented in the form of binary labels. Figure 4 shows the scalability of the BNB-based model. As the training data increased, the model-fitting time fluctuated moderately around 0.002 seconds but significantly increased when the training data size was >104. In addition, 104 training data points with 10 validation data points had the highest validation score and the third shortest model-fitting time (Figure 5).

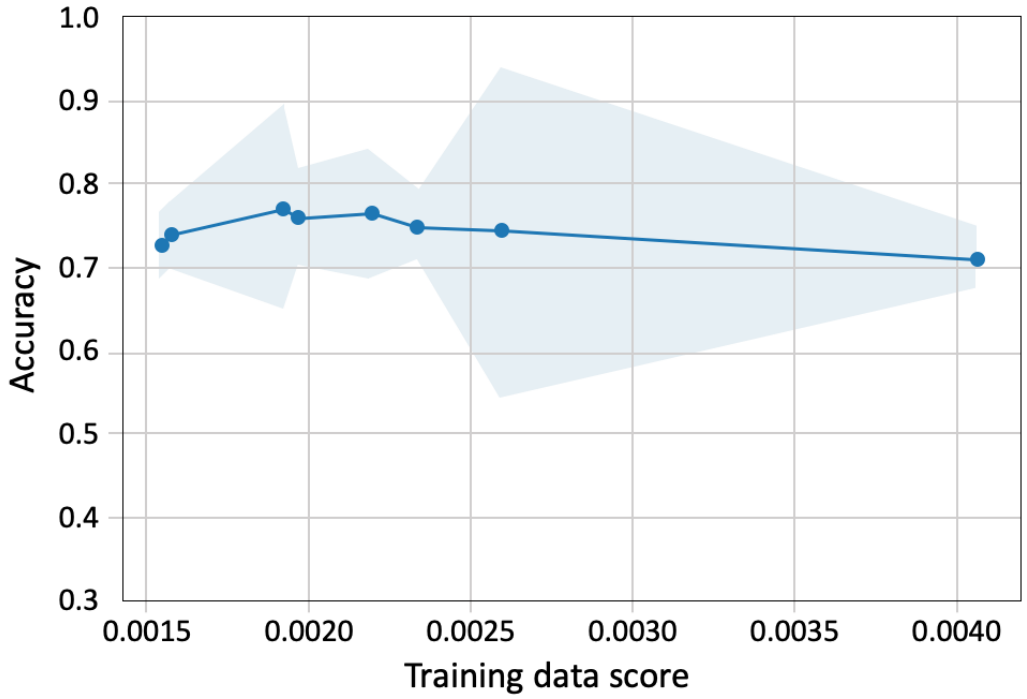
Figure 3. Learning curve of the Bernoulli naïve Bayes.



XSL•FO
RenderX

Figure 4. Scalability of the Bernoulli naïve Bayes.

Figure 5. Performance of the Bernoulli naïve Bayes.



Model Enhancement (Step 4)

To optimize the final performance of our model, we enhanced the BNB-based models using feature addition (step 4.1) and rule-based judgment (step 4.2). In feature addition, we gathered 37 keywords provided by the experts and combined them with the 160 words chosen in step 2.2 to form a new feature space (Figure 2). The experts included 6 participating dispatchers and 2 emergency physicians. After they had listened to the 114 audio recordings, they were asked, “Which keyword in an emergency call indicates whether a patient is a PAMT or non-PAMT patient?” They then provided keywords based on their personal experience. The 37 keywords were expected to expand the important feature set, which may be limited by the small amount of data. The feature space created by the union of 160 and 37 words was used to develop enhanced models. Although important features must be included, their contribution to the classification may be small if their frequencies are not significant. Therefore, a rule-based judgment (step 4.2) was designed to highlight the importance of the 37 suggested

XSL•FO
RenderX

keywords. Specifically, any text used in the validation that contained at least 2 of the 37 words provided by the experts was classified as PAMT. Texts that did not fit this rule were further examined by a BNB classifier (Figure 2).

The enhanced BNB-based model was compared with various derivative models based on combinations of different steps. The 4 derivative versions of the BNB-based model are presented in Table 2. Model A comprised manually selected features and rule-based judgment. Model B was a classical text classification model that included TF-IDF feature extraction and selection with BNB classification. Model C comprised feature engineering steps and manual feature addition with BNB classification. Finally, we named the best version as the PAMT model. It comprises steps 1 to 4.2, including text preprocessing, feature engineering, model classification, and both model enhancement approaches.

Table 2. BNB-based models of different combinations of steps.

Model	Performance						Steps included[a]				BNB[b] classification
	SENS[c] (%)	SPEC[d] (%)	PPV[e] (%)	NPV[f] (%)	ACC[g] (%)	Youden index	1	2	4.1	4.2	
Model A	54.7	82.1	56.8	80.9	73.9	0.368	✓		✓	✓	
Model B	53.0	86.7	67.0	81.6	76.6	0.397	✓	✓			✓
Model C	54.0	87.3	67.8	82.1	77.3	0.413	✓	✓	✓		✓
PAMT[h] model	68.0	78.0	60.6	85.8	75.0	0.460	✓	✓	✓	✓	✓

[a]Step 1, text preprocessing; step 2, term frequency–inverse document frequency feature extraction and selection; step 4.1, manual feature addition; step 4.2, rule-based judgment.

[b]BNB: Bernoulli naïve Bayes.

[c]SENS: sensitivity.

[d]SPEC: specificity.

[e]PPV: positive predictive value.

[f]NPV: negative predictive value.

[g]ACC: accuracy.

[h]PAMT: prehospital-activated major trauma.

Human Participants

For a reference comparison with the PAMT model, we conducted a survey to collect severe trauma judgments from 6 volunteer dispatchers. They were from the fire departments of Taipei City and New Taipei City (Table 3). The participants were asked to listen to 114 road accident audio clips. As we focused on text analysis, the participants were not allowed to receive any information other than the text. Therefore, the audio clips were transcribed into a computer-synthesized voice using a text-to-speech tool. The audio clips were played randomly in both female and male voices. In this way, the tone, speed, and emotions of the speech were neutralized. While listening to the clips, each participant classified the cases as PAMT or non-PAMT depending on their personal experience and intuition. They also shared information regarding their certainty (certain or uncertain) in each case.

Table 3. Profiles of the participating dispatchers.

Participant	Sex	Age (years), range	Service city	EMT[a] experience (year)	Dispatch experience (year)
A	Male	30-39	New Taipei City	13	6
B	Female	40-49	New Taipei City	10	2
C	Male	30-39	New Taipei City	14	1
D	Male	30-39	New Taipei City	10	1
E	Male	30-39	Taipei City	10	4
F	Male	30-39	Taipei City	9	4

[a]EMT: emergency medicine technician.

Data Analysis

The analysis determined the accuracy, positive predictive value, negative predictive value, sensitivity, and specificity of the PAMT model prediction and average judgments of the participants [33,34].

Accuracy refers to the proportion of correctly predicted PAMT and non-PAMT cases. The proportion of cases with true-predicted PAMT and non-PAMT results can be presented as positive predictive value and negative predictive value, respectively. sensitivity and specificity represent the ability of a classification system to correctly identify PAMT and non-PAMT cases, respectively. The Youden index was

XSL•FO
RenderX

calculated using different models and can be expressed as the sum of sensitivity and specificity minus 1.

All 114 cases were categorized into certainty levels from 0 to 6, depending on how many participants regarded a case as *certain*. For example, a case with certainty level 4 indicated that 4 participants were certain of their judgment, whereas the other two were not. The accuracy was also calculated for different certainty levels.

Data management and statistical analyses were performed using Python (Python Software Foundation) and Excel (Microsoft Corporation).

Results

Sample

In total, 114 patients were included in the final analysis. The transcribed texts ranged from 84 to 652 characters, with a mean of 241.4 (SD 106.7) characters; the mean character count of PAMT cases was greater than that of non-PAMT cases (266, SD 102 vs 227, SD 107). The transcribed computer-synthesized audio ranged from 24 to 145 seconds in length, with a mean of 58.9 (SD 24.5) seconds, and the mean call length of PAMT cases was longer than that of non-PAMT (64, SD 24 vs 54, SD 24 seconds) cases (Multimedia Appendix 4).

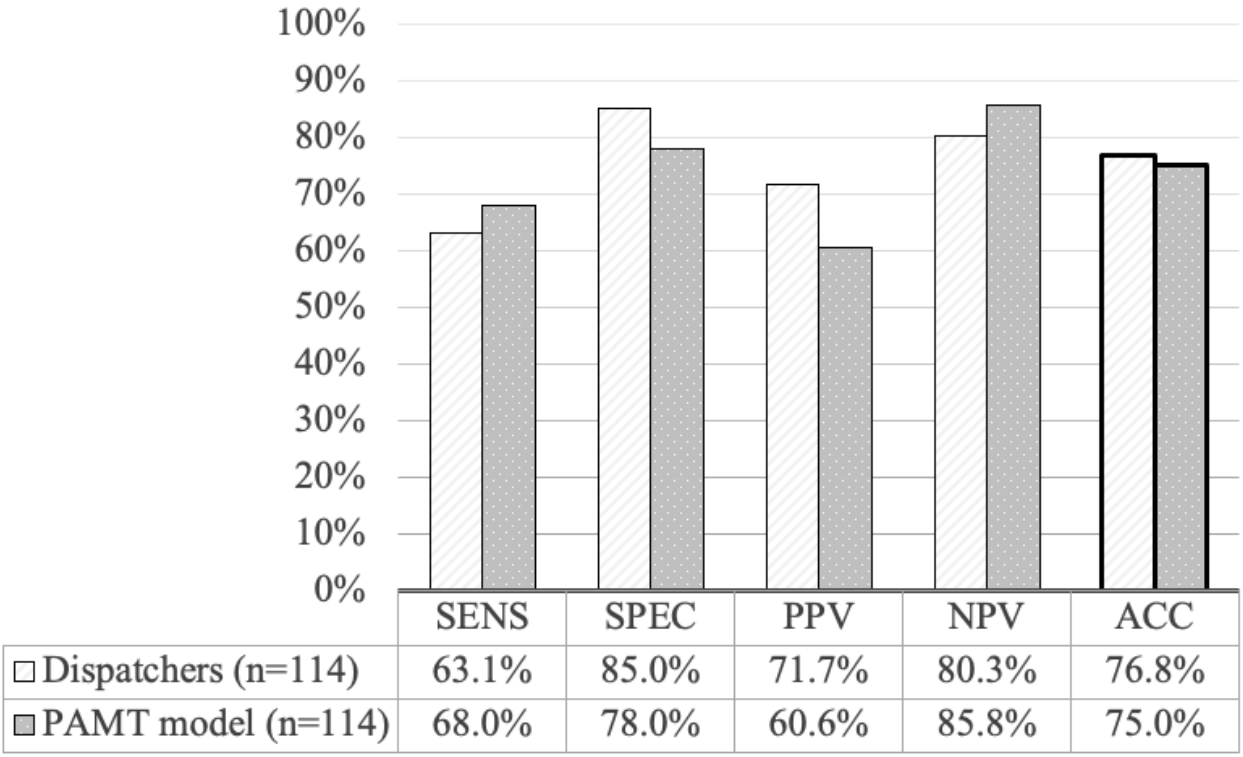
Outcome Data

In this study, the machine learning model was trained on a random sample of 104 cases and validated on the remaining 10 cases. RRS-CV was conducted 100 times to obtain greater unbiased validation results; moreover, no external data were used to test the performance of the trained models. According to Table 1, BNB outperformed the other models because it had the highest overall metrics: accuracy (76.6%) and Youden index (0.397). The mean sensitivity, specificity, positive predictive value, and negative predictive value for BNB were 53.0%, 86.7%, 67.0%, and 81.6%, respectively. As there was still room for improvement, model enhancement was performed based on BNB to increase the performance. The enhanced BNB-based model, known as the PAMT model, exhibited the best performance. Its Youden index was 0.460, and it achieved a mean sensitivity, specificity, positive predictive value, negative predictive value, and accuracy of 68.0%, 78.0%, 60.6%, 85.8%, and 75.0%, respectively (Table 2). The performance of model C, which was only enhanced by adding the features provided by the 6 volunteer dispatchers, was ranked after the PAMT model. The mean sensitivity, specificity, positive predictive value, negative predictive value, accuracy, and Youden index of model C were 54.0%, 87.3%, 67.8%, 82.1%, 77.3%, and 0.413, respectively. Model A contained only the features provided by the experts and was classified based on rule-based judgment. It achieved the worst results (sensitivity 54.7%; specificity 82.1%; positive predictive value 56.8%; negative predictive value 80.9%; accuracy 73.9%; Youden index 0.368).

In contrast, the mean sensitivity, specificity, positive predictive value, negative predictive value, and accuracy of the 6 participants were 63.1%, 85.0%, 71.7%, 80.3%, and 76.8%, respectively (Multimedia Appendix 5). The PAMT model with the best performance had a higher sensitivity and negative predictive value but a lower specificity, positive predictive value, and accuracy than the participants. Overall, the PAMT model did not surpass the performance of the participating dispatchers (Figure 6).

XSL•FO
RenderX

Figure 6. Overall performance of participating dispatchers versus prehospital-activated major trauma (PAMT) model. ACC: accuracy; NPV: negative predictive value; PAMT: prehospital activated major trauma; PPV: positive predictive value; SENS: sensitivity; SPEC: specificity.

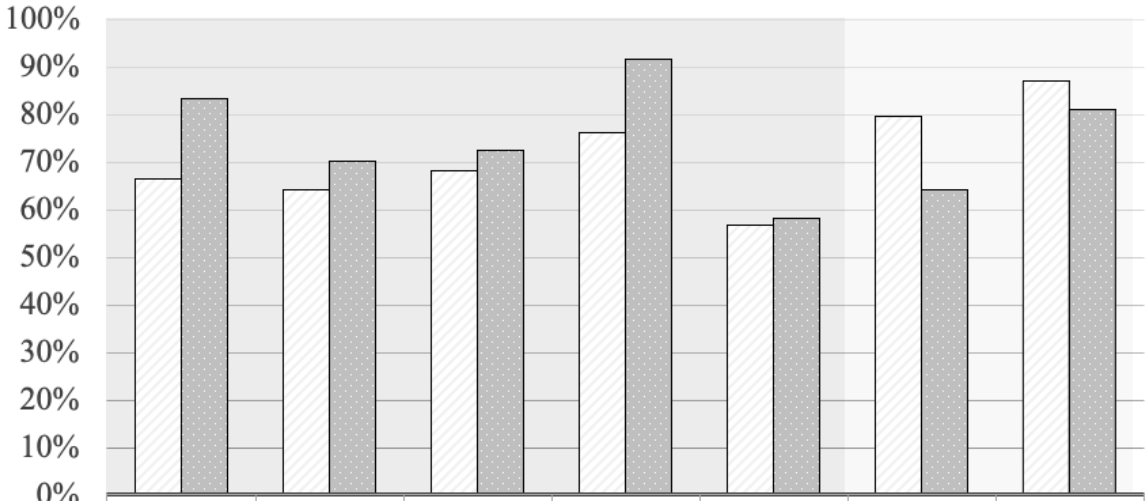


	SENS	SPEC	PPV	NPV	ACC
Dispatchers (n=114)	63.1%	85.0%	71.7%	80.3%	76.8%
PAMT model (n=114)	68.0%	78.0%	60.6%	85.8%	75.0%

In the subgroup analysis, as shown in Figure 7, the mean accuracy of the participants at certainty levels from 0 to 6 was 66.7%, 64.3%, 68.2%, 76.4%, 56.9%, 79.8%, and 87.1%. The mean accuracy of the PAMT model at certainty levels from 0 to 6 was 83.3%, 70.4%, 72.7%, 91.7%, 58.3%, 64.3%, and 81.3%. After all cases were categorized based on different certainty levels, the accuracy of the participants for levels 0 to 6 generally increased, except for level 4, whereas the accuracy of the PAMT model did not show such a linear pattern. The results of the PAMT model did not display a clear trend; that is, they were affected by the certainty level because the BNB model classified cases according to the feature distribution. If we define levels 5 and 6 as *certain cases* and levels 0 to 4 as *uncertain cases*, we can observe that, although the accuracy of the PAMT model was lower than that of the participants in *certain* cases (77.52% vs 85.48%), it was greater than the accuracy of the participants in *uncertain* cases (73.57% vs 66.34%; Figure 7).

XSL•FO
RenderX

Figure 7. Accuracy of predicting prehospital-activated major trauma (PAMT) by participating dispatchers and PAMT model over different certainty levels.



	Level 0 (n=2)	Level 1 (n=14)	Level 2 (n=11)	Level 3 (n=12)	Level 4 (n=12)	Level 5 (n=14)	Level 6 (n=49)
Dispatchers	66.7%	64.3%	68.2%	76.4%	56.9%	79.8%	87.1%
PAMT model	83.3%	70.4%	72.7%	91.7%	58.3%	64.3%	81.3%

Discussion

Principal Findings

Our study makes 3 major contributions to the field. First, this is the primary study to use a machine learning–based model to identify severely injured patients during the dispatch phase. Second, the overall performance of the model was similar to that of human dispatchers (Figure 6). Third, the model produced favorable results for cases in which dispatchers were uncertain (Figure 7).

With no suitable previous studies as a reference, we enrolled 6 volunteer dispatchers in our study. Their judgment was regarded as a reference for comparison with the models. Although such a small sample size cannot represent all dispatchers, we were still able to observe heterogeneity in human performance. As shown in Multimedia Appendix 5, three participants (A, B, and E) had a high specificity and low sensitivity, whereas the other three (C, D, and F) had more balanced figures between specificity and sensitivity. We can speculate that different experiences may affect judgment, and that the policy each participant chose, either aggressive or conservative, also made a difference. With the assistance of the proposed model, which is more stable and adjustable, it is possible to narrow the range of human discrepancies and decrease the uncertainty.

The proposed machine learning models are text classification models. As important words were repeatedly mentioned in often short and intermittent emergency calls (Multimedia Appendix 1), the frequency-based feature extraction method, TF-IDF, demonstrated the ability to select representative words in severe trauma calls. In addition, feature correlation analysis was performed for these words (Multimedia Appendix 6). Features with high correlation coefficients were words that frequently appeared together in Mandarin, or in the question required to be asked during a call. In contrast, low correlation words indicated that they appeared independently. Despite varying degrees of correlation, all the selected features are meaningful and have the potential to be keywords for judging PAMT. Therefore, the occurrence of these words was the main input for the machine learning models. Furthermore, we analyzed the length of the texts and the accuracy between the PAMT and participants. As each text was represented by a feature vector formed by word occurrences, the original length may be one of the factors affecting accuracy. Multimedia Appendix 7 presents further results.

To explore why machine learning performed better in classifying uncertain cases, we compared the words suggested by experts and the words selected by the model. Of the 37 words provided by the experts, 23 were regarded as keywords specifically for PAMT. In Multimedia Appendix 8, we compare these 23 words and the top 23 decisive words selected by the model that were most likely to occur in the PAMT texts. In the left column, most words are aggregated in the “Patient status” and “Patient basic information” categories; few are in the “Geographic information” and “Auxiliary words and other information” categories. In contrast, the words in the right column are grouped not only in the “Patient status” category but also in “Geographic information” and “Auxiliary words and other information.” This phenomenon shows that the participants focused more on the situations and injury mechanisms of the patients, whereas the proposed model was able to capture other information such as the location of an accident or wording in a conversation. In uncertain cases, there may be fewer obvious keywords for PAMT, which is possibly why the proposed model is more helpful.

In addition to the PAMT model, we tried different feature combinations and classification approaches to develop three other models. Models A, B, and C refer to manual feature

XSL•FO
RenderX

addition with rule-based judgment, TF-IDF feature engineering with BNB classification, and TF-IDF feature engineering plus manual feature addition with BNB classification, respectively (Table 2). The PAMT model consisted of steps 2.1 to 4.2. It is important to consider sensitivity and specificity while developing a triage tool; therefore, we chose the model with the highest Youden index as our final model, which was the PAMT model. The sensitivity of the PAMT model was also the highest, making it suitable for use as a triage tool.

Our results demonstrate that it is necessary to combine machine learning (steps 2.1 and 2.2) and human experience (steps 4.1 and 4.2) to develop a prehospital dispatching triage tool (Table 2). A purely manual model using the features provided by experts with rule-based judgment, such as model A, or a classical machine learning–based text classification model, such as model B, did not perform sufficiently well. Although the features of model C are composed of the TF-IDF selection and are provided by experts, without rule-based judgment to increase the importance of these keywords, it failed to outperform the PAMT model. Rule-based judgment makes the added feature of experts suggesting words more significant in classification, which is a complement of limited data. Although the best classification performance of the BNB model indicates that the occurrence of words in a call is key to identifying PAMT cases, there is currently no machine learning model that can completely replace human dispatchers.

Comparison With Prior Work

Abundant research has been conducted on field triage and prognosis prediction using prehospital data for the early recognition of severely injured trauma patients. However, the dispatching accuracy has seldom been addressed in previous studies [35,36]. The predictors used in these studies, either physiological data or injury mechanisms, were difficult to acquire through telephone calls. In the few studies regarding the accuracy of dispatching, most dealt with helicopter emergency medical services dispatching [6,37-40]. In another study, all trauma emergency calls were included and compared between clinicians and nonclinicians in a prehospital critical care team in Scotland [41]. The sensitivity of the two groups, in the study, for identifying major trauma (injury severity score>15) were 0.112 and 0.259 and the specificity was 0.998 and 0.995. Our model had a significantly higher sensitivity and lower specificity. However, the results varied as the gold standards differed. We chose the judgment of the on-scene EMT as the gold standard as it represents comprehensive prehospital information, whereas injury severity score is prognostic data that can only be obtained in the hospital. Moreover, a higher sensitivity, which avoids undertriage, allows us to apply the model as an early triage tool that can determine the priority of dispatching based on patient severity.

For a machine learning dispatch support system, a commercialized model for the recognition of OHCA through dispatching was proposed [17]. The model consists of an automatic speech recognition and textual analysis. In 2 retrospective studies conducted in Denmark and Sweden, positive results were reported in terms of both accuracy and time [18,19]. In a randomized controlled trial in Denmark, the performance of this model surpassed human recognition; however, no significant improvements were found in dispatchers' ability to recognize OHCA with model assistance [20]. Although there are numerous differences in recognizing OHCA and severe trauma, this model also uses machine learning–based text analysis. Another machine learning–based voice analysis model was proposed to recognize the emotional state of OHCA callers [42]. Although the goal differed from our approach, the study also had a small sample size, and the data source was the audio of emergency calls. It is reasonable to expect that the performance of future models may improve with a combination of semantic and emotional analyses.

Limitations

Our study had several limitations. First, human intervention is required for text preprocessing. Conversations in the recordings were often in fragments without a complete grammatical structure and contained specific terms. A customized dictionary for word segmentation, stop word removal, and synonym grouping must be constructed according to the specific medical domain, regional features, and culture. Although there are references, applying these procedures requires researchers to fully comprehend phone conversations [29]. We assumed that everyone who understood the conversation would process the audio and text materials in the same way; otherwise, the features we used in the later steps would be different. This limitation can be overcome by replacing this step with an automatic program [10]. Second, the dispatchers listened to 114 audio clips before providing the keywords. Although they did not know the answers and were asked to provide their opinions based on their experience, they might have chosen words from the audio that they had just heard. Third, owing to strict protocols and the administration's concern for this novel study concept, the recordings of the emergency calls were not allowed to be copied, and only a limited number of crews were permitted to access them over a short period. Given that the preprocessing steps were labor intensive, we could not enroll a large sample size. To compensate for this shortage, we randomly sampled the PAMTs during the entire year of 2018 in the Taipei Trauma Registry. For text classification, the main factors that affect the results may not be determined by data size [43]. Another innovative study with a small amount of data contributed to specific fields [42]. This study serves as a proof of concept and aims to reveal the potential of this methodology for target applications. Nevertheless, some advanced text classification models, such as deep learning models with semantic feature extraction methods [44], may be limited by the size and characteristics of the data. Therefore, research should be conducted on a larger scale with more participants and integrated data to develop a more mature model for actual deployment.

Conclusions

The results of our study suggest that the applied machine learning model is not superior to dispatchers in identifying road accident calls in severe trauma cases; however, the model can assist dispatchers when they lack confidence in the judgment of the calls. A study conducted on a larger scale is required for further model development and validation.

XSL•FO
RenderX

Acknowledgments

This study was funded by the Taiwan Ministry of Science and Technology (grants MOST 108-2314-B-002-130-MY3 and MOST 109-2314-B-002-154-MY2) and the National Taiwan University Hospital (grant UN110-052). The authors wish to acknowledge the 6 volunteer dispatchers, Chi-Li Liu, Ding-Chuan Wang, Jerry Lai, Shun-Wei Chung, Yu-Chi Chang, and Yu-Ju Hsu, for participating in this study. This would not have been possible without their assistance. The authors appreciate the excellent performance of emergency medical technicians and the quality assurance of the Ambulance Division and Dispatch Center of the Taipei Fire Department. Their commitment and accomplishments have significantly improved prehospital care. The authors would also like to thank the staff and Dr Chin-Hao Chang of the National Taiwan University Hospital Statistical Consulting Unit for his help with the statistics.

Authors' Contributions

KCC and YCC contributed equally as first authors, and WCC and AYC contributed equally as corresponding authors. KCC contributed to the formal analysis, visualization, and original draft. YCC contributed to the data curation, formal analysis, methodology, and original draft. JTS contributed to the conceptualization, data curation, methodology, resources, visualization, and review and editing of the manuscript. CYO contributed to data curation. CHH contributed to data curation and resources. MCT contributed to data curation and resources. MHMM contributed to funding acquisition, project administration, resources, and supervision. WCC contributed to conceptualization, data curation, methodology, visualization review, editing, and supervision. AYC contributed to conceptualization, formal analysis, funding acquisition, project administration, editing, and supervision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Example texts.
[DOCX File , 18 KB - jmir_v24i6e30210_app1.docx]

Multimedia Appendix 2
Equations and Python script used in the model.
[DOCX File , 29 KB - jmir_v24i6e30210_app2.docx]

Multimedia Appendix 3
Comparison of machine learning models.
[DOCX File , 52 KB - jmir_v24i6e30210_app3.docx]

Multimedia Appendix 4
Descriptive statistics for audio and text files.
[DOCX File , 18 KB - jmir_v24i6e30210_app4.docx]

Multimedia Appendix 5
Profiles and predictive performances of the participating dispatchers.
[DOCX File , 15 KB - jmir_v24i6e30210_app5.docx]

Multimedia Appendix 6
Feature correlation analysis.
[DOCX File , 55 KB - jmir_v24i6e30210_app6.docx]

Multimedia Appendix 7
Relation of the length of text and accuracy.
[DOCX File , 19 KB - jmir_v24i6e30210_app7.docx]

Multimedia Appendix 8
Representative keywords for prehospital-activated major trauma (PAMT) chosen by experts and the PAMT model.
[DOCX File , 23 KB - jmir_v24i6e30210_app8.docx]

XSL•FO
RenderX

References

1. Injuries and violence: the facts 2014. World Health Organization. 2014. URL: https://apps.who.int/iris/bitstream/handle/10665/149798/9789241508018_eng.pdf [accessed 2022-05-19]
2. Harmsen AM, Giannakopoulos GF, Moerbeek PR, Jansma EP, Bonjer HJ, Bloemers FW. The influence of prehospital time on trauma patients outcome: a systematic review. Injury 2015 Apr;46(4):602-609. [doi: 10.1016/j.injury.2015.01.008] [Medline: 25627482]
3. Chen CH, Shin SD, Sun JT, Jamaluddin SF, Tanaka H, Song KJ, et al. Association between prehospital time and outcome of trauma patients in 4 Asian countries: a cross-national, multicenter cohort study. PLoS Med 2020 Oct 6;17(10):e1003360 [FREE Full text] [doi: 10.1371/journal.pmed.1003360] [Medline: 33022018]
4. Drennan IR, Geri G, Brooks S, Couper K, Hatanaka T, Kudenchuk P, Basic Life Support (BLS), Pediatric Life Support (PLS) and Education, Implementation and Teams (EIT) Taskforces of the International Liaison Committee on Resuscitation (ILCOR), BLS Task Force, Pediatric Task Force, EIT Task Force. Diagnosis of out-of-hospital cardiac arrest by emergency medical dispatch: a diagnostic systematic review. Resuscitation 2021 Feb;159:85-96. [doi: 10.1016/j.resuscitation.2020.11.025] [Medline: 33253767]
5. Zhelev Z, Walker G, Henschke N, Fridhandler J, Yip S. Prehospital stroke scales as screening tools for early identification of stroke and transient ischemic attack. Cochrane Database Syst Rev 2019 Apr 09;4(4):CD011427 [FREE Full text] [doi: 10.1002/14651858.CD011427.pub2] [Medline: 30964558]
6. Bohm K, Kurland L. The accuracy of medical dispatch - a systematic review. Scand J Trauma Resusc Emerg Med 2018 Nov 09;26(1):94 [FREE Full text] [doi: 10.1186/s13049-018-0528-8] [Medline: 30413213]
7. Gianola S, Castellini G, Biffi A, Porcu G, Fabbri A, Ruggieri MP, Italian National Institute of Health guideline working group. Accuracy of pre-hospital triage tools for major trauma: a systematic review with meta-analysis and net clinical benefit. World J Emerg Surg 2021 Jun 10;16(1):31 [FREE Full text] [doi: 10.1186/s13017-021-00372-1] [Medline: 34112209]
8. Richards CT, Wang B, Markul E, Albarran F, Rottman D, Aggarwal NT, et al. Identifying key words in 9-1-1 calls for stroke: a mixed methods approach. Prehosp Emerg Care 2017;21(6):761-766 [FREE Full text] [doi: 10.1080/10903127.2017.1332124] [Medline: 28661784]
9. Riou M, Ball S, Williams TA, Whiteside A, O'Halloran KL, Bray J, et al. The linguistic and interactional factors impacting recognition and dispatch in emergency calls for out-of-hospital cardiac arrest: a mixed-method linguistic analysis study protocol. BMJ Open 2017 Jul 09;7(7):e016510 [FREE Full text] [doi: 10.1136/bmjopen-2017-016510] [Medline: 28694349]
10. Trujillo A, Orellana M, Acosta MI. Design of emergency call record support system applying natural language processing techniques. In: Proceedings of the 6th Conference on Information and Communication Technologies of Ecuador. 2019 Presented at: TIC.EC '19; November 27-29, 2019; Cuenca City, Ecuador p. 53-65 URL: https://doi.org/10.1007/978-3-030-35740-5_4 [doi: 10.1007/978-3-030-35740-5_4]
11. Choi SW, Ko T, Hong KJ, Kim KH. Machine learning-based prediction of Korean triage and acuity scale level in emergency department patients. Healthc Inform Res 2019 Oct;25(4):305-312 [FREE Full text] [doi: 10.4258/hir.2019.25.4.305] [Medline: 31777674]
12. Chen CH, Hsieh JG, Cheng SL, Lin YL, Lin PH, Jeng JH. Early short-term prediction of emergency department length of stay using natural language processing for low-acuity outpatients. Am J Emerg Med 2020 Nov;38(11):2368-2373. [doi: 10.1016/j.ajem.2020.03.019] [Medline: 32216994]
13. Sterling NW, Patzer RE, Di M, Schrager JD. Prediction of emergency department patient disposition based on natural language processing of triage notes. Int J Med Inform 2019 Sep;129:184-188. [doi: 10.1016/j.ijmedinf.2019.06.008] [Medline: 31445253]
14. Zhang X, Kim J, Patzer RE, Pitts SR, Patzer A, Schrager JD. Prediction of emergency department hospital admission based on natural language processing and neural networks. Methods Inf Med 2017 Oct 26;56(5):377-389. [doi: 10.3414/ME17-01-0024] [Medline: 28816338]
15. Lucini FR, Fogliatto FS, da Silveira GJ, Neyeloff JL, Anzanello MJ, Kuchenbecker RS, et al. Text mining approach to predict hospital admissions using early medical records from the emergency department. Int J Med Inform 2017 Apr;100:1-8. [doi: 10.1016/j.ijmedinf.2017.01.001] [Medline: 28241931]
16. Fernandes M, Mendes R, Vieira SM, Leite F, Palos C, Johnson A, et al. Risk of mortality and cardiopulmonary arrest in critical patients presenting to the emergency department using machine learning and natural language processing. PLoS One 2020 Apr 2;15(4):e0230876 [FREE Full text] [doi: 10.1371/journal.pone.0230876] [Medline: 32240233]
17. AI for Patient Consultations. Corti. URL: https://www.corti.ai/ [accessed 2021-09-12]
18. Blomberg SN, Folke F, Ersbøll AK, Christensen HC, Torp-Pedersen C, Sayre MR, et al. Machine learning as a supportive tool to recognize cardiac arrest in emergency calls. Resuscitation 2019 May;138:322-329 [FREE Full text] [doi: 10.1016/j.resuscitation.2019.01.015] [Medline: 30664917]
19. Byrsell F, Claesson A, Ringh M, Svensson L, Jonsson M, Nordberg P, et al. Machine learning can support dispatchers to better and faster recognize out-of-hospital cardiac arrest during emergency calls: a retrospective study. Resuscitation 2021 May;162:218-226 [FREE Full text] [doi: 10.1016/j.resuscitation.2021.02.041] [Medline: 33689794]
20. Blomberg SN, Christensen HC, Lippert F, Ersbøll AK, Torp-Petersen C, Sayre MR, et al. Effect of machine learning on dispatcher recognition of out-of-hospital cardiac arrest during calls to emergency medical services: a randomized clinical

XSL•FO
RenderX

trial. JAMA Netw Open 2021 Jan 04;4(1):e2032320 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.32320] [Medline: 33404620]
21. Uysal AK, Gunal S. The impact of preprocessing on text classification. Inf Process Manag 2014 Jan;50(1):104-112 [FREE Full text] [doi: 10.1016/j.ipm.2013.08.006]
22. Sparck Jones K. A statistical interpretation of term specificity and its application in retrieval. J Doc 1972 Jan;28(1):11-21 [FREE Full text] [doi: 10.1108/eb026526]
23. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. In: Proceedings of the 2013 International Conference on Learning Representations. 2013 Presented at: ICLR '13; May 2-4, 2013; Scottsdale, AZ, USA.
24. Kamath CN, Bukhari SS, Dengel A. Comparative study between traditional machine learning and deep learning approaches for text classification. In: Proceedings of the 2018 ACM Symposium on Document Engineering. 2018 Presented at: DocEng '18; August 28-31, 2018; Halifax, Canada p. 1-11 URL: https://doi.org/10.1145/3209280.3209526 [doi: 10.1145/3209280.3209526]
25. Mccallum A, Nigam K. A Comparison of Event Models for Naive Bayes Text Classification. Association for the Advancement of Artificial Intelligence. 1998. URL: https://www.cs.cmu.edu/~knigam/papers/multinomial-aaaiws98.pdf [accessed 2022-05-19]
26. Trstenjak B, Mikac S, Donko D. KNN with TF-IDF based framework for text categorization. Procedia Eng 2014;69:1356-1364 [FREE Full text] [doi: 10.1016/j.proeng.2014.03.129]
27. Lodhi H, Saunders C, Shawe-Taylor J, Cristianini N, Watkins C. Text classification using string kernels. J Mach Learn Res 2002 Feb;2:419-444.
28. Li Y, Wang X, Xu P. Chinese text classification model based on deep learning. Future Internet 2018 Nov 20;10(11):113 [FREE Full text] [doi: 10.3390/fi10110113]
29. Ma WY, Chen KJ. Introduction to CKIP Chinese word segmentation system for the first international Chinese Word Segmentation Bakeoff. In: Proceedings of the 2nd SIGHAN Workshop on Chinese Language Processing. 2003 Presented at: SIGHAN '03; July 11-12, 2003; Sapporo, Japan p. 168-171 URL: https://doi.org/10.3115/1119250.1119276 [doi: 10.3115/1119250.1119276]
30. Li PH, Fu TJ, Ma WY. Why Attention? Analyze BiLSTM deficiency and its remedies in the case of NER. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence. 2020 Presented at: AAAI '20; February 7-12, 2020; New York, NY, USA p. 8236-8244. [doi: 10.1609/aaai.v34i05.6338]
31. Venkatesh, Ranjitha KV. Classification and optimization scheme for text data using machine learning naïve Bayes classifier. In: Proceedings of the 2018 IEEE World Symposium on Communication Engineering. 2018 Presented at: WSCE '18; December 28-30, 2018; Singapore, Singapore p. 33-36 URL: https://doi.org/10.1109/WSCE.2018.8690536 [doi: 10.1109/wsce.2018.8690536]
32. Sheshasaayee A, Thailambal G. Comparison of classification algorithms in text mining. Int J Pure Appl Math 2017;116(22):425-433.
33. Chatterjee A, Gerdes MW, Prinz A, Martinez S. A comparative study to analyze the performance of advanced pattern recognition algorithms for multi-class classification. In: Proceedings of the 2020 Conference on Emerging Technologies in Data Mining and Information Security. 2020 Presented at: IEMIS '20; July 2-4, 2020; Kolkata, India p. 111-124 URL: https://doi.org/10.1007/978-981-15-9774-9_11 [doi: 10.1007/978-981-15-9774-9_11]
34. Chatterjee A, Gerdes MW, Martinez SG. Identification of risk factors associated with obesity and overweight-a machine learning overview. Sensors (Basel) 2020 May 11;20(9):2734 [FREE Full text] [doi: 10.3390/s20092734] [Medline: 32403349]
35. Sewalt CA, Venema E, Wiegers EJ, Lecky FE, Schuit SC, den Hartog D, et al. Trauma models to identify major trauma and mortality in the prehospital setting. Br J Surg 2020 Mar;107(4):373-380 [FREE Full text] [doi: 10.1002/bjs.11304] [Medline: 31503341]
36. Atiksawedparit P, Rattanasiri S, Sittichanbuncha Y, McEvoy M, Suriyawongpaisal P, Attia J, et al. Prehospital prediction of severe injury in road traffic injuries: a multicenter cross-sectional study. Injury 2019 Sep;50(9):1499-1506 [FREE Full text] [doi: 10.1016/j.injury.2019.05.028] [Medline: 31174870]
37. Giannakopoulos GF, Bloemers FW, Lubbers WD, Christiaans HM, van Exter P, de Lange-de Klerk ES, et al. Criteria for cancelling helicopter emergency medical services (HEMS) dispatches. Emerg Med J 2012 Jul;29(7):582-586. [doi: 10.1136/emj.2011.112896] [Medline: 21785150]
38. Wilmer I, Chalk G, Davies GE, Weaver AE, Lockey DJ. Air ambulance tasking: mechanism of injury, telephone interrogation or ambulance crew assessment? Emerg Med J 2015 Oct;32(10):813-816. [doi: 10.1136/emermed-2013-203204] [Medline: 25527473]
39. Coats TJ, Newton A. Call selection for the Helicopter Emergency Medical Service: implications for ambulance control. J R Soc Med 1994 Apr;87(4):208-210 [FREE Full text] [Medline: 8182675]
40. Cameron S, Pereira P, Mulcahy R, Seymour J. Helicopter primary retrieval: tasking who should do it? Emerg Med Australas 2005 Aug;17(4):387-391. [doi: 10.1111/j.1742-6723.2005.00762.x] [Medline: 16091103]

XSL•FO
RenderX

41. Sinclair N, Swinton PA, Donald M, Curatolo L, Lindle P, Jones S, et al. Clinician tasking in ambulance control improves the identification of major trauma patients and pre-hospital critical care team tasking. Injury 2018 May;49(5):897-902. [doi: 10.1016/j.injury.2018.03.034] [Medline: 29622470]
42. Chin KC, Hsieh TC, Chiang WC, Chien YC, Sun JT, Lin HY, et al. Early recognition of a caller's emotion in out-of-hospital cardiac arrest dispatching: an artificial intelligence approach. Resuscitation 2021 Oct;167:144-150. [doi: 10.1016/j.resuscitation.2021.08.032] [Medline: 34461203]
43. Ong MS, Magrabi F, Coiera E. Automated categorisation of clinical incident reports using statistical text classification. Qual Saf Health Care 2010 Dec;19(6):e55. [doi: 10.1136/qshc.2009.036657] [Medline: 20724392]
44. Lee JY, Dernoncourt F. Sequential short-text classification with recurrent and convolutional neural networks. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016 Presented at: NAACL HLT '16; June 12-17, 2016; San Diego, CA, USA p. 515-520. [doi: 10.18653/v1/n16-1062]

Abbreviations

BNB: Bernoulli naïve Bayes
EMT: emergency medical technician
OHCA: out-of-hospital cardiac arrest
PAMT: prehospital-activated major trauma
RRS-CV: repeated random subsampling cross-validation
TF-IDF: term frequency–inverse document frequency

Edited by T Leung; submitted 05.05.21; peer-reviewed by A Chatterjee, M Torii, J Hüsers; comments to author 11.08.21; revised version received 28.09.21; accepted 22.04.22; published 10.06.22.

Please cite as:
Chin KC, Cheng YC, Sun JT, Ou CY, Hu CH, Tsai MC, Ma MHM, Chiang WC, Chen AY
Machine Learning–Based Text Analysis to Predict Severely Injured Patients in Emergency Medical Dispatch: Model Development and Validation
J Med Internet Res 2022;24(6):e30210
URL: https://www.jmir.org/2022/6/e30210
doi:10.2196/30210
PMID:35687393

©Kuan-Chen Chin, Yu-Chia Cheng, Jen-Tang Sun, Chih-Yen Ou, Chun-Hua Hu, Ming-Chi Tsai, Matthew Huei-Ming Ma, Wen-Chu Chiang, Albert Y Chen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.06.2022. 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

Identifying Medication-Related Intents From a Bidirectional Text Messaging Platform for Hypertension Management Using an Unsupervised Learning Approach: Retrospective Observational Pilot Study

Anahita Davoudi[1*], MS, PhD; Natalie S Lee[2*], MD, MPH, MSHP; ThaiBinh Luong[3], PhD; Timothy Delaney[4], BA; Elizabeth Asch[5], BSc; Krisda Chaiyachati[4,5,6*], MD, MPH, MSHP; Danielle Mowery[1*], MS, PhD

[1]Department of Biostatistics, Epidemiology, and Informatics, Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA, United States

[2]Division of General Internal Medicine, Department of Medicine, The Ohio State University Wexner Medical Center, Columbus, OH, United States

[3]Penn Medicine Predictive Healthcare, University of Pennsylvania Health System, Philadelphia, PA, United States

[4]Center for Health Care Innovation, University of Pennsylvania, Philadelphia, PA, United States

[5]Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, PA, United States

[6]Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[*]these authors contributed equally

Corresponding Author:
Danielle Mowery, MS, PhD
Department of Biostatistics, Epidemiology, and Informatics
Institute for Biomedical Informatics
University of Pennsylvania
3700 Hamilton Walk
Philadelphia, PA, 19104-6021
United States
Phone: 1 2157466677
Email: dlmowery@pennmedicine.upenn.edu

Abstract

Background: Free-text communication between patients and providers plays an increasing role in chronic disease management, through platforms varying from traditional health care portals to novel mobile messaging apps. These text data are rich resources for clinical purposes, but their sheer volume render them difficult to manage. Even automated approaches, such as natural language processing, require labor-intensive manual classification for developing training data sets. Automated approaches to organizing free-text data are necessary to facilitate use of free-text communication for clinical care.

Objective: The aim of this study was to apply unsupervised learning approaches to (1) understand the types of topics discussed and (2) learn medication-related intents from messages sent between patients and providers through a bidirectional text messaging system for managing participant blood pressure (BP).

Methods: This study was a secondary analysis of deidentified messages from a remote, mobile, text-based employee hypertension management program at an academic institution. We trained a latent Dirichlet allocation (LDA) model for each message type (ie, inbound patient messages and outbound provider messages) and identified the distribution of major topics and significant topics (probability >.20) across message types. Next, we annotated all medication-related messages with a single medication intent. Then, we trained a second medication-specific LDA (medLDA) model to assess how well the unsupervised method could identify more fine-grained medication intents. We encoded each medication message with n-grams (n=1-3 words) using spaCy, clinical named entities using Stanza, and medication categories using MedEx; we then applied chi-square feature selection to learn the most informative features associated with each medication intent.

Results: In total, 253 participants and 5 providers engaged in the program, generating 12,131 total messages: 46.90% (n=5689) patient messages and 53.10% (n=6442) provider messages. Most patient messages corresponded to BP reporting, BP encouragement, and appointment scheduling; most provider messages corresponded to BP reporting, medication adherence, and confirmatory statements. Most patient and provider messages contained 1 topic and few contained more than 3 topics identified using LDA.

XSL•FO
RenderX

In total, 534 medication messages were annotated with a single medication intent. Of these, 282 (52.8%) were patient medication messages: most referred to the medication request intent (n=134, 47.5%). Most of the 252 (47.2%) provider medication messages referred to the medication question intent (n=173, 68.7%). Although the medLDA model could identify a majority intent within each topic, it could not distinguish medication intents with low prevalence within patient or provider messages. Richer feature engineering identified informative lexical-semantic patterns associated with each medication intent class.

Conclusions: LDA can be an effective method for generating subgroups of messages with similar term usage and facilitating the review of topics to inform annotations. However, few training cases and shared vocabulary between intents precludes the use of LDA for fully automated, deep, medication intent classification.

International Registered Report Identifier (IRRID): RR2-10.1101/2021.12.23.21268061

(J Med Internet Res 2022;24(6):e36151) doi:10.2196/36151

KEYWORDS

chatbots; secure messaging systems; unsupervised learning; latent Dirichlet allocation; natural language processing

Introduction

Background

Digital health technology has enabled patient-clinician communication to move beyond the confines of a face-to-face clinician-patient encounter. For years, the patient portal has been the primary route of electronic patient communication with providers [1-5]. Increasingly, there are mobile health apps that also enable communication between patients and providers through SMS, particularly for management of chronic conditions like hypertension [6,7]. SMS platforms add to the increasing number of channels for facilitating low-threshold, frequent, asynchronous communication between patients and clinical teams. However, enhancing communication pathways also adds burden to clinical teams, which are often already struggling to manage patient message volume through more conventional paths, such as the patient health portal. These pathways can be a source of clinician burnout due to technostress, time pressure, and workflow-related issues [8,9].

Despite the importance of patient message data for clinical care, the current state for clinical review of patient message data is largely manual [10,11]. Natural language processing (NLP) and machine learning (ML)–based systems could facilitate triaging of messages to the right clinical sources (providers, nurses, medical assistants, billing, etc) for appropriate decision-making [12], thereby minimizing the burden of messages on already-strained personnel.

Natural Language Processing and Digital Health Technology

There are some promising examples in the literature of NLP and supervised ML approaches to more efficiently filter and review messages for clinical purposes [13-15]. Chen et al [15] developed HypoDetect (Hypoglycemia Detector) to automatically identify hypoglycemia incidents within messaging threads reported by US veterans with diabetes through SMS. They trained and tested three supervised ML algorithms to classify each thread as containing a hypoglycemia incident (positive) or not (negative). Liu et al [14] trained and tested a deep learning approach—Longformer-masked language models using Hugging Face—to classify conversation stages of messages and behaviors present in messages from both texters and volunteers. Stenner et al [13] developed PASTE (Patient-Centered Automated SMS Tagging Engine), a rule-based NLP system for encoding medication-related messages from MyMediHealth, which is a medication management system for scheduling and administering medications and sending reminders to patient cell phones. Although all three examples cited are promising ways to harness NLP and ML for clinical patient-facing applications, a large amount of labeled data is necessary to train and test robust NLP or ML models. Stenner et al [13] focused narrowly on NLP for medication-related concepts and, therefore, used existing libraries of annotated data. However, this strategy is not possible for identification of novel intents. Thus, in the cases of HypoDetect and Shout, which represented the application of ML to novel topics, researchers manually annotated 3000 and 8844 messages, respectively, to begin training their models.

Unsupervised ML approaches have been shown to subgroup texts with similar word usage and could reduce the burden of manual annotation for training clinical models for intent classification. Therefore, we investigated the utility of unsupervised methods for deriving message intents, in particular, latent Dirichlet allocation (LDA). LDA is an unsupervised, generative statistical model for learning subgroups of observations within a data set based on similarities among observations [16]. LDA has been leveraged to derive insights into patient communication data in patient portals [17,18], but its usage to classify message intent, particularly in mobile text messaging technologies, has been largely unexplored. The purpose of this proof-of-concept study was to explore the extent to which LDA might be applied to patient-provider messages to accurately identify domains (ie, topics) of clinical relevance (ie, those topics that were deemed informative for clinical action that are indicative of intents). In this study, we applied LDA to a database of patient-provider messages exchanged in a mobile app designed for remote hypertension monitoring and examined the usability of LDA-derived topic classes for identifying and classifying novel intents.

Methods

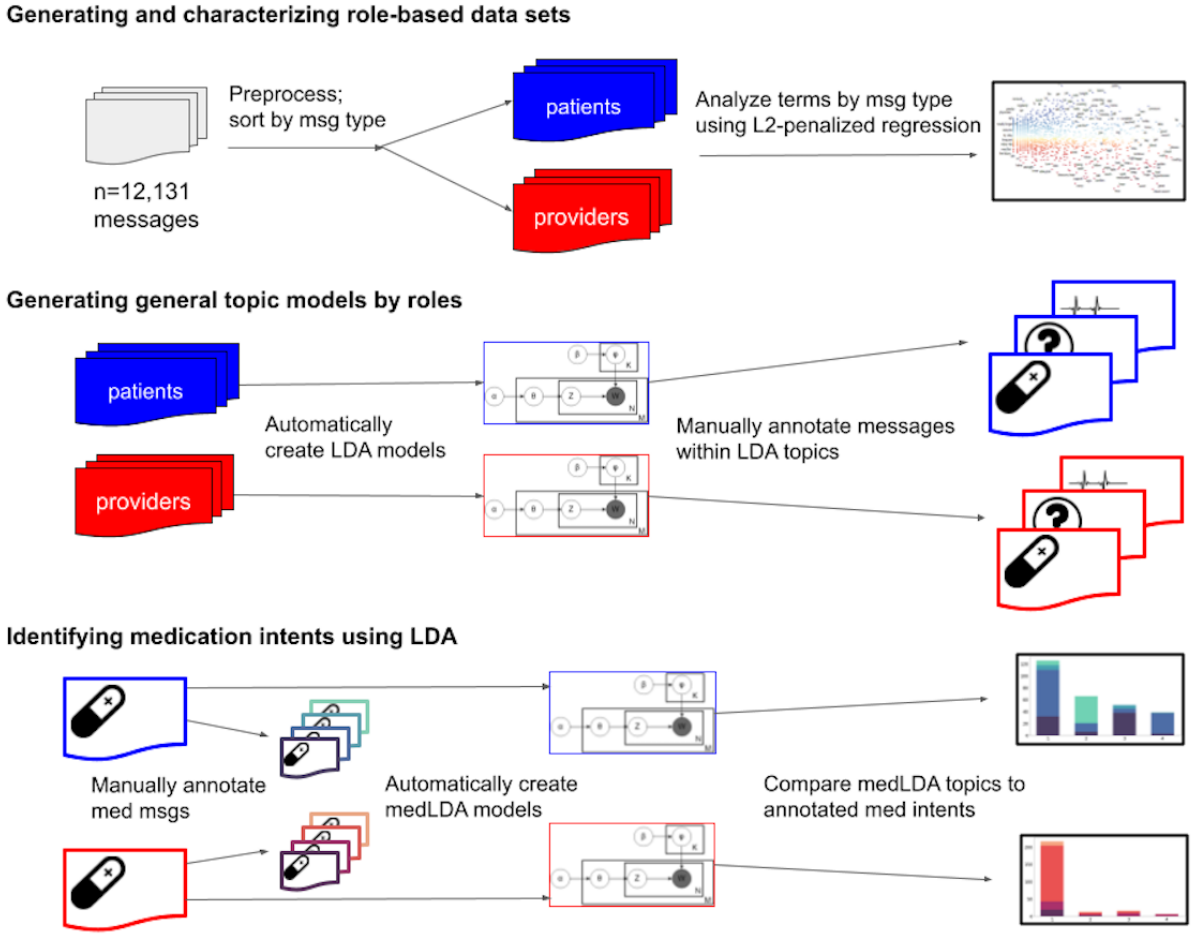
Overview

In this study, we retrospectively studied messages exchanged through a third-party mobile app. Participants were adults enrolled in Penn Medicine's Employee Hypertension

XSL•FO
RenderX

Management Program (eHTN) from 2015 to 2019. To protect the privacy of our study participants, all data were deidentified using a text deidentification system called De-Id prior to the study and the analyses [19]. In Figure 1, we outline our methodology, and we describe our analytical framework in subsequent sections.

Figure 1. Study workflow. LDA: latent Dirichlet allocation; med: medication; medLDA: medication-specific LDA model; msg: messages.



Ethics Approval

This study was approved by the Institutional Review Board at the University of Pennsylvania (approval No. 834667).

Employee Hypertension Management Program

The eHTN is a primarily remote hypertension management program for Penn Medicine employees with uncontrolled hypertension. Through the eHTN, employees diagnosed with uncontrolled hypertension at an initial office visit receive a prescription medication, a treatment plan for blood pressure (BP) management, and a BP cuff for home-based measurements. A critical component of the program is out-of-office communication of BP readings, typically once every 2 weeks, plus unlimited bidirectional text message conversation between the clinical team and patients for issues related to BP management. These patient-clinician interactions were facilitated through a proprietary Health Insurance Portability and Accountability Act–compliant, bidirectional text messaging mobile app. Topics discussed reflected the spectrum of issues related to hypertension management, including questions about BP, medication-related questions or complaints, refill requests, and BP machine or cuff issues, as well as questions about the app itself. There were no technical restrictions to the conversation, though content that was less relevant to hypertension management (eg, a statin medication refill request) was generally redirected by the BP clinical team. Conversations could be initiated by patients (eg, a medication question) or by clinicians (eg, to request an updated BP reading). Our study period was the duration of app usage, from June 2015 to November 2019.

Generating and Characterizing Role-Based Data Sets

To automatically identify topical themes by the roles of patients and providers, we classified messages into two data sets: inbound (ie, patient-generated messages) and outbound (ie, provider-generated messages). For each data set, we removed messages that appeared to be automatic messages generated by the app, such as patient enrollment (eg, “Registered to use the PROGRAM app”) or calendar events (eg, “Calendar event created: Plan Check-In”). Among inbound messages, patients could report BP readings as structured data elements (eg, “Annotation: Pulse 76”) and could add contextual information as free text (eg, “left arm”): for example, “Annotation: Pulse 95. Ran down the steps.” We removed all app-specific prefixes (eg, “Annotation:”) from the messages, but kept the remaining parts of the text messages in the model. Next, we preprocessed each message by changing words to lowercase (eg, “Meds”

XSL•FO
RenderX

reduced to “meds”), removing stop words (eg, “of” and “the”), and stemming terms (eg, “scheduling” stems to “schedul”). To better understand commonality and differences between the patient and provider messages, we analyzed and visualized individual words between data sets according to their frequency and informativeness. We computed and visualized the word frequencies according to their scaled L2-penalized regression coefficients using scattertext [20].

Generating LDA Models by Roles: Topics and Subtopics

Next, we aimed to automatically learn topics dispersed within each message data set. For each processed data set (ie, patient and provider independently), we applied LDA. LDA is an unsupervised, generative statistical model for learning subgroups of observations within a data set based on similarities among observations [16]. We leveraged LDA to identify subgroups of messages with similar term usage and derive topics that might correlate to high-value intents hypothesized a priori (eg, medication reorders and appointment scheduling requests). To generate useful LDA models, we experimented with hyperparameters of α and β by varying their values from .01 to 1, their symmetry parameters, and the number of topics from 5 to 75 [21]. Our goal was to optimize the topics such that they should be precise, but the words that compose a topic can be diverse and be comprised of most words in the corpus. After manual inspection, the parameters for each model were set as follows to provide the most precise and diverse range of semantically coherent topics: α was set as asymmetric to ensure that document-topic density would result in more specific topic distributions per document; β was set as symmetric to ensure that word-topic density would result in less specific word distributions per topic. We also limited the number of topics to 50 after observing that the exact composition of terms listed across models with topics over 50 were identical with near zero weights, suggesting that LDA was unable to identify any additional distinct topics beyond 50 topics. In LDA models, each unit of analysis (eg, each message) is assigned LDA-derived topics with associated probabilities. For each message, the probabilities of each topic sum to 1; therefore, a message could have one or more significant subtopics. For example, a message may be assigned LDA topics 1, 2, and 5 with high probabilities associated with each topic. We defined a *primary topic* as the topic with the highest probability; a *secondary subtopic* was defined as any topic with a probability equal to or greater than .20. We chose a threshold of .20 based on the observation that most messages had single-digit probabilities or ones that were close to 0, and when messages contained multiple themes they often coincided with probabilities greater than .20. Note that a main topic can have a probability of less than .20 because some messages may or may not have a significant topic (ie, above .20 probability). Additionally, a given message could be assigned to two or more topics if both topics exceeded a probability of .20 (eg, if a message has a probability of including topic 15 of >.20 and topic 37 of >.20, the message will appear in both topics 15 and 37). For each message, we identified both primary and secondary topics.

Comparing LDA-Generated Topics With Manually Derived Intents

To explore the clinical validity of the LDA-derived topics, we manually annotated a subset of messages with their intents, which is the term used to describe the goal or main idea of the text. The codebook for manual annotation was loosely based on an annotation schema from prior work on clinician-patient text communication [22]; briefly, two research team members (TL and NL) developed a common annotation codebook for messages exchanged through a different text message–based platform for remote hypertension management. The codebook was refined by review of over 1200 text messages exchanged in the program between October 1, 2020, and January 31, 2021; the review was conducted by our team members (TL and NL) for a separate internal pilot study. The codebook was further refined based on iterative discussion and review of the study text data by three team members (TL, NL, and DM). We attempted to apply this schema to each LDA-derived topic, but determined that the patient messages still contained heterogeneity (ie, 1 topic does not correlate to a single intent) and variability of intents.

Given these factors, we attempted to apply LDA to a more limited data set. We focused on the short messages that had a single LDA intent, occurred frequently, and appeared clinically useful; we chose medication-related intents for this last category, given the clinical importance of identifying medication-related communication. Each of three reviewers (TL, DM, and NL) manually reviewed the data set and annotated only those messages with a single intent related to medication; any messages that were deemed to have more than one intent were excluded from review, with the exception of messages where the second intent was a pleasantry. All medication messages were reviewed among the team to resolve disagreements through consensus. For all messages annotated with a single medication intent, we generated another medication-specific LDA (medLDA) model and attempted to reclassify the resulting messages according to the set k topics: k=4 for both patient and provider messages, based on manual annotation. Across each k topic, we report the majority class intent for that topic number and apply the heuristic of classifying each message within a particular topic to the majority class intent. We report the recall, precision, and F1 score of applying this heuristic for classifying each medication intent class [13,23].

Visualizing Sublanguage of Medication Intents

We aimed to automatically capture the sublanguage of medication intents by identifying the most significant language features for each medication intent. To identify lexical features, each text message was preprocessed using spaCy by removing stop words, reducing case, and encoding n-grams. We applied term frequency–inverse document frequency to extract the most informative lexical features. To identify semantic features, we encoded the named entities of problems, treatments, and tests using the i2b2 (Informatics for Integrating Biology and the Bedside) named entity recognition (NER) model from the Stanza package in Python (version 3; Python Software Foundation) [24]. To standardize medication-related details, we encoded RxNorm categories and semantic medication categories of drug

XSL•FO
RenderX

name, strength, route, frequency, form, dose amount, intake time, duration, dispense amount, refill, and necessity using the MedEx package [25]. Examples of semantic features can be found in Table 1. The stop words were removed in the preprocessing step, and question marks were represented by the word "question." Within each subclass, we computed the word frequencies (n=1-3 words) and identified the most informative words using chi-square feature selection. We selected the lexical and semantic features (ie, n-grams, Stanza NER, and MedEx) that were most significantly associated with each medication intent ($P<.05$). We report the sublanguage features associated with each medication intent.

Table 1. Examples of lexical and semantic features.

Package type and category	Example[a]
Stanza	
Problem	"Musinex, proair inhaler and delsym for *cough*"
Test	"Took meds between 8:30 and 9:00 am after my *MRI*[b]"
Treatment	"*Meds*[c] taken late"
MedEx[d]	
Drug product name (DPN)	"Sorry Lisinopril was stopped *HCTZ*[e] 25 mg added / daily Metoprolol decreased to 50 mg / daily."
Drug ingredient (DIN)	"Sorry *Lisinopril* was stopped HCTZ 25 mg added / daily *Metoprolol* decreased to 50 mg / daily."
Drug brand name (DBN)	"Hi have only 11 tablets of *amlodipine* left. Are you able to issue me a new prescription?"
Drug dose form (DDF)	"Dr doubled my Lisinopril and removed the water *pill* See OV[f]"
Dose (DOSE)	"Sorry Lisinopril was stopped HCTZ *25 mg* added / daily Metoprolol decreased to *50 mg* / daily."
Dose amount (DOSEAMT)	"Hi have only *11 tablets* of amlodipine left. Are you able to issue me a new prescription?"
Frequency (FREQ)	"Sorry Lisinopril was stopped HCTZ 25 mg added / *daily* Metoprolol decreased to 50 mg / *daily*."
Route (RUT)	"Lossrtan is giving me SOB[g] chest pressure on *inhalation* is this l? Happens about 20 min after I take it and lasts thru day."
Duration (DRT)	"Lossrtan is giving me SOB chest pressure on inhalation is this l? Happens about *20 min* after I take it and lasts thru day."
Dose unit	"Good Morning can you call me in a refill for my bp[h]*pills* please?"

[a]Keywords are italicized.
[b]MRI: magnetic resonance imaging.
[c]meds: medications.
[d]The MedEx semantic type for each term is included in parentheses.
[e]HCTZ: hydrochlorothiazide.
[f]OV: office visit.
[g]SOB: shortness of breath.
[h]bp: blood pressure.

Results

In this retrospective observational study, we studied text messages exchanged through a third-party mobile app between 253 participants who were enrolled in the eHTN and their clinicians. Of the patients who participated, 96.0% (243/253) were actively engaged in the program, sending at least one inbound message. Of the total 12,131 messages collected, 46.90% (n=5689) were generated by patients and 53.10% (n=6442) were generated by providers (Table 2).

Using word frequency and L2-penalized regression coefficients, we identified distinct language use by role (ie, patient and provider; Figure 2). Within the patient messages (blue), we observed terms with higher coefficients and higher frequency, including temporal expressions (eg, "morning," "evening," "tonight," "gm" [good morning], and "hr" [hour]), medical terms (eg, "rx" [prescription] and "pulse"), and confirmations (eg, "okay" and "thx" [thanks]). Within the provider messages (red), we observed terms with lower coefficients and higher frequency including salutations (eg, "mrs" and "mr"), directive verbs (eg, "check record," "sign," "send," "confirm," "recheck," and "look"), and positive and negative sentiment (eg, "nice," "great," and "worry").

When we applied LDA to the patient and provider text data sets, we observed a broad distribution of messages, with certain topics occurring more frequently for both patients and providers (Figure 3 and Table 3). In Figure 3, among the 5689 patient messages, the majority of messages occurred within topic 1 (n=1117, 19.63%; eg, "thank," "ok," "great," and "nice"), topic 17 (n=412, 7.24%; "pulse," "mg," "daili" [daily], "take," "tab" [tablet], "dose," "losartan," and "amlopidin" [amlodipine]), and topic 7 (n=395, 6.94%; "bp," "read," "hi," "record," "check,"

XSL•FO
RenderX

and “today”). Among the 6442 provider messages, the majority of messages occurred within topic 47 (n=1249, 19.39%; “record,” “please,” “bp,” “update,” “read,” “hi,” and “check”), topic 12 (n=665, 10.32%; “good,” “look,” “work,” “bp,” “keep,” and “great”), and topic 42 (n=419, 6.50%; “call,” “schedul” [schedule], “hi,” “appt” [appointment], “please,” “appoint” [appointment], “come,” and “see follow” [see at follow-up]).

In Table 3, among 5689 patient messages, the majority were assigned 1 significant topic (n=2851, 50.11%), but it was also common for messages to be more heterogeneous, with 2 (n=1893, 33.27%) or 3 (n=564, 9.91%) co-occurring primary and secondary topics (Figure S1 in Multimedia Appendix 1). Similar heterogeneity was observed among provider messages; the majority of the 6442 messages were assigned to 1 topic (n=3311, 51.40%), but many had 2 (n=2466, 38.28%) or 3 (n=503, 7.81%) co-occurring topics (Figure S2 in Multimedia Appendix 1).

Table 2. Statistics according to patient and provider messages.

Message type	Patient messages (n=5689), mean (SD)	Provider messages (n=6442), mean (SD)
Words per message	17.01 (23.40)	26.75 (28.79)
Sentences per message	2.71 (2.02)	3.11 (2.11)
Messages per user	23.84 (26.22)	521.83 (1588.20)

Figure 2. Characteristics of messages shown with a scatterplot image using word frequency and L2-penalized regression coefficients. Terms with higher usage are colored according to patients (blue) and providers (red). Terms with intermediate colors, such as green, yellow, and orange, reflect coefficients with values that have less of an association with patient or provider usage. Coef: coefficient; Reg: regression.

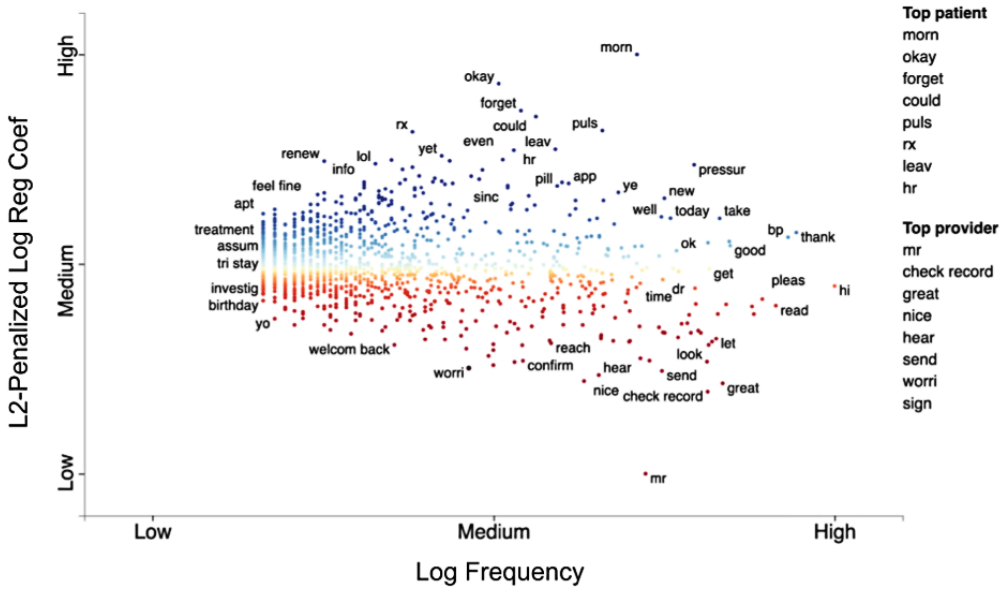
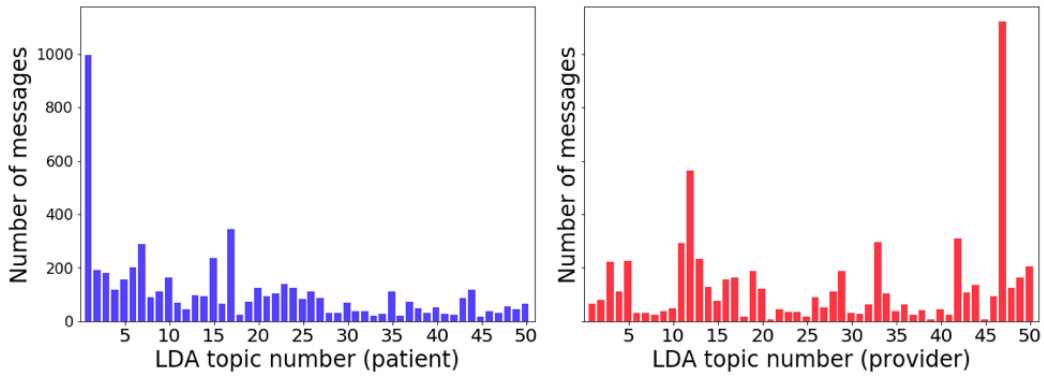


Figure 3. Distribution of patient (left) and provider (right) messages according to major topics. LDA: latent Dirichlet allocation.



XSL•FO
RenderX

Table 3. Distribution of patient and provider messages according to shared significant subtopics within each main topic.

Number of LDA[a] topics by data set	Messages, n (%)
Patient (n=5689)	
1	2851 (50.11)
2	1893 (33.27)
3	564 (9.91)
4	49 (0.86)
5	0 (0)
Provider (n=6442)	
1	3311 (51.40)
2	2466 (38.28)
3	503 (7.81)
4	22 (0.34)
5	0 (0)

[a]LDA: latent Dirichlet allocation.

In Table 4, we depict the distribution of messages that were manually classified according to medication intent. Among the 282 patient medication messages, the intent of the majority of messages was medication request (n=134, 47.5%), followed by medication taking (n=79, 28.0%) and medication location (n=54, 19.1%). Among the 252 provider medication messages, the intent of the majority of messages was medication question (n=173, 68.7%), followed by medication question response (n=41, 16.3%). We observed lexical and semantic sublanguage features associated with each medication intent category according to the patient and provider data sets. Among the patient medication intents, for the medication location intent, terms associated with drug dispensaries (eg, “pharmacy” and “apothecary”), hospitals (eg, “hup” [Hospital of the University of Pennsylvania]), and street locations (eg, “Spruce” and “Market”) were common. For the medication question intent, we observed semantic types associated with MedEx drug names (eg, “DPN” [drug product name], “DBN” [drug brand name], “DDF” [drug dose form], and “DIN” [drug ingredient]), course (eg, “start” and “stop”), use of a question mark, and Stanza problem entity. The medication request and medication-taking intents commonly included MedEx categories (eg, “DOSE” and “FREQ” [frequency]), terms for refills and verbs (eg, “need” and “taking”), and Stanza treatment entities. Among the provider medication intents, for the medication change intent, terms associated with temporal expressions (eg, “tomorrow” and “week”), MedEx categories (eg, “DBN” and “DOSE”), and program references were common. Among the medication question and medication refill question intents, we observed terms associated with refill requests (eg, “refill,” “refilled,” and “need”), use of a question mark, and Stanza treatment entity. Among the medication question response intents, we observed terms associated with references and change (eg, “baseline” and “increasing”) as well as side effects.

In Figure 4, we depict the outcomes of applying the medLDA model for both patient and provider messages. Among the patient messages manually classified according to medication intent and automatically classified within a topic, we observed a majority medication intent within each topic: medication request (topic 1), medication location (topic 2), medication taking (topic 3), and medication request (topic 4). Among the provider messages, the majority of medication intents within each topic were medication question (topics 1 and 2) and medication question response (topics 3 and 4).

In Table 5, by applying a majority intent class heuristic to classify each message within a topic number, we computed performance for predicting each medication intent category by message type. For patient messages, we observed high recall and moderate precision for medication location (recall=0.833; precision=0.682) and medication request (recall=0.843; precision=0.685). In contrast, we observed moderate recall and high precision for medication taking (recall=0.481; precision=0.745). For provider messages, we observed excellent recall and high precision for medication question (recall=0.965; precision=0.726). Conversely, we observed low recall and moderate precision for medication question response (recall=0.342; precision=0.636). All other classes could not be predicted with this approach.

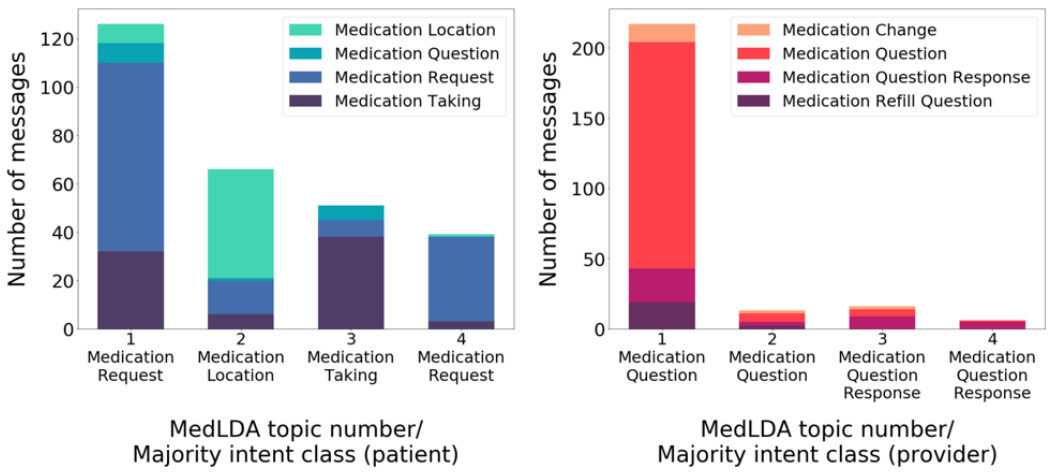
XSL•FO
RenderX

Table 4. Distribution of medication intent categories with examples from patient and provider messages.

Message type and medication intent category	Messages, n (%)	Example message[a]	Sublanguage features[a]
Patient (n=282)			
Medication request	134 (47.5)	"Yes I am. Sent in a new prescription for the 10 mg when we changed the dosage because I *needed* to *refill* my pills."	*need*, taking, *refill*, dose, script
Medication taking	79 (28.0)	"Sorry Lisinopril was *stopped HCTZ*[b]*25 mg* added / daily Metoprolol decreased to 50 mg / *daily 25 mg started* today"	taking, *dose_freq*[c], doseamt[d], *din*[e]*_dose*, *started*, *stopped*, dose_treatment
Medication location	54 (19.1)	"You sent it to *apothecary* at 3737 *market st*?"	*apothecary_market_st*, pcam[f]_pharmacy, pah[g]_pharmacy, hup[h]_pharmacy, cvs, ravdin[i]
Medication question	15 (5.3)	"So at what point would / should I start 5 mg of amlodipine or another *drug?*"	*dpn*[j]*_question*, night, feeling, really_tired
Provider (n=252)			
Medication question	173 (68.7)	"Hi - we got your *refill* request - have you been taking your blood pressure medicine everyday?"	need_refilled, *refill*, *question*, taking
Medication question response	41 (16.3)	"I have not heard of amlodipine causing *loose stool* - if anything very rarely it can cause some *constipation*."	typical_side_effects, *side_effect*, feel, effect_din, morning
Medication refill question	21 (8.3)	"Do you need refills on anything? do you need the enalipril refilled too? ok what do you *need refilled*?"	refill, refill_needed, meds[k]_need, refill_test refilled_treatment, *need_refilled*
Medication change	17 (6.7)	"Hi talked to Dr [**NAME**] *lets increase* your amlodipine to 10mg and see what your readings are like in a couple of weeks...."	see_tomorrow, dose_question, week, *lets_increase*, dpn_dose_program

[a]Italics indicate encoded features identified and shared by the example sentence and sublanguage features.
[b]HCTZ: hydrochlorothiazide.
[c]freq: frequency.
[d]doseamt: dose amount.
[e]din: drug ingredient.
[f]pcam: Perelman Center for Advanced Medicine.
[g]pah: Pennsylvania Hospital.
[h]hup: Hospital of the University of Pennsylvania.
[i]ravdin: Ravdin building.
[j]dpn: drug product name.
[k]meds: medications.

Figure 4. Distribution of medication intents among patient messages (left) and provider messages (right) in the medLDA model. medLDA: medication-specific latent Dirichlet allocation.



XSL•FO
RenderX

Table 5. Performance of majority class by topic classification.

Message type and medication intent category	Recall	Precision	F1 score
Patient			
Medication location	0.833	0.682	0.749
Medication question	__[a]	—	—
Medication request	0.843	0.685	0.756
Medication taking	0.481	0.745	0.585
Provider			
Medication change	—	—	—
Medication question	0.965	0.726	0.829
Medication question response	0.342	0.636	0.445
Medication refill question	—	—	—

[a]The class could not be predicted with this approach.

Discussion

Principal Findings

We developed and applied an unsupervised method, LDA, to facilitate the review and derivation of patient and provider intents from a large data set of messages produced using an asynchronous, bidirectional communication platform. We learned that LDA can be leveraged to identify subgroups of messages, but with some limitations.

First, we successfully applied a data-driven approach using single-word frequencies scaled by L2-penalized regression coefficients and detected distinct word usage by role (ie, patients and providers). Patients used temporal expressions (eg, “morning” and “evening”) to initiate requests, medical terms (eg, “rx” and “pulse”) to communicate medication and BP reporting, and confirmations (eg, “okay” and “thx”) to convey information understanding. Providers commonly used salutations (eg, “mrs” and “mr”) to initiate communication with patients, directive verbs (eg, “check record,” “sign,” “send,” “confirm,” “recheck,” and “look”) to instruct the patient, and positive and negative sentiments (eg, “nice,” “great,” and “worry”) to encourage patients to continue program engagement. We did not conduct a formal sentiment analysis.

These initial insights informed the decision to develop LDA models based on roles—patient and provider—to identify potentially distinct topics among messages. In the summary overview of the 50 LDA-derived topics, the results initially seemed promising. In LDA models for both patient and provider messages, the majority of messages occurred within a few prominent major topics, and the terms comprising each topic appeared sensible. For example, among the most common topics from provider-generated messages, we observed topics with terms indicative of BP checking and reporting (ie, topic 47: “bp” and “check”), BP reporting encouragement (ie, topic 12: “good,” “work,” “keep,” and “great”), and appointment scheduling (ie, topic 42: “call,” “appt,” and “come”). Among the most common topics from patient-generated messages, we observed topics with terms suggestive of confirmation and gratitude (ie, topic 1: “thank” and “nice”), medication adherence and BP reporting (ie, topic 17: “pulse,” “mg,” “tab,” and “dose”), and BP reporting (ie, topic 7: “bp,” “record,” and “check”). However, when we set out to validate the LDA-derived topics via manual annotation, we still observed significant heterogeneity of intents within LDA-derived topics. For example, in topic 17, the patient-derived topics composed of words like “mg,” “tab,” and “dose” contained diverse messages citing medication dosage, ranging from medication adherence reports to questions about medication, as well as nonmedication-related messages (eg, BP values). The topics were also not comprehensive, as several other topics contained medication-related messages. We hypothesized that the data set might be too heterogeneous for LDA; in particular, we were concerned that the messages were too complex (ie, messages often contained more than one intent) for the effective application of LDA.

Therefore, we focused our efforts on a curated subset of the data. We manually identified messages containing a single medication-related intent and annotated them according to four different medication-related intent categories, for both patient and provider messages (4 topics each). We aimed to determine how well the medLDA model could identify these medication intents as distinct topics. We observed that, generally, for each medLDA model–derived topic, there was a dominant intent, resulting in moderate precision for some classes (ie, patient: medication location, medication request, and medication taking; provider: medication question and medication question response). However, each topic still contained heterogeneity in intent, and the distribution of intents was skewed across topics. There are several potential explanations for these observations. One is that in our manually annotated reference standard, the distribution of messages with a single medication intent were largely skewed within both the patient-generated and provider-generated medication messages. Among the provider-generated medication intents, the medication question intent was predominant (~70%) among messages. Among patient-generated medication intents, the medication request or medication-taking intents were common, and they tended to co-occur among medication LDA topics. Studies of patient portal message classification also demonstrate somewhat skewed

XSL•FO
RenderX

distributions in message type [26,27]. These skewed distributions and shared common vocabulary terms may explain why the medLDA models were not able to perfectly discern each medication intent across topics. Another consideration is that in narrowing our data set to single-intent messages related to medication, there was a several factor–fold reduction in our data set. Thus, the resulting data set may have been too small for LDA to learn nuanced pattern differences.

Our findings suggest that LDA may still hold promise for automatically discerning novel topics within a large corpus of text data. However, there is likely a "just right" database for its application, one where the unit of analysis contains one or two primary intents, and different intents are well represented (ie, the database is very large). Currently, the use of LDA for analyzing text data may be limited to identifying broad differences in language use patterns, with the caveat that lexical similarity does not necessarily mean intent similarity. For example, De et al [28] applied LDA topic modeling to narrow down classes of patient messages—those relating to fatigue, prednisone, and patient visits—to identify commonly occurring themes within those message classes. Our work demonstrates that LDA is limited for providing further clinical insight. Further investigation of terms and semantic categories encoded by MedEx and Stanza provided some insights of shared and distinct concepts; however, more powerful language models might be necessary to discern intents with subtle semantic differences that are important for clinical contexts.

Limitations and Future Work

Our pilot study has several limitations. Notably, we conducted our analysis with a single data set generated from a particular patient-provider engagement program. However, we believe that unsupervised learning approaches can be beneficial for streamlining the mining of free-text data with customization to each individual program or application. As a result, this work is important because it highlights potential approaches for incorporating unstructured learning into this process. Customizations could be achieved with a larger annotated corpus and more powerful language models. Also, our unit of analysis was each message; we could have chosen to analyze other units (eg, sentences) that could improve LDA performance. However, this would not resemble real messaging practices, which often contain more than one sentence and intent. Another limitation is that our intent categories were clinically oriented, that is, they were based on clinically actionable intents. The model may have performed differently with a different reference standard framework (eg, negative, positive, or neutral sentiment). In the future, we will develop patient-provider language models, such as Bidirectional Encoder Representations from Transformers models, that might improve our ability to capture and leverage differences between message types to improve automatic intent classification.

Conclusions

We demonstrated how unsupervised learning can be applied to group text messages and identified medication-related messages within a bidirectional text messaging system for hypertension management. While LDA was useful in generating coarse categories, more detailed intent annotation is needed to develop reliable NLP-based intent classifiers that drive clinical actions and address subtopic heterogeneity.

Acknowledgments

This project was supported, in part, by a grant from the National Institute on Aging (grant P30-AG034546), which provided financial support for AD, TL, TD, and EA. NSL was funded by the Department of Veterans Affairs through the National Clinician Scholars Program. KC reported receiving grant support from Independence Blue Cross, Inc, for this work. DM received funding from the National Institutes of Health for this work (grant UL1-TR001878).

Data Availability

The term list and associated probabilities within each LDA topic and medLDA topic can be found at our GitHub site [29].

Authors' Contributions

AD, NSL, TL, KC, EA, and DM conceptualized the study design. TL and NSL provided and helped interpret data. DM, NSL, and TL coded the text data, and AD analyzed the data. AD and DM created the data visualization. DM and AD wrote the first draft of the manuscript. NSL and AD contributed equally as first coauthors. KC and DM jointly supervised the work as co-senior authors. All authors revised the manuscript and approved the final version.

Conflicts of Interest

KC reported receiving grant support from the National Institutes of Health (grants K08-AG065444 and P50-CA-244690), the Patient-Centered Outcomes Research Institute, the RAND Corporation, and Roundtrip, Inc; personal fees from the Villanova School of Business; board membership for Primary Care Progress, Inc; and consultancy fees from Verily, Inc, that are outside of the submitted work. DM reported receiving grant support from the National Institutes of Health (grants P30-AR069589 and P50-MH127511) and the University of Pittsburgh/Pittsburgh Health Alliance, all outside of this work.

Multimedia Appendix 1

XSL•FO
RenderX

Supplementary materials.
[DOCX File , 211 KB - jmir_v24i6e36151_app1.docx]

References

1. Kallmerten PS, Chia LR, Jakub K, Turk MT. Patient portal use by adults with heart failure: An integrative review. Comput Inform Nurs 2021 Apr 08;39(8):418-431. [doi: 10.1097/CIN.0000000000000733] [Medline: 33782319]
2. Sun R, Korytkowski MT, Sereika SM, Saul MI, Li D, Burke LE. Patient portal use in diabetes management: Literature review. JMIR Diabetes 2018 Nov 06;3(4):e11199 [FREE Full text] [doi: 10.2196/11199] [Medline: 30401665]
3. Han H, Gleason KT, Sun C, Miller HN, Kang SJ, Chow S, et al. Using patient portals to improve patient outcomes: Systematic review. JMIR Hum Factors 2019 Dec 19;6(4):e15038 [FREE Full text] [doi: 10.2196/15038] [Medline: 31855187]
4. 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]
5. Masterman M, Cronin RM, Davis SE, Shenson JA, Jackson GP. Adoption of secure messaging in a patient portal across pediatric specialties. AMIA Annu Symp Proc 2016;2016:1930-1939. [Medline: 28269952]
6. Majeed-Ariss R, Baildam E, Campbell M, Chieng A, Fallon D, Hall A, et al. Apps and adolescents: A systematic review of adolescents' use of mobile phone and tablet apps that support personal management of their chronic or long-term physical conditions. J Med Internet Res 2015 Dec 23;17(12):e287 [FREE Full text] [doi: 10.2196/jmir.5043] [Medline: 26701961]
7. Campos CL, Jones D, Snively BM, Rocco M, Pedley C, Atwater S, et al. Text messaging and home blood pressure monitoring for patients with uncontrolled hypertension: Proposal for a feasibility pilot randomized controlled trial. JMIR Res Protoc 2021 May 14;10(5):e18984 [FREE Full text] [doi: 10.2196/18984] [Medline: 33988513]
8. Ye J. The impact of electronic health record-integrated patient-generated health data on clinician burnout. J Am Med Inform Assoc 2021 Apr 23;28(5):1051-1056 [FREE Full text] [doi: 10.1093/jamia/ocab017] [Medline: 33822095]
9. Thomas Craig KJ, Willis VC, Gruen D, Rhee K, Jackson GP. The burden of the digital environment: A systematic review on organization-directed workplace interventions to mitigate physician burnout. J Am Med Inform Assoc 2021 Apr 23;28(5):985-997 [FREE Full text] [doi: 10.1093/jamia/ocaa301] [Medline: 33463680]
10. Shimada SL, Petrakis BA, Rothendler JA, Zirkle M, Zhao S, Feng H, et al. An analysis of patient-provider secure messaging at two Veterans Health Administration medical centers: Message content and resolution through secure messaging. J Am Med Inform Assoc 2017 Sep 01;24(5):942-949 [FREE Full text] [doi: 10.1093/jamia/ocx021] [Medline: 28371896]
11. Robinson JR, Valentine A, Carney C, Fabbri D, Jackson GP. Complexity of medical decision-making in care provided by surgeons through patient portals. J Surg Res 2017 Jun 15;214:93-101 [FREE Full text] [doi: 10.1016/j.jss.2017.02.077] [Medline: 28624066]
12. Sulieman L, Robinson JR, Jackson GP. Automating the classification of complexity of medical decision-making in patient-provider messaging in a patient portal. J Surg Res 2020 Nov;255:224-232 [FREE Full text] [doi: 10.1016/j.jss.2020.05.039] [Medline: 32570124]
13. Stenner SP, Johnson KB, Denny JC. PASTE: Patient-centered SMS text tagging in a medication management system. J Am Med Inform Assoc 2012;19(3):368-374 [FREE Full text] [doi: 10.1136/amiajnl-2011-000484] [Medline: 21984605]
14. Liu Z, Peach RL, Lawrance EL, Noble A, Ungless MA, Barahona M. Listening to mental health crisis needs at scale: Using natural language processing to understand and evaluate a mental health crisis text messaging service. Front Digit Health 2021;3:779091 [FREE Full text] [doi: 10.3389/fdgth.2021.779091] [Medline: 34939068]
15. Chen J, Lalor J, Liu W, Druhl E, Granillo E, Vimalananda VG, et al. Detecting hypoglycemia incidents reported in patients' secure messages: Using cost-sensitive learning and oversampling to reduce data imbalance. J Med Internet Res 2019 Mar 11;21(3):e11990 [FREE Full text] [doi: 10.2196/11990] [Medline: 30855231]
16. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res 2003;3:993-1022 [FREE Full text]
17. Sulieman L, Yin Z, Malin BA. Why patient portal messages indicate risk of readmission for patients with ischemic heart disease. In: Proceedings of the AMIA Annual Symposium. 2019 Presented at: The AMIA Annual Symposium; November 16-20, 2019; Washington, DC p. 828-837 URL: http://europepmc.org/abstract/MED/32308879
18. Noteboom C, Al-Ramahi M. What are the gaps in mobile patient portal? Mining users feedback using topic modeling. In: Proceedings of the 51st Hawaii International Conference on System Sciences. 2018 Presented at: The 51st Hawaii International Conference on System Sciences; January 3-6, 2018; Waikoloa Village, Hawaii p. 564-573 URL: https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/a1092b54-69de-433e-89f4-504654658b92/content [doi: 10.24251/hicss.2018.072]
19. Gupta D, Saul M, Gilbertson J. Evaluation of a deidentification (De-Id) software engine to share pathology reports and clinical documents for research. Am J Clin Pathol 2004 Feb;121(2):176-186. [doi: 10.1309/E6K3-3GBP-E5C2-7FYU] [Medline: 14983930]
20. scattertext. spaCy. URL: https://spacy.io/universe/project/scattertext [accessed 2022-02-16]
21. Wexler A, Davoudi A, Weissenbacher D, Choi R, O'Connor K, Cummings H, et al. Pregnancy and health in the age of the internet: A content analysis of online "birth club" forums. PLoS One 2020;15(4):e0230947 [FREE Full text] [doi: 10.1371/journal.pone.0230947] [Medline: 32287266]

XSL•FO
RenderX

22. Davoudi A, Lee NS, Chivers C, Delaney T, Asch EL, Reitz C, et al. Patient interaction phenotypes with an automated remote hypertension monitoring program and their association with blood pressure control: Observational study. J Med Internet Res 2020 Dec 03;22(12):e22493 [FREE Full text] [doi: 10.2196/22493] [Medline: 33270032]
23. Hripcsak G, Rothschild AS. Agreement, the f-measure, and reliability in information retrieval. J Am Med Inform Assoc 2005;12(3):296-298 [FREE Full text] [doi: 10.1197/jamia.M1733] [Medline: 15684123]
24. Zhang Y, Zhang Y, Qi P, Manning CD, Langlotz CP. Biomedical and clinical English model packages for the Stanza Python NLP library. J Am Med Inform Assoc 2021 Aug 13;28(9):1892-1899 [FREE Full text] [doi: 10.1093/jamia/ocab090] [Medline: 34157094]
25. Xu H, Stenner SP, Doan S, Johnson KB, Waitman LR, Denny JC. MedEx: A medication information extraction system for clinical narratives. J Am Med Inform Assoc 2010;17(1):19-24 [FREE Full text] [doi: 10.1197/jamia.M3378] [Medline: 20064797]
26. Haun JN, Lind JD, Shimada SL, Martin TL, Gosline RM, Antinori N, et al. Evaluating user experiences of the secure messaging tool on the Veterans Affairs' patient portal system. J Med Internet Res 2014 Mar 06;16(3):e75 [FREE Full text] [doi: 10.2196/jmir.2976] [Medline: 24610454]
27. Boffa M, Weathers A, Ouyang B. Analysis of patient portal message content in an academic multi-specialty neurology practice. Neurology 2015 Apr;84(14 Supplement):S11.005.
28. De A, Huang M, Feng T, Yue X, Yao L. Analyzing patient secure messages using a fast health care interoperability resources (FHIR)-based data model: Development and topic modeling study. J Med Internet Res 2021 Jul 30;23(7):e26770 [FREE Full text] [doi: 10.2196/26770] [Medline: 34328444]
29. LDA_textbot_analysis. GitHub. 2022. URL: https://github.com/semantica-NLP/LDA_textbot_analysis [accessed 2022-06-06]

Abbreviations

amlopidin: amlodipine
appt, appoint: appointment
BP: blood pressure
daili: daily
DBN: drug brand name
DDF: drug dose form
DIN: drug ingredient
DPN: drug product name
eHTN: Employee Hypertension Management Program
FREQ: frequency
gm: good morning
hr: hour
hup: Hospital of the University of Pennsylvania
HypoDetect: Hypoglycemia Detector
i2b2: Informatics for Integrating Biology and the Bedside
LDA: latent Dirichlet allocation
medLDA: medication-specific latent Dirichlet allocation
ML: machine learning
NER: named entity recognition
NLP: natural language processing
PASTE: Patient-Centered Automated SMS Tagging Engine
rx: prescription
schedul: schedule
see follow: see at follow-up
tab: tablet
thx: thanks

XSL•FO
RenderX

Edited by A Mavragani; submitted 04.01.22; peer-reviewed by R Benson, G Liu, F He, SS Amritphale, L Mbuagbaw; comments to author 01.02.22; revised version received 01.04.22; accepted 17.05.22; published 29.06.22.

Please cite as:
Davoudi A, Lee NS, Luong T, Delaney T, Asch E, Chaiyachati K, Mowery D
Identifying Medication-Related Intents From a Bidirectional Text Messaging Platform for Hypertension Management Using an Unsupervised Learning Approach: Retrospective Observational Pilot Study
J Med Internet Res 2022;24(6):e36151
URL: https://www.jmir.org/2022/6/e36151
doi:10.2196/36151
PMID:35767327

©Anahita Davoudi, Natalie S Lee, ThaiBinh Luong, Timothy Delaney, Elizabeth Asch, Krisda Chaiyachati, Danielle Mowery. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 29.06.2022. 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

Re-engineering a Clinical Trial Management System Using Blockchain Technology: System Design, Development, and Case Studies

Yan Zhuang[1,2], PhD; Luxia Zhang[1,2], MPH, MD; Xiyuan Gao[3], BSc; Zon-Yin Shae[4], PhD; Jeffrey J P Tsai[4], PhD; Pengfei Li[2], MSc; Chi-Ren Shyu[5], PhD

[1]National Institute of Health Data Science, Peking University, Beijing, China

[2]Advanced Institute of Information Technology, Peking University, Hangzhou, China

[3]Department of Statistics, University of Missouri, Columbia, MO, United States

[4]Department of Computer Science and Information Engineering, Asia University, Taichung, Taiwan

[5]Institute for Data Science and Informatics, University of Missouri, Columbia, MO, United States

Corresponding Author:
Chi-Ren Shyu, PhD
Institute for Data Science and Informatics
University of Missouri
22G Heinkel Building
University of Missouri
Columbia, MO, 65211-2060
United States
Phone: 1 573 882 3884
Email: shyuc@missouri.edu

Abstract

Background: A clinical trial management system (CTMS) is a suite of specialized productivity tools that manage clinical trial processes from study planning to closeout. Using CTMSs has shown remarkable benefits in delivering efficient, auditable, and visualizable clinical trials. However, the current CTMS market is fragmented, and most CTMSs fail to meet expectations because of their inability to support key functions, such as inconsistencies in data captured across multiple sites. Blockchain technology, an emerging distributed ledger technology, is considered to potentially provide a holistic solution to current CTMS challenges by using its unique features, such as transparency, traceability, immutability, and security.

Objective: This study aimed to re-engineer the traditional CTMS by leveraging the unique properties of blockchain technology to create a secure, auditable, efficient, and generalizable CTMS.

Methods: A comprehensive, blockchain-based CTMS that spans all stages of clinical trials, including a sharable trial master file system; a fast recruitment and simplified enrollment system; a timely, secure, and consistent electronic data capture system; a reproducible data analytics system; and an efficient, traceable payment and reimbursement system, was designed and implemented using the Quorum blockchain. Compared with traditional blockchain technologies, such as Ethereum, Quorum blockchain offers higher transaction throughput and lowers transaction latency. Case studies on each application of the CTMS were conducted to assess the feasibility, scalability, stability, and efficiency of the proposed blockchain-based CTMS.

Results: A total of 21.6 million electronic data capture transactions were generated and successfully processed through blockchain, with an average of 335.4 transactions per second. Of the 6000 patients, 1145 were matched in 1.39 seconds using 10 recruitment criteria with an automated matching mechanism implemented by the smart contract. Key features, such as immutability, traceability, and stability, were also tested and empirically proven through case studies.

Conclusions: This study proposed a comprehensive blockchain-based CTMS that covers all stages of the clinical trial process. Compared with our previous research, the proposed system showed an overall better performance. Our system design, implementation, and case studies demonstrated the potential of blockchain technology as a potential solution to CTMS challenges and its ability to perform more health care tasks.

(J Med Internet Res 2022;24(6):e36774) doi:10.2196/36774

XSL•FO
RenderX

KEYWORDS

blockchain; clinical trials; clinical trial management system; electronic data capture; smart contract

Introduction

Clinical trials are considered to be the cornerstone of the development of new drugs or treatments because they have investigated the safety and efficacy of new therapeutics using a standard protocol [1]. As conducting clinical trials involves complex processes, good management is critical for success [2]. The clinical trial management system (CTMS) is a set of software tools used for managing clinical trial processes including but not limited to protocol development, site selections, patient recruitment, study conduct, data collection, data analysis, and study closeout. With the increasing adoption of the CTMS, many substantial benefits such as accessing up-to-date information, improving data quality, and boosting overall study efficiency have simplified the traditional labor-intensive management process [3-5]. A complete CTMS design must be secure, cost-efficient, compliant with regulations, traceable, and auditable to manage the process in each phase of the study [5-9]. However, the current CTMS market is fragmented and lacks thorough designs with all the required features and management tools [2,7]. According to the 2019 Unified Clinical Operations Survey provided by Veeva (a global life science service), nearly all respondents (99%) had issues with their current CTMS, and 90% of the respondents reported a significant deficiency in at least 1 CTMS application [10,11]. Emerging technologies, such as blockchain, are believed to potentially re-engineer CTMSs and provide a comprehensive solution [12].

Blockchain is an open-source distributed ledger technology that has been proven in the areas of security, stability, and robustness in real-world applications, including cryptocurrencies [13-15]. A blockchain consists of continuously generated blocks containing validated transactions, time stamps, and block IDs used for chaining to the previous block. It is considered to be a revolutionary technology, as it has unique features such as immutability to ensure data consistency; a peer-to-peer system with public auditability (all blockchain transactions can be audited by any user at any time) to provide regulatory compliance; anonymity (all users are represented by a unique hash string) to protect patient privacy [16]; and a smart contract, which is a self-executing programmable computer protocol that can be designed for different applications. These features are a perfect fit for health care applications [17-19]. However, most blockchain designs used for health care applications remain in the conceptual stage, and there are several technical challenges such as scalability constraints [20-24]. Quorum blockchain, a private blockchain developed by JP Morgan that requires participating users to gain permissions from the blockchain initiator before joining, has enhanced security, scalability, and efficiency based on the original blockchain [25,26]. The performance of the Quorum blockchain in areas such as transaction throughput and transaction latency has been evaluated as extraordinarily improved (compared with the original blockchain) using the Raft consensus mechanism for the validation process without compromising its unique properties [25].

We have implemented a blockchain platform that provides unique software designs for key components of CTMSs to achieve better management and monitoring of clinical trials with the following applications: (1) an auditable, sharable, and transparent electronic trial master file (eTMF); (2) a fast patient recruitment model with an automated matching mechanism through the smart contract and a simplified enrollment using a digital signature validated by the blockchain; (3) a timely electronic data capture (EDC) system that ensures data consistency, traceability, and security through blockchain's properties; (4) a reproducible data analytics module that keeps records of data and code use; and (5) a secure, auditable, and efficient payment and reimbursement model. We conducted case studies for each application to empirically prove its feasibility and test its scalability, stability, and efficiency.

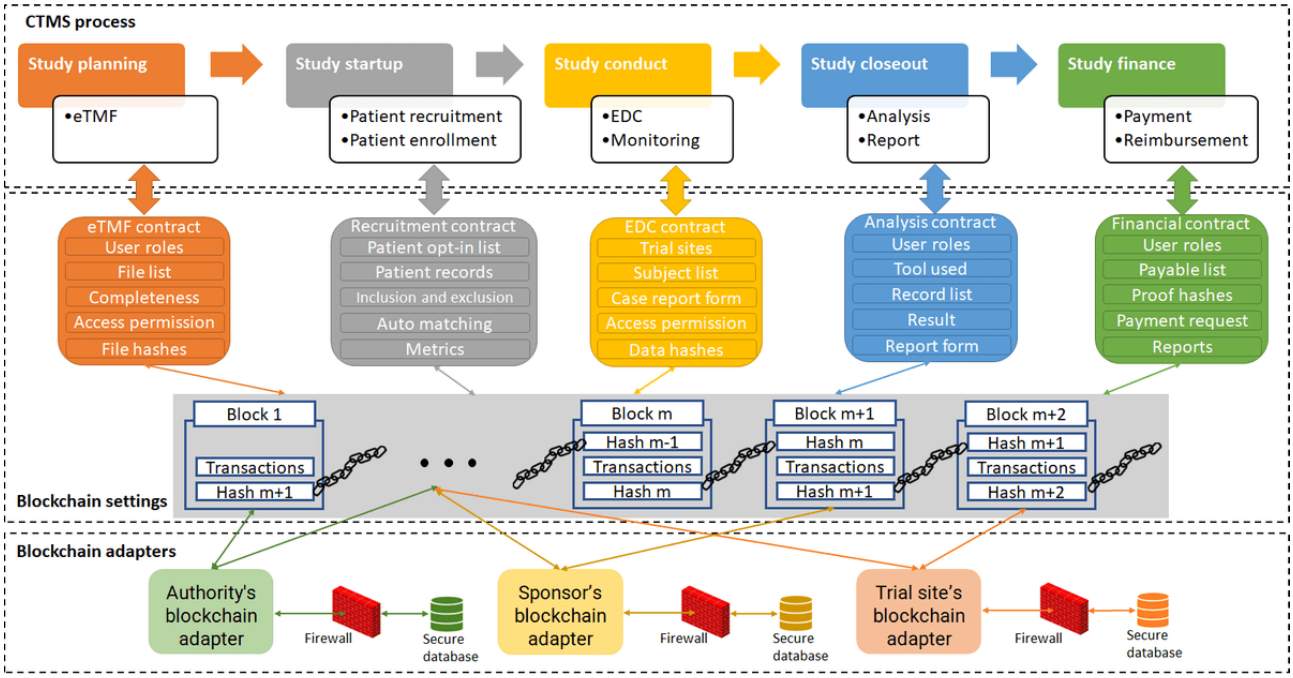
Methods

Overview

Figure 1 depicts the overall architecture and main smart contract designs that span five different stages of the clinical trial process: (1) study planning targets on the eTMF for clinical trial protocol development; (2) following the protocol's establishment, study start-up focuses on recruiting participants for clinical trials; (3) while the clinical trial is in progress, study conduct develops EDC for data collection and monitoring the safety and efficacy of the treatment; (4) during the closing phase, study closeout collaborates with statistical tools to provide a reproducible analytics report; and (5) study finance adopts the blockchain's nature of cryptocurrency for payment and reimbursement.

XSL•FO
RenderX

Figure 1. The overall architecture of 5 different clinical trial processes. Different applications are implemented by smart contracts defined through blockchain initiation. Participating sites require blockchain adapters to interact with the blockchain system and the secure database protected by local health information technology regulations. CTMS: clinical trial management system; EDC: electronic data capture; eTMF: electronic trial master file.



This architecture is generalizable to all different clinical trials; therefore, the participating site can use the same CTMS to manage simultaneous clinical trials by switching trial IDs obtained by the sponsors, whereas the registration on the blockchain-based CTMS remains constant. It is noteworthy to mention that CTMSs may require additional functions such as protocol development, which are not included in our system design, as the current procedures for protocol development are sophisticated enough [27] with no need to adopt a new approach such as blockchain to reinstate the existing process, although most present tools can be integrated with our proposed blockchain-based CTMS without extensive arrangement.

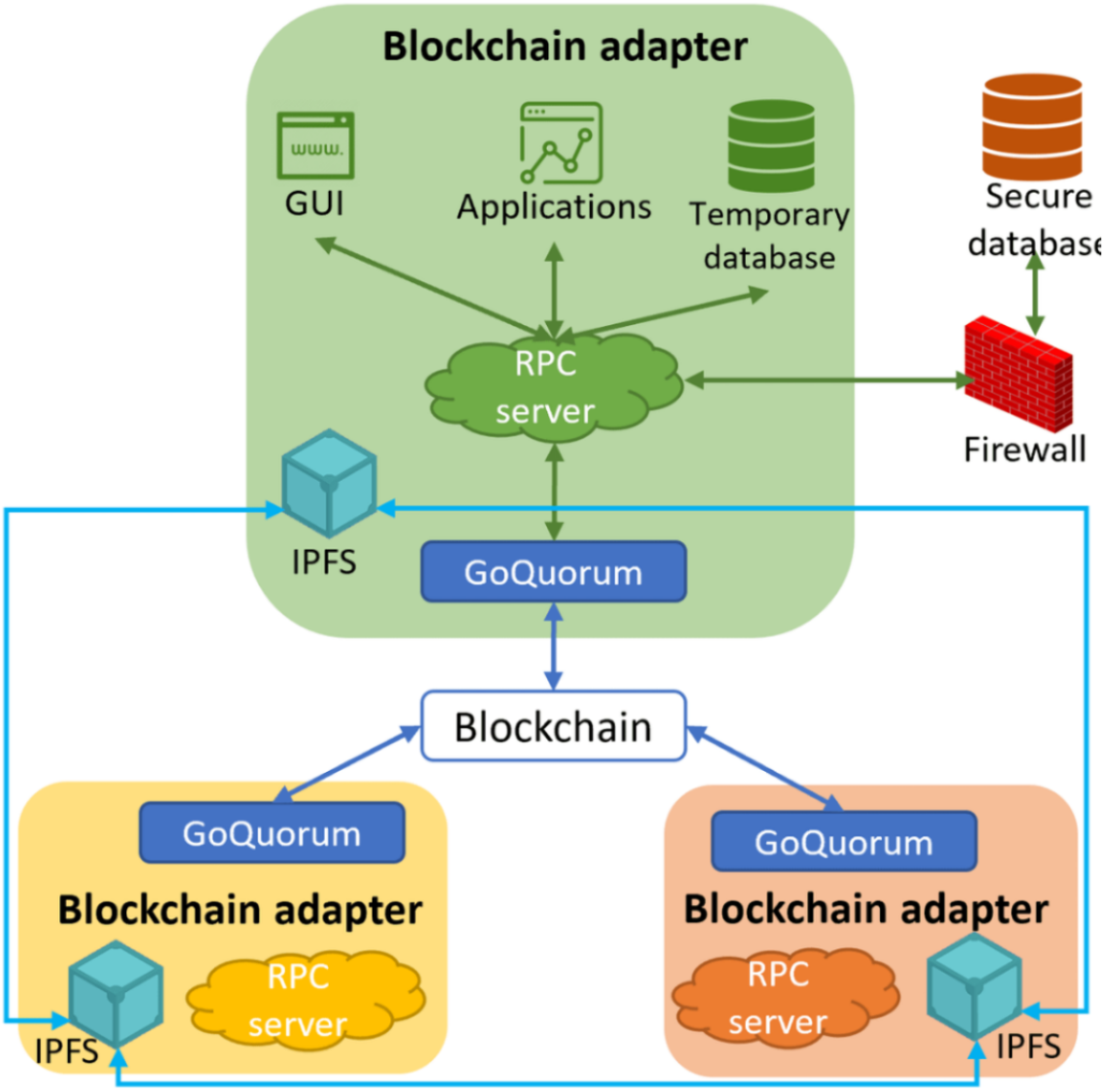
The remainder of this paper is organized as follows: (1) the environment setup specifies the details of the hardware and software required to construct the system, notably the blockchain adapters shown in Figure 1; (2) the following sections describe the blockchain settings as smart contracts for each stage of the CTMS process, as shown in Figure 1; and (3) as shown in the *Results* section, we conducted case studies on study planning, study start-up, and study conduct to test the blockchain features and overall performance, such as scalability.

Environment Setup

In this study, we used a laptop equipped with 16 GB of RAM, an i5 processor, and a 1 TB hard drive to represent the authority's node and 5 Intel NUC machines, each equipped with 16 GB of RAM, an Intel i3 processor, and a 1.5 TB hard drive to represent the clinical trial sites' and sponsors' nodes. These machines were set up at 2 different locations under different networks. Owing to the regulatory compliance required by each participating site, we converted each blockchain node into a blockchain adapter that abides by local health information technology regulations [28]. As shown in Figure 2, each blockchain adapter installed the Ubuntu operating system, which in turn runs GoQuorum, an Ethereum-based Quorum blockchain client. Once the authority node started the client, the Quorum blockchain with the Raft consensus mechanism was built automatically. Then, the blockchain adapter will be added to the blockchain by the authority and will be able to communicate with other blockchain adapters as well as the local secured database protected by health IT when the participating site obtains permission to join the system. Tools can be installed on the blockchain adapter and integrated with the blockchain through a remote procedure call server. For example, a team of professionals such as medical experts, statisticians, clinical research coordinators, and medical writers can use the blockchain adapters for protocol development. Existing tools can still be used as anticipated. The sole exception (limited to development scenarios) was the ability to store a log file in the blockchain after each use. In all other aspects, users can take advantage of blockchain's unique features such as immutability to ensure file consistency and traceability to acknowledge the users who edited the file, as well as decentralization to improve the efficiency of working distributively without changing the existing legacy process. Each adapter has an interplanetary file system (IPFS) installed, which is an innovative peer-to-peer distributed file system. Each file stored in the IPFS is assigned a unique cryptographic hash for indexing and ensuring consistency. Compared with other distributed file systems, the IPFS has shown great improvement in efficiency, scalability, and stability [29]. However, the design concept of the IPFS lacks the capability of access control and file use tracking [30]. However, this makes it a perfect match for blockchain. The IPFS can be used for data storage, whereas blockchain serves as a content management system.

XSL•FO
RenderX

Figure 2. Blockchain adapter design and connections. All adapters have the same setup with an RPC server connecting local applications and databases, an IPFS that connects to other IPFS on each adapter, and a GoQuorum application programming interface that connects to the blockchain. GUI: graphical user interface; IPFS: interplanetary file system; RPC: remote procedure call.



A unique public-private key pair will be generated for each user such as participants, investigator, sponsor, and others, after the user registers a blockchain account through a site's blockchain adapter. Patients and potential participants must register on-site so that the administrators from the trial sites can prove their identities and map their local patient ID to the blockchain account with their consent. A hash value of the public key, also known as the blockchain account address, will be used to represent the user's identity. A private key will be used as a digital signature. All transactions must be signed by the sender's private key before they can be recorded in the blockchain. Each group, such as the financial management team, has an umbrella account in addition to separate individual user accounts, each of which maps to the umbrella account for each member so that the entire group can share permission when authentication to the group is made. Potential participants must visit trial sites to opt in to the system and generate their blockchain account so that the trial site can verify their identities. Instead of memorizing the key pair, a username and password or biometric authentication mechanism can be used on a graphical user interface (GUI) for users to log into the blockchain system.

To build the blockchain-based CTMS, we made the following assumptions: (1) each participating site, including the sponsor, trial sites, site institutional review boards, and the Food and Drug Administration, is required to provide at least 1 blockchain node, which can be any electronic device that can install the Quorum blockchain; (2) the authority (eg, Food and Drug Administration) has initiated the blockchain system so that all that the participating site requires is to obtain permission from the authority before joining the system by proving their identity; and (3) each participating site has an administrator to operate the system.

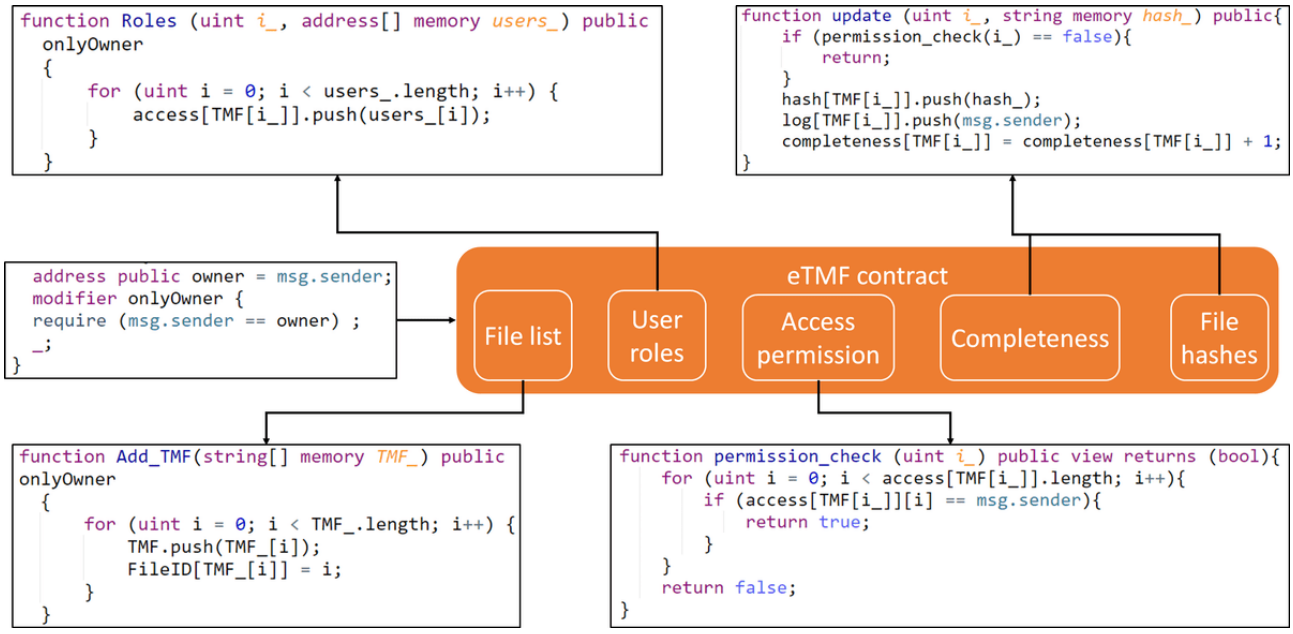
Study Planning

With the increasing adoption of electronic documents for clinical trials, planning, sharing, and managing documents have become increasingly critical and intricate [31]. The eTMF is a form of content management system used to manage and collaborate in

XSL•FO
RenderX

a timely fashion on essential clinical documents throughout the life cycle of clinical trials. However, several persistent challenges exist in most eTMF designs, such as the inability to audit unlocatable files; inaccurate metrics for timeliness, quality, or completeness; inconsistency caused by loss or alteration of the information; and collaboration issues caused by different trial master file (TMF) standards. Our eTMF design contains a smart contract used to control file access, validate file consistency, and manage collaboration in TMF development and the IPFS network used for file storage and file indexing. Figure 3 shows a portion of the source code of the smart contract for each function.

Figure 3. A portion of the source code of the electronic trial master file (eTMF) contract design. These codes show the main logic of each function. All smart contract functions are predefined, and users can use graphical user interfaces to call the functions.



The TMF document list and other expected artifacts list must be identified in the eTMF smart contract at the beginning of the study planning phase. Sponsors must assign files to team members so that they can work jointly by adding their blockchain accounts to the smart contract associated with the file ID from each TMF. All TMFs are encrypted using OpenSSL and a randomly generated key pair before being stored in the IPFS [32]. All users can download the file from the IPFS using the file hash, but only users who have permission from the sponsor can retrieve the decrypt key from the smart contract to decrypt the file. When a team member works on a certain file, the blockchain adapter from the member's site automatically sends a flag to the smart contract to block other team members from working on the same file. When the team member finishes editing the file, the blockchain adapter will encrypt the new version of the file with a random new pair of keys, upload the encrypted file to the IPFS, obtain a new hash value from the IPFS, and send the decrypt key, hash value, and negative flag to the blockchain to update the file registration information. The completeness metric (the percentage of expected artifacts that are completed) will be updated automatically.

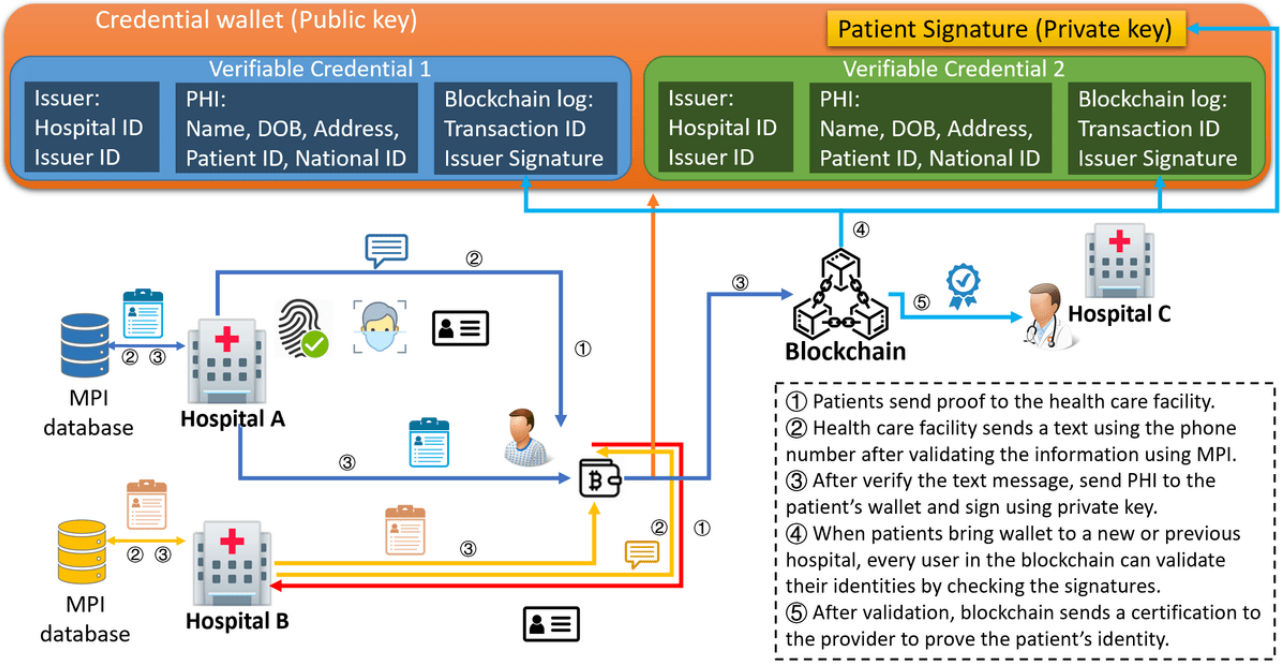
Using blockchain technology for eTMF can provide the following unique features: (1) consistency—each version of a file will have a hash value stored in the blockchain, and any changes to the file will result in a mismatch of its new hash with the original hash; (2) traceability and auditability—each team member must work on the file sequentially so that any changes can be traceable to the editing user through blockchain transaction history [33] (users can audit who has changed the file by checking the log files in the blockchain, but only the sponsors, or the authority, know the real identity of the user); (3) efficiency—using IPFS as file storage is efficient compared with other file transferring processes because team members can collaboratively work on the same file; and (4) security—with blockchain's security setting, all transactions are considered secure so that only the recipients can receive the correct decrypt key for the file.

Study Start-up

After the study team has selected trial sites and defined target enrollment metrics, clinical trials must meet the recruitment goal. Patient recruitment has been recognized as a key to success. However, 86% of clinical trials fail to meet their recruitment goals on time. We refined our earlier work, which was a blockchain-based recruitment model using a smart contract for automated matching [13] for use under the CTMS study start-up scheme, as shown in Figure 4. We have developed a patient credential wallet for patients to store their key pairs as well as the credentials issued by the health care facilities on their local device. Verifiable credentials contain issuers' information such as hospital ID and issuer ID, patients' protected health information, and the ID of the blockchain transaction that was made and signed by the issuer when the patient presents their blockchain account on-site. In this configuration, each health care facility has the protected health information mapped to the patients' public key, but the patients' private keys are maintained locally.

XSL•FO
RenderX

Figure 4. Trial sites must register participants and input primary medical history to the smart contract. The smart contract will automatically send notifications to the matched patients, asking for authentication through their mobile device using their fingerprint. DOB: date of birth; MPI: master patient index; PHI: protected health information.



Users who want to participate in clinical trials must follow the same procedure outlined for patients and participants. They also need to provide permission for the use of their electronic health records (EHRs) for future matching purposes. The hospital administrator must input the basic user information into the recruitment contract, including demographic information and primary diagnoses from past visits. As soon as the sponsor inputs the recruitment inclusion and exclusion criteria into the smart contract, the smart contract can automatically screen potential participants by matching the basic information. After the initial screening is accomplished, hospitals can perform precise matching by checking the full EHRs of matched users. When a user is fully matched, the sponsor will send a transaction to the user to ask for enrollment. Future on-site visits are still required, but the enrollment process can be operated by sending out the consent form and asking the user to sign it using their private key [34], which will send a confirmation transaction to the smart contract. The smart contract also contains personalized metrics such as time consumption, cost, and retention, used to evaluate the performance of the team in the recruitment process and the timeliness of decisions to increase productivity.

The features of blockchain technology are a great fit for the needs of recruitment and enrollment for the following reasons: (1) transparency can improve the awareness of clinical trials for patients, (2) auditability ensures the legitimacy of clinical trials, (3) anonymity protects patient privacy, (4) asymmetric encryption eases the process for patient enrollment, and (5) the automated matching mechanism operating via a smart contract can significantly reduce the time required for recruitment.

Study Conduct

Data collection is one of the most important processes for the evaluation and monitoring of aspects of the experimental condition (eg, drug effect) as clinical trials are conducted. Compared with the traditional paper-based case report form (CRF), which serves the sole purpose of recording information, EDC systems are used to collect data electronically, reduce data errors, improve the efficiency of the collation process, and enable faster data access. However, there are several challenges faced by both the paper form and the EDC system, such as security concerns, data inconsistency, and untimely (slow) data input. All clinical trials were monitored, which was a process of data and safety monitoring. The Data and Safety Monitoring Board comprises a group of professionals from different fields such as biostatistics, medicine, and ethics, who monitor patient safety and treatment efficacy. The legacy data monitoring method is source data verification (SDV), which is resource intensive and accounts for up to 30% of the total clinical trial budget. We designed an EDC contract to effectively collect data, reduce the need for SDV, and monitor patient safety persistently.

After participants submit their consent to the blockchain during the recruitment phase, the system administrator from each trial site must register them in the participant list in the EDC contract to map their blockchain account to the trial ID and their local patient IDs. Figure 5A shows a customized CRF converted through a smart contract shown in Figure 5B. Data fields and types, such as selection and input, can be defined in the smart contract and retrieved by blockchain adapters for conversion into a GUI-based CRF. After each participant's site visit, the investigator needs to input the records into the electronic CRF (eCRF). The records will then be automatically encrypted, hashed, and stored in the IPFS using the blockchain adapter of the site. Figure 6 shows the encryption, storing, and retrieving process after the data are input through the GUI. The smart contract will validate whether the trial site has permission to store the participant's data, after which the visit ID and decrypt key will be sent through Quorum blockchain's private

XSL•FO
RenderX

transaction. This ensures that the data contained in the private transaction are encrypted, and only the recipient can decrypt using their private key, or the information can be made available to the sponsor by the site's administrator. The sponsor's blockchain adapter will automatically retrieve the decrypt key and hash from the blockchain, decrypt the records, and hash the records to compare with the hash stored in the blockchain. Mismatching hashes will create an alert to the trial site and for the sponsor to initiate further investigation. This can eliminate data inconsistency caused by falsification. However, most EDCs require manual input, and human data errors can also cause a data inconsistency issue. We have implemented a data extraction application on each blockchain adapter to automatically extract the CRF-required data from the visit records before storing it in the secured EHR database to reduce the risk of human errors. However, most CRFs have partial data fields that are trial oriented and are not included in the EHR, meaning that manual input is still needed. Although blockchain's immutability features were intended to be designed as unchangeable for all records, some modifications may still occur owing to unintentional human error. However, the updated (erroneous) records cannot replace the previous input and will contain a pointer to the former hash of the data record for future validation. In this blockchain-based CTMS system, safety monitoring relies on the investigators to report through the EDC so that the safety monitoring team can evaluate only true issues of data and safety.

Figure 5. (A) The graphical user interface for principal investigators containing a sample electronic case report form (eCRF) coded through the smart contract and a sample timeline for the participant. (B) The smart contract is used for defining data fields and types of the eCRF. Blockchain adapters will retrieve the information from the smart contract and generate the eCRF.

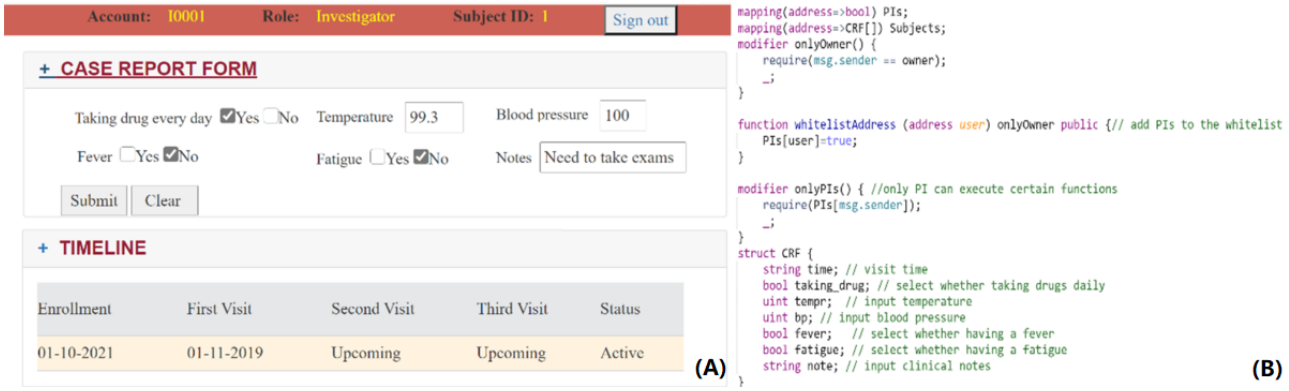
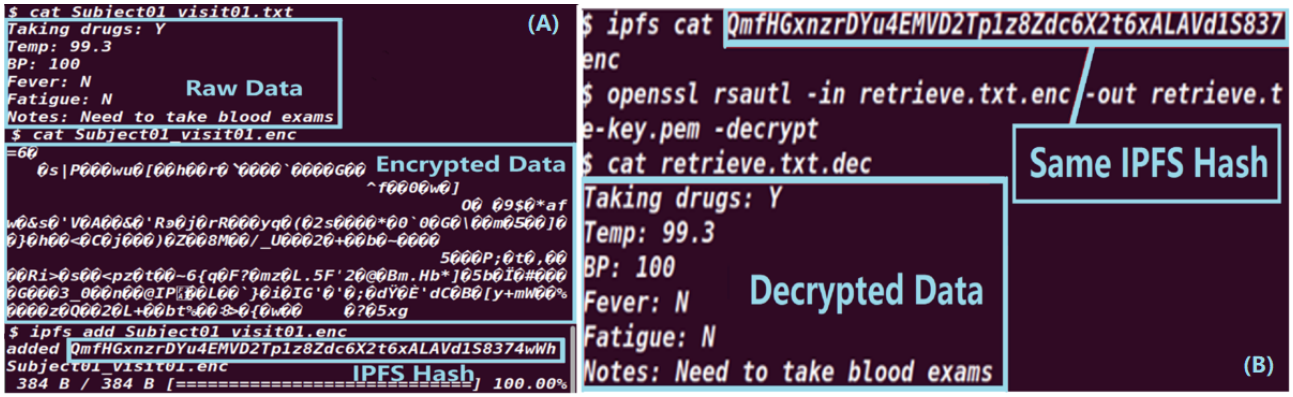


Figure 6. (A) The investigator's blockchain adapter retrieves the data through the graphical user interface, encrypts the data using the investigator's public key, and stores the encrypted data into the interplanetary file system (IPFS). (B) The sponsor's blockchain adapter retrieves the encrypted data through the IPFS and decrypts the data using the private key.



In this module, using blockchain and an IPFS for EDC has the following benefits: (1) immutability ensures data consistency from the data input through data analysis to reduce the need for SDV; (2) traceability improves the auditability as to who, when, and how the records were changed; (3) the efficiency of the IPFS permits fast data retrieval; and (4) the security property of blockchain protects patient privacy and data security. With the addition of the automated extraction mechanism added to the blockchain adapters, the efficiency and accuracy of the data collection process are significantly enhanced.

Study Closeout

When the last participant completes their site visit, the clinical trial will enter the closeout phase. There will be a closeout checklist that can be collaboratively completed by sponsors and the team using the eTMF. The clinical trial database can be locked to prevent future changes after validation of the final data. Statistical analysis must be conducted to evaluate the outcomes of clinical trials. In the blockchain-based CTMS system, we created several R scripts for several statistical models in each blockchain adapter and added the names of the available statistical methods in the smart contract. The statisticians can use the existing script or use their preferred statistical tool to analyze the final data, after which they can generate the final

XSL•FO
RenderX

statistical report. The source code must be encrypted and stored in the IPFS for validation purposes. The team members or authority can request the decrypt key from the sponsor and reproduce the results using the source code and clinical trial data.

Barriers to analyzing clinical trials are mainly those of selective reporting [35], incomplete reporting data [36], and a lack of appropriate statistical methods [37]. Blockchain provides solutions to the challenges in this stage through its immutability and auditability features, which help to ensure the completeness of reports. The analyzed data and applied methods will store a log file in the blockchain so that the study group and the authority can reproduce the results at any time to validate the completeness and appropriateness of audits of the analyzing methods.

Study Finance

Numerous components can add to the cost of a clinical trial, such as regulatory services, start-up, and medical writing, of which all can create challenges in financial management. In this module, we use payment and reimbursement to the trial sites and patients [38] as an example of the potential use of blockchain technology as a financial management tool. The validation of the payment or reimbursement requests as to when and how the recipient is paid is a time-consuming process, making on-time payments challenging [39]. In this module, we designed a smart contract and a collaborative validation network in the blockchain-based CTMS.

Before a clinical trial begins, the study team should define a list of payable entities (as well as payable items) and input this list into the smart contract. This can standardize the payable items and reduce the risk of hidden fees. Each trial site may have different rates for the same payable item. The rates must also be defined through a smart contract accessible only by the sponsor and trial site. Compensation for the patient is normally based on the time required for the participant to take part in the study. After each visit, the trial site must send a request transaction containing the time spent and the payable items to the blockchain, store the encrypted proof in the IPFS, and send the decrypt key and hash to the sponsor. The clinical trial financial management team can validate the proof and send the payment requests to the sponsor. A transaction that contains a payment receipt will be sent from the sponsor to the trial site and marks the status of the request as paid in the trial site's GUI. The payment to or reimbursement of the trial site has a similar process, as trial sites send request transactions that contain payable items to the sponsor and wait for the approval. However, payable items may not cover all requested payments. Trial sites need to follow the same request process with *additional items* in the payable items. Sponsors can collaborate to validate the proof and price the additional items to make payments.

Using blockchain technology for financial management has the following benefits: (1) a customizable charging standard for different trial sites as long as the sponsor agrees (all payable items and rates are preferred to be defined in the smart contract for an expedited validation process); (2) the traceability feature ensures that all requests and payments are traceable by the requester and the recipient (all the proof needs to be stored in the IPFS); (3) the immutability feature ensures that the request, payment, and proof of payment are not modifiable after the payment is made; and (4) the security property of blockchain protects user privacy.

Results

Overview

We implemented the blockchain-based CTMS and installed it on 6 blockchain nodes representing 1 authority, 2 sponsors, and 3 trial sites. Each blockchain node has been converted into a blockchain adapter. We generated 2 clinical trials with 1000 participants at each trial site for each study. We conducted 3 case studies to simulate the processes described in the study planning, start-up, and conduct sections to evaluate the feasibility and performance of the system. These studies were also conducted to assess the key components of the processes discussed in the *Study Closeout* and the *Study Finance* sections, such as ensuring the consistency of data recorded from statistical tools and reimbursement or payment forms.

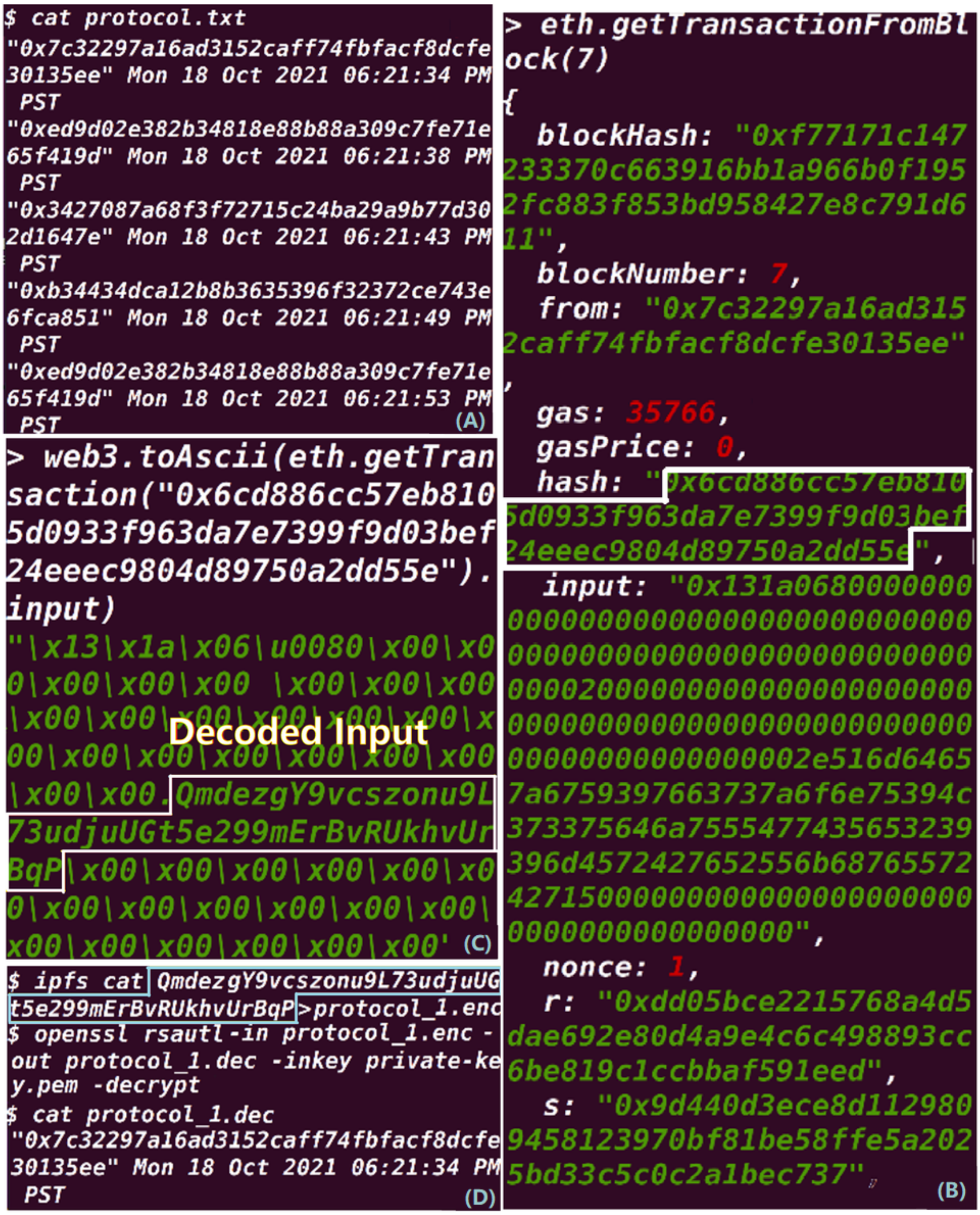
Study Planning

During this stage, the key benefit of the blockchain system is to record all changes in the essential files and ensure file consistency. The case study simulates the TMF collaboration process, as all experts are working on the same file named *protocol.txt* and sharing an *umbrella ID* for encryption purposes. The goal of this case study is to test the capability of (1) handling file conflicts while collaborating on master files, (2) ensuring traceability and auditability of file changes using blockchain's properties, and (3) storing files into and retrieving files from the IPFS efficiently. We created 2 accounts for each node representing the 2 experts from each participating site to simulate the TMF collaboration process. The script is designed as follows: (1) an expert retrieves the file hash from the smart contract and the file from the IPFS using the hash, (2) the expert writes their blockchain ID (for tracing validation purposes) to the file and keeps the file open for 10 seconds, (3) encrypts the file and stores the new hash to the smart contract, and (4) repeats this script 20 times. The script is deployed on each blockchain adapter, and all scripts run simultaneously. If the file is being opened, the file will not be retrievable. In this case, the script will keep running until successfully executed.

After 12 minutes and 38 seconds, the scripts were successfully executed, and all 60 records were moved into the final protocol file as shown in Figure 7A. Records can be traced from the blockchain by tracing transactions. For example, the first record is recorded in the transaction inside block 7 as shown in Figure 7B, where we can extract the transaction ID from the block and check the details using the ID, as shown in Figure 7C. Then, the encrypted file can be retrieved from the IPFS using the hash stored in the transaction and decrypted using the decrypt key under the umbrella ID as shown in Figure 7D.

XSL•FO
RenderX

Figure 7. (A) Part of the final master file with all experts' blockchain IDs and input times. (B) Trace the input record from the blockchain by checking the block number and transaction ID. (C) Decode the details of the transaction input. (D) Retrieve and decrypt the file using the hash stored in the blockchain transaction.



Study Start-up

In addition to blockchain features such as transparency, auditability, and anonymity, the key contribution of this module is to provide an automated matching mechanism that can filter out the potential participants that matched the recruitment criteria. We have repeated the case study from our previous recruitment work [13] using the Quorum-based approach to assess the accuracy and performance and comparing it with our previous Ethereum-based approach. The case study simulates the participant matching and recruitment process. The goal of this case study is to test the accuracy and efficiency of an automated matching engine and anonymity during the recruitment process.

We selected 6000 patient records of breast cancer from the Surveillance, Epidemiology, and End Results database and evenly distributed them into 3 clinical sites. Common inclusion and exclusion criteria from 10 recruiting clinical trials for breast cancer were selected to simulate the recruitment process. We created a script to populate the smart contract with criteria and patient records, such as demographic information and primary diagnosis. After calling the automated matching function, a total of 1145 of 6000 patients' blockchain accounts were matched

XSL•FO
RenderX

in 1.39 seconds, which was slightly better than the 2.13 seconds that resulted from our previous Ethereum approach.

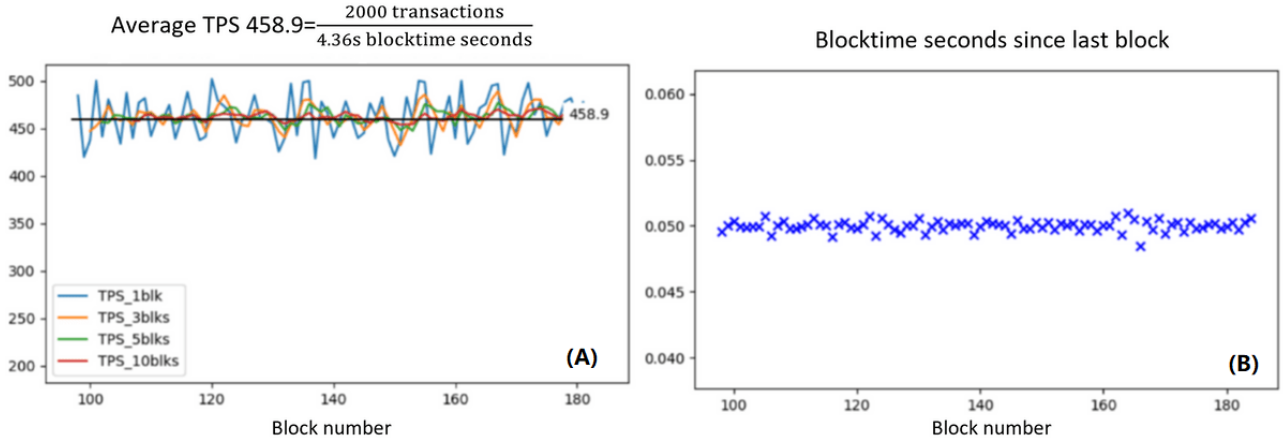
Study Conduct

The case study simulated the data collection process during a clinical trial using a sample eCRF designed through a smart contract. A script was created to mimic the data capture process: (1) the script randomly generated data for the data fields defined by the eCRF from the 3 trial site adapters; (2) the trial site adapters encrypt the data file using a random public key, store the encrypted data file in the IPFS and obtain the hash value, and send decrypt key and hash value to the sponsor through a private transaction; and (3) the sponsors' adapters retrieve data from the IPFS and decrypt the data files. The goal of this case study is to test the following aspects: (1) data consistency from the input to the retrieval, (2) the successful rate and accuracy of the transactions for data collection, and (3) the scalability and efficiency of the system.

We ran the script on each participant from each blockchain adapter every second for an hour. There were 1.2 million transactions written into the blockchain with an average latency of 1.73 seconds and 335.4 transactions per second (TPS), a key measurement of blockchain scalability. The remainder of the transactions were held in the buffer to sequentially push them into the blockchain. It took nearly 18 hours to send 21.6 million transactions generated by the script into the blockchain with a 100% success rate. All records have been precisely collected. Figure 8 shows the blockchain performance after submitting 2000 transactions simultaneously. The average TPS was 458.9 (SD 21.224), but it gradually decreased and stabilized during the simulation. The TPS was not associated with the block generation time from our simulation results.

As script 3 is purely off-chain, the stability is based on the performance of the IPFS and the specifications of the adapter's devices. We did not included script 3 in our stability test, as many researchers have proven the performance of the IPFS [40]. To test system robustness, we manually shut down the sender's blockchain adapter after the transaction and found that the recipient could still retrieve the data.

Figure 8. Scalability and stability test results of the first 2000 simultaneous transactions. (A) Transactions per second (TPS) values were calculated using every 1, 3, 5, and 10 blocks. (B) Time consumption of generating a new block.



Discussion

Principal Findings

In addition to the scalability test, we have evaluated various blockchain features that are critical at different stages of a clinical trial. Blockchain demonstrates auditability, transparency, and immutability in the *Study Planning* section. We manually submitted malicious transactions to tamper with the current eTMF using a random blockchain account outside the umbrella ID. These transactions were automatically filtered out by the blockchain with no responses. All transactions are publicly auditable, and the recorded data cannot be changed. From our simulation, blockchain plus IPFS is also efficient for file storage and retrieval. In the study start-up case study, we mainly tested the feasibility and efficiency of subject matching through smart contracts. The simulation results show that the smart contract can match potential participants accurately and efficiently without exposing the patients' identities. During study conduct case study, we evaluated data consistency and scalability efficiency and robustness of the blockchain. TPS is a key measurement of blockchain's scalability, and Quorum blockchain shows a better performance compared with Ethereum. All legitimate transactions have been successfully executed and recorded in the blockchain. The blockchain also shows robustness when a single node fails in our simulation.

Limitations

The main limitation of the proposed architecture is that health care facilities must cooperate to provide blockchain adapters to join the system. As blockchain adapters need to communicate with secure databases protected by local health care facilities' firewalls and store classified documents and patient records outside the firewall to the IPFS, health care facilities need to follow the local health information technology regulations to set up the blockchain adapters. Although there are no hardware requirements for blockchain adapters, the device specifications may affect their performance. From our simulation experience, too much transaction generation may take up memory and crush the blockchain node. From a previous study that evaluated the scalability of Quorum blockchain using powerful cloud service as 8 blockchain nodes, their testing result of 8 nodes with Raft consensus mechanism has a similar TPS with slightly lower latency which is 1.4 seconds compared with our 1.7 seconds

XSL•FO
RenderX

[41]. Both studies have empirically proved the Quorum blockchain as stable, robust, and scalable. Another limitation is that the designed eCRF is intended only for simulation purposes, and the data are randomly generated to test the scalability of the system. The real eCRF may have more complex designs, but because data are collected and transferred through the IPFS while blockchain only serves as a key distributor, access controller, and log auditor, there should not be significant changes in the sizes of blockchain transactions that cause concerns about the feasibility, scalability, and stability of the blockchain system.

Future Work

Our future work will continue to investigate the needs of the clinical trial process and add more comprehensive functions to the proposed blockchain-based CTMS architecture, such as adding machine learning tools to monitor patient conditions persistently and predict side effects and overall outcomes. The current safety monitoring process described in the *Study Conduct* section relies on the EDC process. However, the Data and Safety Monitoring Board convenes only when the clinical trial has been conducted for a while and the data have met a certain point. Adding artificial intelligence components to the Study Conduct module can achieve more efficient monitoring. We will also investigate more potential in CTMS design using blockchain technology, such as integrating secure multiparty computation with blockchain for computational applications such as subject matching and using the cryptocurrency concept to build a novel CTMS that will help ensure timely validation and payment.

Conclusions

In this study, we described a blockchain-based CTMS that covers 4 different stages of clinical trials. Through our simulation process, we empirically proved the feasibility of each application in the blockchain architecture. Compared with the scalability test on the Ethereum blockchain from our previous research, Quorum blockchain shows an overall better performance. The unique contribution of this work is the exploration of the benefits of blockchain technology in targeting the needs of CTMSs. This covers several essential functions (each of which is a part of the clinical trial process) using a distinctive blockchain adapter design to support an efficient, secure, traceable, transparent, and auditable management system. Our system design, implementation, and simulation results demonstrate the potential of blockchain to create a CTMS, and we suggest that this should serve as a notice for the health IT community to consider this emerging technology.

Acknowledgments

This research was supported in part by the University of Missouri Informatics and Data Science Research Initiatives (YZ and CRS); grant 72125009 from the National Natural Science Foundation of China (LZ); grants 2019BD017, 2020BD004, and 2020BD005 from the PKU-Baidu Fund (YZ, PL, and LZ); and grant MOST 110-2321-B-468-001 from the Ministry of Science and Technology in Taiwan (ZYS).

Conflicts of Interest

None declared.

References

1. Hills M, Armitage P. The two-period cross-over clinical trial. Br J Clin Pharmacol 1979 Jul 04;8(1):7-20 [FREE Full text] [doi: 10.1111/j.1365-2125.1979.tb05903.x] [Medline: 552299]
2. Saumell Y, Torres O, Batista M, Sánchez L. Validation of instruments for assessing drug safety management during the conduction of clinical trials. Int J Health Policy Manag 2018 Jul 01;7(7):623-629 [FREE Full text] [doi: 10.15171/ijhpm.2017.140] [Medline: 29996582]
3. Hufford M, Peterson D, Paty J, Shiffman S. System for clinical trial subject compliance. Google Patents. URL: https://patents.google.com/patent/US8433605B2/en [accessed 2022-06-06]
4. Bleicher P, Stamos N, Klofft J, Dale R. Clinical trial data management system and method. Google Patents. 2004. URL: https://patents.google.com/patent/EP1082693B1/en [accessed 2022-06-06]
5. Frigoletto FD, Lieberman E, Lang JM, Cohen A, Barss V, Ringer S, et al. A clinical trial of active management of labor. N Engl J Med 1995 Sep 21;333(12):745-750. [doi: 10.1056/NEJM199509213331201] [Medline: 7643880]
6. Park YR, Yoon YJ, Koo H, Yoo S, Choi CM, Beck SH, et al. Utilization of a clinical trial management system for the whole clinical trial process as an integrated database: system development. J Med Internet Res 2018 Apr 24;20(4):e103 [FREE Full text] [doi: 10.2196/jmir.9312] [Medline: 29691212]
7. Dagalur S. CTMS: what you should know. Applied Clinical Trials. 2016. URL: http://www.appliedclinicaltrialsonline.com/clinical-trial-management-systems-what-you-should-know [accessed 2022-06-06]
8. Bérard C, Cloutier L, Cassivi L, Systems D. Evaluating clinical trial management systems: a simulation approach. Industrial Manag Data Syst 2012;112(1):146-164. [doi: 10.1108/02635571211193680]
9. Benchoufi M, Porcher R, Ravaud P. Blockchain protocols in clinical trials: transparency and traceability of consent. F1000Res 2017 Jul 4;6:66. [doi: 10.12688/f1000research.10531.3]
10. Veeva 2019 Unified clinical operations survey report. Veeva. URL: https://www.veeva.com/wp-content/uploads/2019/06/Veeva_ClinicalOperationsSurvey_2019.pdf [accessed 2022-06-06]

XSL•FO
RenderX

11. Bose A, Das S. Trial analytics--a tool for clinical trial management. Acta Pol Pharm 2012;69(3):523-533 [FREE Full text] [Medline: 22594267]
12. Benchoufi M, Ravaud P. Blockchain technology for improving clinical research quality. Trials 2017 Jul 19;18(1):335 [FREE Full text] [doi: 10.1186/s13063-017-2035-z] [Medline: 28724395]
13. Zhuang Y, Sheets LR, Shae Z, Chen Y, Tsai JJ, Shyu C. Applying blockchain technology to enhance clinical trial recruitment. AMIA Annu Symp Proc 2019;2019:1276-1285 [FREE Full text] [Medline: 32308925]
14. Szabo N. Smart contracts: building blocks for digital markets. EXTROPY J Transhumanist Thought 1996;16 [FREE Full text]
15. Nakamoto S. Bitcoin: a peer-to-peer electronic cash system. SSRN J 2019 [FREE Full text] [doi: 10.2139/ssrn.3977007]
16. Khurshid A, Holan C, Cowley C, Alexander J, Harrell DT, Usman M, et al. Designing and testing a blockchain application for patient identity management in healthcare. JAMIA Open 2021 Jul;4(3):ooaa073 [FREE Full text] [doi: 10.1093/jamiaopen/ooaa073] [Medline: 34505001]
17. Zhuang Y, Sheets LR, Chen Y, Shae Z, Tsai JJ, Shyu C. A patient-centric health information exchange framework using blockchain technology. IEEE J Biomed Health Inform 2020 Aug;24(8):2169-2176. [doi: 10.1109/jbhi.2020.2993072]
18. Zhang P, Schmidt D, White J, Lenz G. Blockchain technology use cases in healthcare. In: Advances in Computer. Amsterdam, Netherlands: Elsevier Science; 2018.
19. Khurshid A. Applying blockchain technology to address the crisis of trust during the COVID-19 pandemic. JMIR Med Inform 2020 Oct 22;8(9):e20477 [FREE Full text] [doi: 10.2196/20477] [Medline: 32903197]
20. Kuo T, Zavaleta Rojas H, Ohno-Machado L. Comparison of blockchain platforms: a systematic review and healthcare examples. J Am Med Inform Assoc 2019 May 01;26(5):462-478 [FREE Full text] [doi: 10.1093/jamia/ocy185] [Medline: 30907419]
21. Zhuang Y, Chen Y, Shae Z, Shyu C. Generalizable layered blockchain architecture for health care applications: development, case studies, and evaluation. J Med Internet Res 2020 Jul 27;22(7):e19029 [FREE Full text] [doi: 10.2196/19029] [Medline: 32716300]
22. Kuo T, Kim H, Ohno-Machado L. Blockchain distributed ledger technologies for biomedical and health care applications. J Am Med Inform Assoc 2017 Nov 01;24(6):1211-1220 [FREE Full text] [doi: 10.1093/jamia/ocx068] [Medline: 29016974]
23. Cornelius CA, Qusay H. Blockchain in healthcare: opportunities, challenges, and possible solutions. Int J Healthcare Inf Syst Informatics 2020;15(3):82-97. [doi: 10.4018/IJHISI.2020070105]
24. Zhang P, Kuo T. The feasibility and significance of employing blockchain-based identity solutions in health care. In: Blockchain Technology and Innovations in Business Processes. Singapore: Springer; 2021.
25. Baliga A, Subhod I, Kamat P, Chatterjee S. Performance evaluation of the quorum blockchain platform. arXiv 2018 [FREE Full text] [doi: 10.48550/arXiv.1809.03421]
26. Cachin C, Vukoli M. Blockchain consensus protocols in the wild. In: Proceedings of the 31st International Symposium on Distributed Computing (DISC 2017). 2017 Presented at: 31st International Symposium on Distributed Computing (DISC 2017); Oct 16-20, 2017; Vienna, Austria.
27. Chan A, Tetzlaff J, Altman D, Laupacis A, Gøtzsche PC, Krleža-Jerić K, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. Ann Intern Med 2013 Feb 05;158(3):200-207 [FREE Full text] [doi: 10.7326/0003-4819-158-3-201302050-00583] [Medline: 23295957]
28. Azaria A, Ekblaw A, Vieira T, Lippman A. MedRec: using blockchain for medical data access and permission management. In: Proceedings of the 2016 2nd International Conference on Open and Big Data (OBD). 2016 Presented at: 2016 2nd International Conference on Open and Big Data (OBD); Aug 22-24, 2016; Vienna, Austria. [doi: 10.1109/OBD.2016.11]
29. Chen Y, Li H, Li K, Zhang J. An improved P2P file system scheme based on IPFS and Blockchain. In: Proceedings of the 2017 IEEE International Conference on Big Data (Big Data). 2017 Presented at: 2017 IEEE International Conference on Big Data (Big Data); Dec 11-14, 2017; Boston, MA, USA. [doi: 10.1109/bigdata.2017.8258226]
30. Nyaletey E, Parizi R, Zhang Q, Choo K. BlockIPFS - Blockchain-enabled interplanetary file system for forensic and trusted data traceability. In: Proceedings of the 2019 IEEE International Conference on Blockchain (Blockchain). 2019 Presented at: 2019 IEEE International Conference on Blockchain (Blockchain); Jul 14-17, 2019; Atlanta, GA, USA. [doi: 10.1109/blockchain.2019.00012]
31. Mitchel J, Hays J. System and method for electronic document management, organization, collaboration, and submission in clinical trials. Google Patents. URL: https://patents.google.com/patent/US20100077218 [accessed 2022-06-06]
32. Viega J, Messier M, Chandra P. Network Security with OpenSSL Cryptography for Secure Communications. Sebastopol, California, United States: O'Reilly Media; 2002.
33. Benchoufi M, Altman D, Ravaud P. From clinical trials to highly trustable clinical trials: blockchain in clinical trials, a game changer for improving transparency? Front Blockchain 2019 Dec 10;2. [doi: 10.3389/fbloc.2019.00023]
34. Zhuang Y, Sheets L, Gao X, Shen Y, Shae Z, Tsai JJ, et al. Development of a blockchain framework for virtual clinical trials. AMIA Annu Symp Proc 2020;2020:1412-1420 [FREE Full text] [Medline: 33936517]
35. Mbuagbaw L, Thabane L, Ongolo-Zogo P, Lang T. The challenges and opportunities of conducting a clinical trial in a low resource setting: the case of the Cameroon mobile phone SMS (CAMPS) trial, an investigator initiated trial. Trials 2011 Jun 09;12:145 [FREE Full text] [doi: 10.1186/1745-6215-12-145] [Medline: 21658262]

XSL•FO
RenderX

36. Koenig F, Slattery J, Groves T, Lang T, Benjamini Y, Day S, et al. Sharing clinical trial data on patient level: opportunities and challenges. Biom J 2015 Jan 18;57(1):8-26 [FREE Full text] [doi: 10.1002/bimj.201300283] [Medline: 24942505]
37. Howard KI, Krause MS, Vessey JT. Analysis of clinical trial data: the problem of outcome overlap. Psychother Theor Rese Pract Train 1994;31(2):302-307. [doi: 10.1037/h0090213]
38. Breitkopf CR, Loza M, Vincent K, Moench T, Stanberry LR, Rosenthal SL. Perceptions of reimbursement for clinical trial participation. J Empir Res Hum Res Ethics 2011 Sep;6(3):31-38 [FREE Full text] [doi: 10.1525/jer.2011.6.3.31] [Medline: 21931235]
39. Grady C. Payment of clinical research subjects. J Clin Investigation 2005 Jul 01;115(7):1681-1687. [doi: 10.1172/jci25694]
40. Benet J. Ipfs-content addressed, versioned, p2p file system. arXiv 2014. [doi: https://doi.org/10.48550/arXiv.1407.3561]
41. Mazzoni M, Corradi A, Di Nicola V. Performance evaluation of permissioned blockchains for financial applications: the ConsenSys Quorum case study. Blockchain Res Application 2022 Mar;3(1):100026 [FREE Full text] [doi: 10.1016/j.bcra.2021.100026]

Abbreviations

CRF: case report form
CTMS: clinical trial management system
eCRF: electronic case report form
EDC: electronic data capture
EHR: electronic health record
eTMF: electronic trial master file
GUI: graphical user interface
IPFS: interplanetary file system
SDV: source data verification
TMF: trial master file
TPS: transactions per second

Edited by T Leung; submitted 30.01.22; peer-reviewed by TT Kuo, R Rastmanesh, P Asprion; comments to author 14.04.22; revised version received 07.05.22; accepted 20.05.22; published 27.06.22.

Please cite as:
Zhuang Y, Zhang L, Gao X, Shae ZY, Tsai JJP, Li P, Shyu CR
Re-engineering a Clinical Trial Management System Using Blockchain Technology: System Design, Development, and Case Studies
J Med Internet Res 2022;24(6):e36774
URL: https://www.jmir.org/2022/6/e36774
doi:10.2196/36774
PMID:35759315

©Yan Zhuang, Luxia Zhang, Xiyuan Gao, Zon-Yin Shae, Jeffrey J P Tsai, Pengfei Li, Chi-Ren Shyu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.06.2022. 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

Assessing the Clinical Robustness of Digital Health Startups: Cross-sectional Observational Analysis

Sean Day[1], BS, MPhil; Veeraj Shah[1,2], BS; Sari Kaganoff[1], MBA; Shannon Powelson[3], BS; Simon C Mathews[4], MD

[1]Rock Health Inc, San Francisco, CA, United States

[2]Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom

[3]Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, United States

[4]Division of Gastroenterology & Hepatology, Johns Hopkins University School of Medicine, Baltimore, MD, United States

Corresponding Author:
Simon C Mathews, MD
Division of Gastroenterology & Hepatology
Johns Hopkins University School of Medicine
1800 Orleans St
Baltimore, MD, 21287
United States
Phone: 1 410 955 5000
Email: smathe14@jhmi.edu

Abstract

Background: The digital health sector has experienced rapid growth over the past decade. However, health care technology stakeholders lack a comprehensive understanding of clinical robustness and claims across the industry.

Objective: This analysis aimed to examine the clinical robustness and public claims made by digital health companies.

Methods: A cross-sectional observational analysis was conducted using company data from the Rock Health Digital Health Venture Funding Database, the US Food and Drug Administration, and the US National Library of Medicine. Companies were included if they sell products targeting the prevention, diagnosis, or treatment phases of the care continuum. Clinical robustness was defined using regulatory filings and clinical trials completed by each company. Public claims data included clinical, economic, and engagement claims regarding product outcomes made by each company on its website.

Results: A total of 224 digital health companies with an average age of 7.7 years were included in our cohort. Average clinical robustness was 2.5 (1.8 clinical trials and 0.8 regulatory filings) with a median score of 1. Ninety-eight (44%) companies had a clinical robustness score of 0, while 45 (20%) companies had a clinical robustness score of 5 or more. The average number of public claims was 1.3 (0.5 clinical, 0.4 economic, and 0.4 engagement); the median number of claims was 1. No correlation was observed between clinical robustness and number of clinical claims (r^2=0.02), clinical robustness and total funding (r^2=0.08), or clinical robustness and company age (r^2=0.18).

Conclusions: Many digital health companies have a low level of clinical robustness and do not make many claims as measured by regulatory filings, clinical trials, and public data shared online. Companies and customers may benefit from investing in greater clinical validation efforts.

(J Med Internet Res 2022;24(6):e37677) doi:10.2196/37677

KEYWORDS

digital health; health tech; software as a medical device (SaaMD); real-world evidence; venture capital

Introduction

The digital health sector has grown rapidly over the past decade [1,2]. There are more than 1900 digital health startup companies in the United States that have raised more than US $2 million in venture funding, which in total have raised more than US $77 billion in venture capital funding since 2011 [3].

Although growth is apparent, the ability to measure impact is not. Several studies have highlighted the need for greater clinical validation [4,5] and found that many solutions were not supported by robust clinical evidence [6,7] and demonstrated mixed results on cost savings and cost-effectiveness [8,9]. In addition, there is evidence that some claims made by digital health companies have been misleading [10-12], with a few

XSL•FO
RenderX

highly publicized cases resulting in legal action by the Federal Trade Commission and state attorneys general [13-15]. Most studies focusing on clinical impact are narrowly defined to specific clinical therapeutic areas (eg, diabetes, cardiac arrhythmia), making it difficult to extrapolate findings to the broader field of digital health [16-22]. Additional limitations include a small sample of companies and the use of publications as a proxy for clinical impact. Other studies in digital health have examined larger clinical trends, such as growth in clinical trials, but these often lack data on clinical and customer focus [23,24].

As a result, the literature is often too narrow or too broad to enable an understanding of clinical impact. In addition, no studies have examined both clinical rigor and public claims made by companies. To address these limitations, we sought to comprehensively examine the topic of clinical robustness in digital health companies by using a more comprehensive definition of clinical rigor and examining companies' public claims across the most in-depth database of US-based digital health companies. These findings provide additional context for all stakeholders in health technology that rely on a more accurate characterization of digital health solutions.

Methods

Population

Companies were identified using the Digital Health Venture Funding Database maintained by Rock Health Inc, a digital health venture fund and advisory firm, which has been used in prior studies [25-27]. The database includes all digital health companies with headquarters in the United States that have raised at least one venture funding round of US $2 million or more since 2011. Our analysis included companies that sell products targeting the prevention, diagnosis, or treatment phases of the care continuum, which raised at least one round of funding between 2011 and 2020. Digital health companies are defined as those that build and sell digital technologies in health care [28,29].

Company Variables

Total venture funding, clinical area(s) of focus, care continuum phase(s), and customer data were collected for each company. Clinical areas represented 1 of 20 specific clinical domains (eg, cardiovascular, nephrology). Care continuum phase was defined according to the following: prevention, diagnosis, or treatment. Customer type referred to the category of buyer for a company's products, such as payer (ie, health insurance companies), biopharma (ie, pharmaceutical or biopharmaceutical companies), and medical devices (ie, companies that manufacture medical devices). Companies can be categorized into multiple categories (ie, companies may address multiple phases of the care continuum or multiple clinical areas). Company variables were gathered from the Digital Health Venture Funding Database, which Rock Health maintains using publicly available information such as company websites, press releases, and US Securities and Exchange Commission filings. Data for companies were collected through August 3, 2021.

Claims Variables

The number and type of claims, defined as unique quantitative statements about product outcomes, made on a company's website were collected in the following categories: clinical, economic, and engagement. The definitions of these claim types are detailed in Table 1. Data were obtained by reviewing all pages of a company's website, excluding links to external pages such as press releases, between May 3, 2021, and August 3, 2021.

Table 1. Types of claims made by digital health companies.

Claim type	Definition	Claim subtype
Engagement	A quantitative statement on how engaged users are with the technology or that it provides a better patient experience	• Number of active users/user retention rate • Measure of user engagement per unit time (ie, monthly, annual)
Economic	A quantitative statement about a product's impact on health-related expenses or revenue for the buyer or end user of the product	• Money saved per stakeholder, including return on investment (either to patient, payer, or provider) or as compared to competition/existing standard of care • New revenue generation for stakeholders • Decrease in health care services utilization
Clinical	A quantitative statement about a product's impact on patient health or well-being	• Diagnostic efficacy • General clinical improvement/reduction in symptoms or condition • Change in objective clinical metric (including validated patient-reported metrics) • Disease cure (reversal or permanent cure of a disease) • Prevention (prevents progression or occurrence of a specific disease) • Improvement in quality of life • Improvement in medication adherence

Clinical Robustness Variables

We collected regulatory data from the Food and Drug Administration (FDA), including the number of 510(k), De Novo, and premarket approval filings (where the company was listed as the "Requester" on fda.gov). In addition, we collected the number and type of clinical trials by searching ClinicalTrials.gov where the company was listed as "Sponsor / Collaborator." Data on both FDA filings and clinical trials were collected between July 1, 2021, and September 2, 2021,

XSL•FO
RenderX

through a combination of web scraping and manual searching. Data collection on claims was completed by at least 3 authors, with blinded cross-review to ensure consistency in data collection.

A "clinical robustness" score was calculated for each company, defined as the sum of the number of regulatory filings and clinical trials. Each regulatory filing and clinical trial was weighted equally in the calculation.

Data and Statistical Analysis

All data were stored in Microsoft Excel (Microsoft Corp), where all descriptive and statistical analyses including coefficient of correlation between variables were calculated.

Results

Population Characteristics

There were 224 companies in the cohort, with an average age of 7.7 years. Collectively, these companies have raised a total of US $8.2 billion in venture capital funding since 2011. The companies spanned 3 phases of the care continuum, with 25 offering solutions for prevention, 106 offering solutions for diagnosis, and 110 companies offering solutions for treatment (Multimedia Appendix 1). Some companies offered products across multiple phases of the care continuum and were therefore counted in multiple categories. Average company funding was similar across phases of care—prevention companies raised US $35.3 million; diagnosis companies raised US $37.8 million; and treatment companies raised US $37.9 million.

Clinical Robustness

The average clinical robustness score for all companies was 2.5 (clinical trials: 1.8; regulatory filings: 0.8). The median clinical robustness score was 1, with 98 companies (44%) having a score of 0 and 34 companies (15%) having a score of 1. Diagnosis companies had the highest average clinical robustness scores (2.8), followed by treatment companies (2.2), and then prevention companies (1.9).

The average clinical robustness score of companies that sold to employers was 3.1, compared to 2.0, 2.2, and 2.7 for those companies that sold to payers, consumers, and providers, respectively (Multimedia Appendix 2). Fifteen of the 18 clinical areas had a higher average number of clinical trials than regulatory filings. The distribution of companies by clinical robustness score can be found in Table 2.

Table 2. Distribution of companies by clinical robustness score.

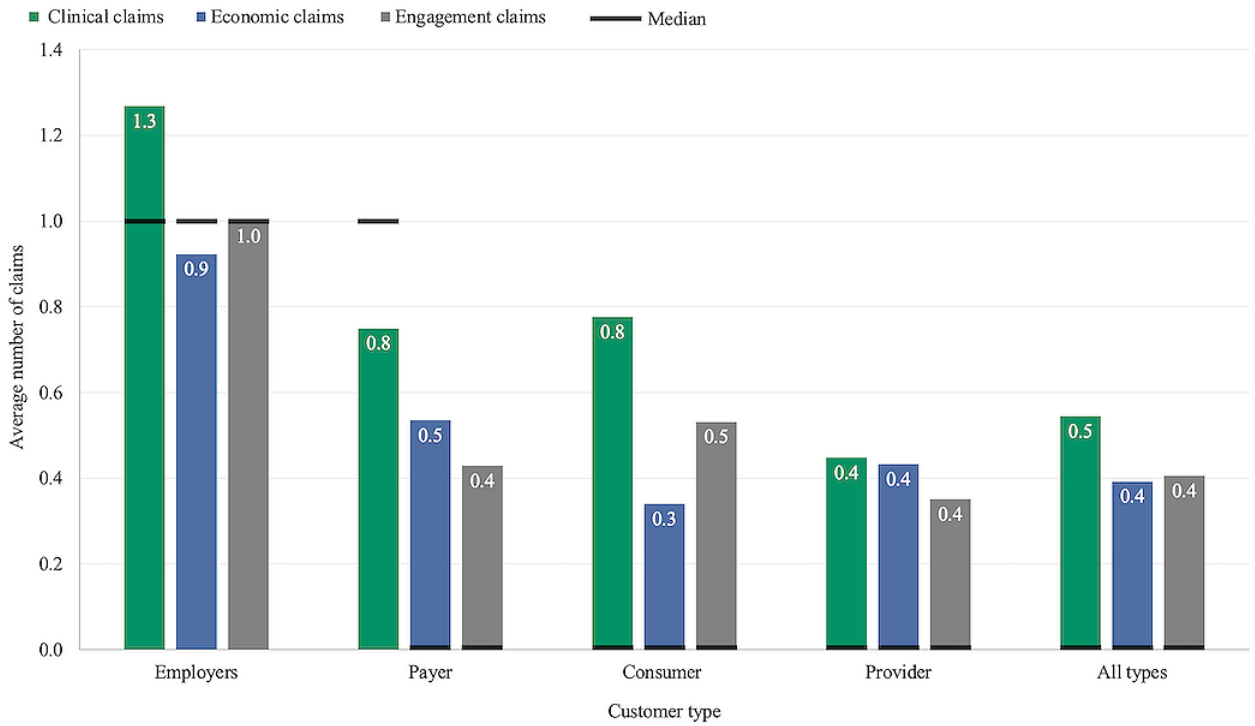
Clinical robustness score	Companies, n (%)
0	98 (44)
1	34 (15)
2	24 (11)
3	15 (7)
4	8 (4)
5	13 (6)
6	7 (3)
7	4 (2)
8	7 (3)
9	1 (0)
≥10	13 (6)

Claims

The average number of total claims for all companies was 1.3 (0.5 clinical, 0.4 economic, and 0.4 engagement) with the median number of total claims equal to 1 (0 clinical, 0 economic, 0 engagement) (Multimedia Appendix 3). The median number of claims of any type was zero for companies that sold to consumers and providers. The median number of economic and engagement claims for companies that sold to payers was also zero. Companies that sold to employers made more clinical, economic, and engagement claims than companies that sold to all other customer types (Figure 1). Many companies (43%) made zero claims.

XSL•FO
RenderX

Figure 1. Average number of claims made by start-ups broken out by customer type.



Clinical Robustness and Claims

There was no correlation between clinical robustness and number of clinical claims (r^2=0.02), clinical robustness and total funding (r^2=0.08), or clinical robustness and company age (r^2=0.18). In addition, there was no strong correlation between clinical areas that had higher clinical robustness scores and clinical areas where companies have received higher average funding (r^2=0.07).

Discussion

Principal Findings

Our findings indicate that many venture-backed startups in digital health have limited clinical robustness as measured by regulatory filings and clinical trials. There was, however, a sizable minority (20%) that had a score of 5 or more, suggesting a small population of rigorously tested solutions (Table 2). Although this subpopulation may portend progress, the lack of meaningful clinical validation for nearly half of digital health companies (44% had a clinical robustness score of 0) highlighted a major gap in health care technology today. The lack of overall correlation between a company's total venture funding and its clinical robustness score similarly highlighted a significant asymmetry in how companies are potentially valued in today's marketplace (ie, no correlation between clinical impact and funding) (Figure 2). However, it is possible that funding amounts reflect future anticipated value rather than current value [30].

Although average clinical robustness was quite low across the population, there was significant variation across clinical areas (eg, high in cardiovascular and nephrology and low in oncology and primary care) (Figure 3). This may reflect the varying levels of technological maturity of digital health solutions across clinical disciplines. Prior literature points to significant differences in technological maturity between well-funded clinical areas such as diabetes [27] and less well-funded areas such as reproductive and maternal health [31].

The average and median number of all claims was also low. These findings indicated that digital health companies largely did not share outcomes publicly. This may have reflected a desire to keep this data private, but more likely represented a lack of meaningful analyses of any impact (clinical, economic, or engagement) since this data could be used as a competitive differentiator if shared publicly. Separately, we identified 32 companies that had one or more clinical claims and a clinical robustness score of 0. These findings may suggest a disconnect between marketing and evidence. Additional future research could examine the links between public claims and clinical trials or regulatory filings (ie, are individual claims directly supported by clinical trials or regulatory filings?).

XSL•FO
RenderX

Figure 2. Clinical robustness, total claims, and total company funding across the digital health landscape.

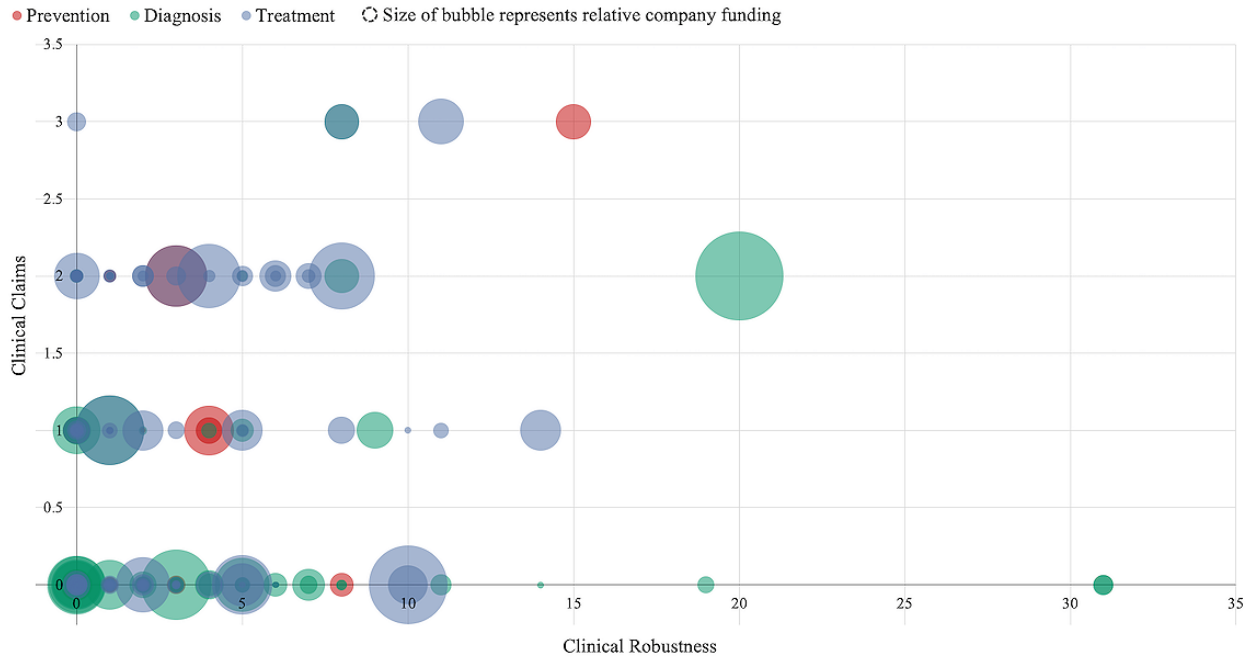
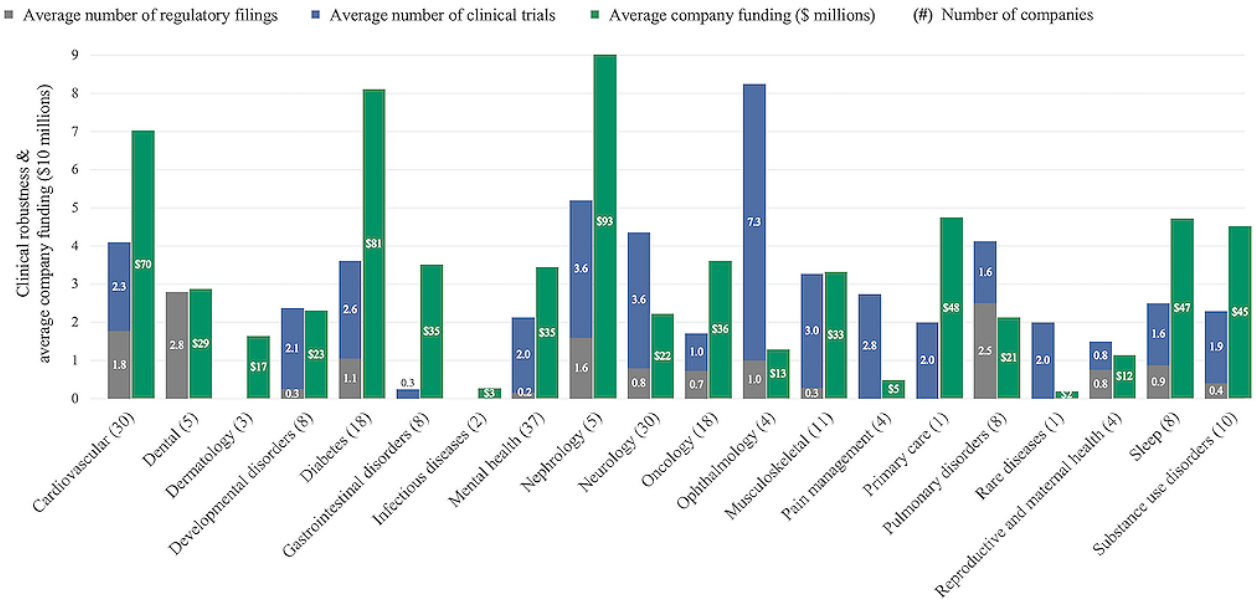


Figure 3. Funding and clinical robustness across various clinical areas.



Companies that sold to employers made more claims and had higher clinical robustness scores compared to other customer types (Figure 1). This may suggest that this customer base is currently the most competitive and/or has the highest entry standards, despite accounting for a relatively small percentage (<12%) of companies in the sample. A competitive employer market for digital health tools is unsurprising given the rising cost burden of health care on employers [32] and evidence that employers are increasingly offering digital health benefits to improve health outcomes [33,34] and contain health care costs [35,36].

Limitations

Our definition of clinical robustness was limited to clinical trials and regulatory findings, which remain proxies for effectiveness and require companies to register their activities. However, we believe these elements offer a better estimation of clinical rigor than publications, which can lag in timing and may not always relate to clinical outcomes. We chose to focus on clinical trials and regulatory filings because they were publicly available data and are generally undertaken to demonstrate an impact on clinical outcomes. Our approach to measuring clinical robustness equally accounts for clinical trials that demonstrate a technology does and does not work. Additionally, our data collection

XSL•FO
RenderX

methodology could have missed clinical trials that were not registered on ClinicalTrials.gov or regulatory filings that were submitted to the FDA under an individual's name rather than the company name (both scenarios are atypical). Future research could incorporate condition- or disease-specific metrics of effectiveness that are then standardized across clinical areas to provide a more accurate measure of clinical impact.

In addition, our cohort only included venture-backed companies above US $2 million in funding. This may have resulted in selection bias by excluding both ends of the spectrum (ie, not including earlier stage companies or large conglomerate technology companies). Although our population excludes these companies, we believe that our sample represents the most comprehensive assessment of outcomes or claims across the digital health landscape and encompasses the majority of funded companies and activity in this space. As a point of reference, the average seed stage deal size in 2021 was US $3.5 million, 75% higher than our minimum funding threshold [37].

Conclusions

Despite the hundreds of digital health companies targeting the myriad of needs across the care continuum, clinical robustness and public communication of claims remains low across much of the sector. These results highlight a significant opportunity for companies to differentiate themselves and for customers to demand greater validation for the products and services they purchase.

Acknowledgments

The authors would like to thank Jasmine DeSilva, Adriana Krasniansky, and Pavan Shah, who are members of the Rock Health research team and manage the Digital Health Venture Funding Database.

Conflicts of Interest

SD, VS, and SK are employees of Rock Health, which supplied data for this analysis. Rock Health is an investor in digital health companies, but no individuals involved in investments were part of this analysis.

Multimedia Appendix 1
Digital health company characteristics across the care continuum, clinical areas, and customer types.
[PDF File (Adobe PDF File), 116 KB - jmir_v24i6e37677_app1.pdf]

Multimedia Appendix 2
Clinical robustness scores, clinical trials, and regulatory filings.
[PDF File (Adobe PDF File), 143 KB - jmir_v24i6e37677_app2.pdf]

Multimedia Appendix 3
Claims made by digital health companies.
[PDF File (Adobe PDF File), 33 KB - jmir_v24i6e37677_app3.pdf]

References

1. Topol EJ. A decade of digital medicine innovation. Sci Transl Med 2019 Jun 26;11(498):eaaw7610. [doi: 10.1126/scitranslmed.aaw7610] [Medline: 31243153]
2. Cohen AB, Dorsey ER, Mathews SC, Bates DW, Safavi K. A digital health industry cohort across the health continuum. NPJ Digit Med 2020 May 12;3(1):68 [FREE Full text] [doi: 10.1038/s41746-020-0276-9] [Medline: 32411829]
3. Rock Health Digital Health Venture Funding Database. URL: https://rockhealth.com/insights/a-new-research-tool-for-a-new-phase-of-digital-health-the-analyses-we-now-support/ [accessed 2020-01-01]
4. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: a path to validation. NPJ Digit Med 2019 May 13;2(1):38 [FREE Full text] [doi: 10.1038/s41746-019-0111-3] [Medline: 31304384]
5. Sedhom R, McShea MJ, Cohen AB, Webster JA, Mathews SC. Mobile app validation: a digital health scorecard approach. NPJ Digit Med 2021 Jul 15;4(1):111 [FREE Full text] [doi: 10.1038/s41746-021-00476-7] [Medline: 34267296]
6. Safavi K, Mathews SC, Bates DW, Dorsey ER, Cohen AB. Top-Funded Digital Health Companies And Their Impact On High-Burden, High-Cost Conditions. Health Aff (Millwood) 2019 Jan;38(1):115-123. [doi: 10.1377/hlthaff.2018.05081] [Medline: 30615535]
7. Cristea IA, Cahan EM, Ioannidis JPA. Stealth research: Lack of peer-reviewed evidence from healthcare unicorns. Eur J Clin Invest 2019 Apr 13;49(4):e13072. [doi: 10.1111/eci.13072] [Medline: 30690709]
8. Wolff J, Pauling J, Keck A, Baumbach J. The economic impact of artificial intelligence in health care: systematic review. J Med Internet Res 2020 Feb 20;22(2):e16866 [FREE Full text] [doi: 10.2196/16866] [Medline: 32130134]
9. Jiang X, Ming W, You JH. The cost-effectiveness of digital health interventions on the management of cardiovascular diseases: systematic review. J Med Internet Res 2019 Jun 17;21(6):e13166 [FREE Full text] [doi: 10.2196/13166] [Medline: 31210136]

XSL•FO
RenderX

10. Larsen ME, Huckvale K, Nicholas J, Torous J, Birrell L, Li E, et al. Using science to sell apps: Evaluation of mental health app store quality claims. NPJ Digit Med 2019 Mar 22;2(1):18 [FREE Full text] [doi: 10.1038/s41746-019-0093-1] [Medline: 31304366]
11. Wisniewski H, Liu G, Henson P, Vaidyam A, Hajratalli NK, Onnela J, et al. Understanding the quality, effectiveness and attributes of top-rated smartphone health apps. Evid Based Ment Health 2019 Feb 11;22(1):4-9 [FREE Full text] [doi: 10.1136/ebmental-2018-300069] [Medline: 30635262]
12. 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]
13. Lumosity to pay $2 million to settle FTC deceptive advertising charges for its 'brain training' program. Federal Trade Commission. 2016. URL: https://www.ftc.gov/news-events/news/press-releases/2016/01/lumosity-pay-2-million-settle-ftc-deceptive-advertising-charges-its-brain-training-program [accessed 2022-06-09]
14. FTC Charges Marketers of ‘Vision Improvement’ App With Deceptive Claims. Federal Trade Commission. 2015. URL: https://www.ftc.gov/news-events/news/press-releases/2015/09/ftc-charges-marketers-vision-improvement-app-deceptive-claims [accessed 2022-06-09]
15. A.G. Schneiderman Announces Settlements With Three Mobile Health Application Developers For Misleading Marketing And Privacy Practices. New York State Office of the Attorney General. 2017. URL: https://ag.ny.gov/press-release/2017/ag-schneiderman-announces-settlements-three-mobile-health-application-developers [accessed 2022-06-09]
16. Kitsiou S, Paré G, Jaana M, Gerber B. Effectiveness of mHealth interventions for patients with diabetes: An overview of systematic reviews. PLoS One 2017 Mar 1;12(3):e0173160 [FREE Full text] [doi: 10.1371/journal.pone.0173160] [Medline: 28249025]
17. 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]
18. 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]
19. Hewitt S, Sephton R, Yeowell G. The effectiveness of digital health interventions in the management of musculoskeletal conditions: systematic literature review. J Med Internet Res 2020 Jun 05;22(6):e15617 [FREE Full text] [doi: 10.2196/15617] [Medline: 32501277]
20. Osborn J, Ajakaiye A, Cooksley T, Subbe CP. Do mHealth applications improve clinical outcomes of patients with cancer? A critical appraisal of the peer-reviewed literature. Support Care Cancer 2020 Mar 4;28(3):1469-1479 [FREE Full text] [doi: 10.1007/s00520-019-04945-4] [Medline: 31273501]
21. Hilty DM, Ferrer DC, Parish MB, Johnston B, Callahan EJ, Yellowlees PM. The effectiveness of telemental health: a 2013 review. Telemed J E Health 2013 Jun;19(6):444-454 [FREE Full text] [doi: 10.1089/tmj.2013.0075] [Medline: 23697504]
22. Feldman N, Back D, Boland R, Torous J. A systematic review of mHealth application interventions for peripartum mood disorders: trends and evidence in academia and industry. Arch Womens Ment Health 2021 Dec 30;24(6):881-892 [FREE Full text] [doi: 10.1007/s00737-021-01138-z] [Medline: 33929636]
23. Chen CE, Harrington RA, Desai SA, Mahaffey KW, Turakhia MP. Characteristics of Digital Health Studies Registered in ClinicalTrials.gov. JAMA Intern Med 2019 Jun 01;179(6):838-840 [FREE Full text] [doi: 10.1001/jamainternmed.2018.7235] [Medline: 30801617]
24. Pathak A, Bain S, Pacifici E. 4230 An app a day: examining clinical evidence for safety and efficacy of diabetes mobile health apps. J Clin Trans Sci 2020 Jul 29;4(s1):53-54. [doi: 10.1017/cts.2020.191]
25. Safavi K, Mathews SC, Bates DW, Dorsey ER, Cohen AB. Top-funded digital health companies and their impact on high-burden, high-cost conditions. Health Aff (Millwood) 2019 Jan;38(1):115-123. [doi: 10.1377/hlthaff.2018.05081] [Medline: 30615535]
26. Cohen AB, Dorsey ER, Mathews SC, Bates DW, Safavi K. A digital health industry cohort across the health continuum. NPJ Digit Med 2020 May 12;3(1):68 [FREE Full text] [doi: 10.1038/s41746-020-0276-9] [Medline: 32411829]
27. Klonoff DC, Evans B, Zweig M, Day S, Kerr D. Is digital health for diabetes in an investment bubble? J Diabetes Sci Technol 2020 Jan 30;14(1):165-169 [FREE Full text] [doi: 10.1177/1932296819867742] [Medline: 31470739]
28. What digital health is (and isn’t). Rock Health. 2013. URL: https://rockhealth.com/what-digital-health-is-and-isnt/ [accessed 2021-12-22]
29. What is digital health? U.S. Food and Drug Administration. 2020. URL: https://www.fda.gov/medical-devices/digital-health-center-excellence/what-digital-health [accessed 2021-12-22]
30. Damodaran A. Valuing young, start-up and growth companies: estimation issues and valuation challenges. SSRN Journal 2009:1-67. [doi: 10.2139/ssrn.1418687]
31. Wiederhold BK. Femtech: digital help for women's health care across the life span. Cyberpsychol Behav Soc Netw 2021 Nov 01;24(11):697-698. [doi: 10.1089/cyber.2021.29230.editorial] [Medline: 34806914]

XSL•FO
RenderX

32. Keehan S, Cuckler G, Poisal J, Sisko A, Smith S, Madison A, et al. National health expenditure projections, 2019-28: expected rebound in prices drives rising spending growth. Health Aff (Millwood) 2020 Apr;39(4):704-714. [doi: 10.1377/hlthaff.2020.00094] [Medline: 32207998]
33. Howarth A, Quesada J, Silva J, Judycki S, Mills PR. The impact of digital health interventions on health-related outcomes in the workplace: A systematic review. Digit Health 2018 May 10;4:2055207618770861 [FREE Full text] [doi: 10.1177/2055207618770861] [Medline: 29942631]
34. Schaetz L, Rimner T, Pathak P, Fang J, Chandrasekhar D, Mueller J, et al. Employee and employer benefits from a migraine management program: disease outcomes and cost analysis. Headache 2020 Oct 16;60(9):1947-1960 [FREE Full text] [doi: 10.1111/head.13933] [Medline: 32799346]
35. Horstman CM, Ryan DH, Aronne LJ, Apovian CM, Foreyt JP, Tuttle HM, et al. Return on investment: medical savings of an employer-sponsored digital intensive lifestyle intervention for weight loss. Obesity (Silver Spring) 2021 Apr 23;29(4):654-661 [FREE Full text] [doi: 10.1002/oby.23117] [Medline: 33759385]
36. Cunningham-Hill M, Dodge-Rice Z, Wilson-Myers C, Sherman C, Neary M, Schueller S. Digital tools and solutions for mental health: an employer's guide. Northeast Business Group on Health and One Mind PsyberGuide. 2020 May 04. URL: https://online.flippingbook.com/view/911432/ [accessed 2022-06-13]
37. Pitchbook-NVCA venture monitor. Pitchbook. URL: https://nvca.org/research/pitchbook-nvca-venture-monitor/ [accessed 2020-05-02]

Abbreviations

FDA: Food and Drug Administration

Edited by A Mavragani; submitted 02.03.22; peer-reviewed by J Andrews, J Segel, A Meyers; comments to author 05.04.22; revised version received 25.05.22; accepted 07.06.22; published 20.06.22.

Please cite as:
Day S, Shah V, Kaganoff S, Powelson S, Mathews SC
Assessing the Clinical Robustness of Digital Health Startups: Cross-sectional Observational Analysis
J Med Internet Res 2022;24(6):e37677
URL: https://www.jmir.org/2022/6/e37677
doi:10.2196/37677
PMID:35723914

©Sean Day, Veeraj Shah, Sari Kaganoff, Shannon Powelson, Simon C Mathews. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.06.2022. 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

mHealth Research for Weight Loss, Physical Activity, and Sedentary Behavior: Bibliometric Analysis

Chieh-Chen Wu[1,2], PhD; Chih-Wei Huang[3], PhD; Yao-Chin Wang[4], PhD; Md.Mohaimenul Islam[3,5], PhD; Woon-Man Kung[1], MD; Yung-Ching Weng[2], PhD; Chun-Hsien Su[6], PhD

[1]Department of Exercise and Health Promotion, College of Kinesiology and Health, Chinese Culture University, Taipei, Taiwan

[2]Department of Healthcare Information and Management, School of Health Technology, Ming Chuan University, Taipei, Taiwan

[3]International Center for Health Information Technology, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan

[4]Graduate Institute of Injury Prevention and Control, Taipei Medical University, Taipei, Taiwan

[5]AESOP Technology, Taipei, Taiwan

[6]Graduate Institute of Sport Coaching Science, College of Kinesiology and Health, Chinese Culture University, Taipei, Taiwan

Corresponding Author:
Chun-Hsien Su, PhD
Graduate Institute of Sport Coaching Science
College of Kinesiology and Health
Chinese Culture University
No. 55, Hwa-Kang Road, Yang-Ming-Shan
Taipei, 11114
Taiwan
Phone: 886 2 2861 0511
Email: chsu@ulive.pccu.edu.tw

Abstract

Background: Research into mobile health (mHealth) technologies on weight loss, physical activity, and sedentary behavior has increased substantially over the last decade; however, no research has been published showing the research trend in this field.

Objective: The purpose of this study was to provide a dynamic and longitudinal bibliometric analysis of recent trends of mHealth research for weight loss, physical activity, and sedentary behavior.

Methods: A comprehensive search was conducted through Web of Science to retrieve all existing relevant documents published in English between January 1, 2010, and November 1, 2021. We developed appropriate research questions; based on the proven bibliometric approaches, a search strategy was formulated to screen the title for eligibility. Finally, we conducted bibliometric analyses to explore the growth rate of publications; publication patterns; and the most productive authors, institutions, and countries, and visualized the trends in the field using a keyword co-occurrence network.

Results: The initial search identified 8739 articles, of which 1035 were included in the analyses. Our findings show an exponential growth trend in the number of annual publications of mHealth technology research in these fields. *JMIR mHealth and uHealth* (n=214, 20.67%), *Journal of Medical Internet Research* (n=71, 6.86%), and *BMC Public Health* (n=36, 3.47%) were the top 3 journals, publishing higher numbers of articles. The United States remained the leading contributor in these areas (n=405, 39.13%), followed by Australia (n=154, 14.87%) and England (n=125, 12.07%). Among the universities, the University of Sydney (n=36, 3.47%) contributed the most mHealth technology research in these areas; however, Deakin University (n=25, 2.41%) and the National University of Singapore (n=23, 2.22%) were in the second and third positions, respectively.

Conclusions: Although the number of papers published on mobile technologies for weight loss, physical activity, and sedentary behavior was initially low, there has been an overall increase in these areas in recent years. The findings of the study indicate that mobile apps and technologies have substantial potential to reduce weight, increase physical activity, and change sedentary behavior. Indeed, this study provides a useful overview of the publication trends and valuable guidance on future research directions and perspectives in this rapidly developing field.

(J Med Internet Res 2022;24(6):e35747) doi:10.2196/35747

XSL•FO
RenderX

KEYWORDS

mobile health; weight loss; physical activity; sedentary behavior; bibliometric analysis; mHealth; weight; behavior; research; literature; bibliometric; journal; trend; app

Introduction

Mobile health (mHealth) has emerged as a potential tool to support physicians and patients in many areas [1,2]. Recent evidence shows that mHealth is an easily accessible and cost-effective tool to assist in improving health outcomes [3]. The widespread availability of mobile phones paves the way to conduct advanced research in health care. mHealth-related research is thriving and gaining in popularity [4]. Over the past few years, it has become clear that mHealth technologies (eg, apps or SMS) can help to reduce weight loss, improve physical activity, and change behavior [5-8]. Previous reviews evaluated the effectiveness of mHealth interventions in these three domains [9-12]. These reviews suggest that mHealth interventions appear to be promising for preventive and therapeutic activities. Given the numerous mHealth-related publications, it is important to analyze these research studies to provide an overview of these domains.

Bibliometric analysis is considered a popular and rigorous statistical method for exploring and analyzing a large volume of scientific literature [13]. It can identify the main themes and emerging trends of certain research topics, and detect knowledge in the literature [14]. Bibliometric analyses that summarized the research landscape in various fields have generated valuable insights [14-18], allowing researchers to study specific research areas by analyzing citations, cocitations, geographical distribution, and word frequency, and by providing insightful conclusions [19]. Thus, bibliometric analyses are contributing to monitoring the development and patterns of effective publications. Moreover, bibliometric analyses help researchers, clinicians, and health care policy makers to collect information to understand the particular area of research and their applications, and to promote interdisciplinary collaborations [20].

The aim of this study was to provide a comprehensive picture of mHealth-related research and the direction of future research to benefit the general population, health care policy makers, and researchers.

Based on the research scope and objectives, we developed the following research questions:

- RQ1: What are the basic characteristics of the publications? How many articles on “mobile technologies” for “weight loss, physical activity, and sedentary behavior” have been published between 2010 and 2021?
- RQ2: Who are the most productive authors/coauthors in these areas, and what were the countries of origin?
- RQ3: Which journal published the most? Which organizations mainly contributed to this area?
- RQ4: What are the most frequent cowords (titles/abstract/keywords) associated with these publications?

Methods

Search Strategy

We searched for potential publications in Web of Science (WoS) with terms related to mHealth technology, weight loss, physical activity, and sedentary behavior. However, we conducted a comprehensive search in WoS only because it is an extensive database of approximately 10,000 prestigious and high-impact research journals. Moreover, WoS has now been widely used and is the most reliable database for conducting bibliometric analyses [20-22]. WoS contains the following information: title, author, institution, country/region, publication year, citation history, funding source, research types, and keywords [23]. A comprehensive search strategy is presented in Multimedia Appendix 1, Table S1.

Inclusion and Exclusion Criteria

In our study, all journal articles about mHealth for these three topics were included for screening. The articles were included in the analysis if they (1) were written in the English language; (2) focused on weight loss, physical activity, and/or sedentary behavior; and (3) were involved in mobile technologies. As mHealth is a leading-edge and rapidly updated research area, research or review articles published in peer-reviewed journals, conference proceedings, and early access articles were included. However, letters, editorials, book chapters, and books were excluded from the bibliometric analysis.

Screening Strategy

Two authors independently screened all the titles and abstracts of retrieved articles and checked the validity of those articles. Any confusion at this stage was resolved by discussing with a third author. Finally, data were collected from selected articles and saved in TXT formats.

Bibliometric Analysis

We aimed to provide a holistic view of mHealth research on these topics to obtain the knowledge structure, potential authors, research trends, most prolific country and institutions, and research hot spots. Bibliometric analysis was used to show bibliometric maps and the graphical representation of such maps.

Growth Rate of Publications

The annual number of publications, annual growth, and average growth rate of publications were calculated in the following ways:

Where P is the total number of publications in the current year, and P_{n-1} is the total number of publications in the previous year.

XSL•FO
RenderX

Publication Patterns

In this study, we also analyzed the publication trends by countries (top 10 most prolific countries), distribution of source journals (top 10 most productive journals), distribution and coauthorship of institutions (top 10 institutions), and distribution of authors (top 10 most productive authors). The rank of the country, journal, institutions, and authors was selected based on the number of publications.

Research Hot Spot Tendencies

We developed citation bursts and a timeline map using the VOSviewer software (Centre for Science and Technology Studies, Leiden University). We also constructed and visualized clusters based on publications between 2010 and 2021. However, each cluster was labeled by the keywords provided by the included articles. The top 100 keywords were selected for mapping with their co-occurrence in 5 clusters. A circle with a label illustrates each node in the keyword map: the bigger circles represent higher frequencies. The color of each circle indicates which cluster it belongs to. Finally, the thickness and length of links between nodes show their association strength.

Results

Publication Outputs

Based on our comprehensive search on WoS, we identified a total of 8739 articles on mHealth technologies in the three areas (weight loss, physical activity, and/or sedentary behavior). After removing 7704 articles, 1035 articles remained (Table 1). The reasons for the exclusion of studies are given in Multimedia Appendix 1, Figure S1. The number of annual publications on mHealth technologies in these domains increased from 7 articles in 2010 to 173 articles in 2021 (Figure 1). Before 2018, the number of annual articles did not reach 100. The average annual growth rate of articles was a maximum of 228.57% in 2012 and showed a –7.93% decline in 2021.

Table 1. The distribution of articles by year between 2010 and 2021.

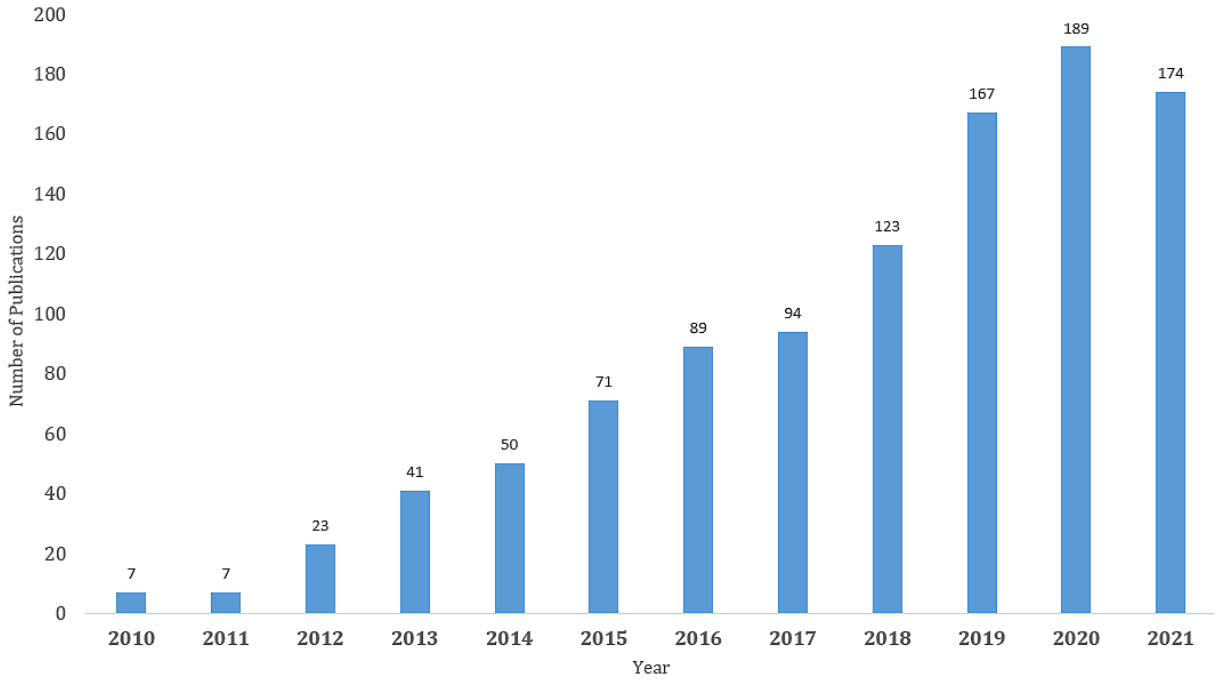
Year	Publication, n	Annual growth, n	AGR[a] (%)
2010	7	N/A[b]	N/A
2011	7	0	0.00
2012	23	16	228.57
2013	41	18	78.26
2014	50	9	21.95
2015	71	21	42.00
2016	89	18	25.35
2017	94	5	5.61
2018	123	29	30.85
2019	167	44	35.77
2020	189	22	13.17
2021	174	–15	–7.93

[a]AGR: average growth rate.

[b]N/A: not applicable.

XSL•FO
RenderX

Figure 1. Number of publications on mobile health technologies in these areas between 2010 and 2021.



Distribution of Source Journals

There were a total of 337 journals that published articles on mHealth technologies in these three domains. However, the Canadian *JMIR mHealth and uHealth* was the most productive journal, publishing 214 (20.67%) articles in these three domains (Table 2). *Journal of Medical Internet Research*, *BMC Public Health*, and *International Journal of Environmental Research and Public Health* were in the second, third, and fourth positions, publishing 71, 36, and 30 articles, respectively, on these topics. The top 10 journals published 441 articles, accounting for 42.6% (441/1035) of all publications in these domains.

Table 2. Top 10 journals that published articles on mobile health technologies for these three domains, 2010-2021.

Rank	Journal	Country	Categories	Publication (N=1035), n (%)
1	JMIR mHealth and uHealth	Canada	Medical informatics	214 (20.67)
2	Journal of Medical Internet Research	Canada	Medical informatics	71 (6.86)
3	BMC Public Health	England	Public, environmental, and occupational health	36 (3.47)
4	International Journal of Environmental Research and Public Health	Switzerland	Public, environmental, and occupational health	30 (2.89)
5	Translational Behavioral Medicine	Switzerland	Public, environmental, and occupational health	21 (2.02)
6	BMJ Open	England	Medicine (general and internal)	18 (1.73)
7	International Journal of Behavioral Nutrition and Physical Activity	England	Nutrition and dietetics	16 (1.54)
8	PLOS One	United States	Multidisciplinary science	16 (1.54)
9	American Journal of Preventive Medicine	United States	Public, environmental, and occupational health	15 (1.42)
10	Digital Health	England	Public, environmental, and occupational health	14 (1.35)

Distribution of Coauthorship of Countries/Regions

Our study showed that researchers from 73 countries and regions conducted research on these topics and published articles in different international peer-reviewed journals (Figure 2). Of the total 1035 articles, the United States contributed the highest number (n=405, 39.13%), followed by Australia (n=154, 14.87%), England (n=125, 12.07%), China (n=68, 6.57%), Spain (n=60, 5.79%), and Canada (n=52, 5.02%) (Table 3).

XSL•FO
RenderX

Figure 2. The coauthorship network of countries/regions that published at least one article in these domains, 2010-2021. Peoples R China: People's Republic of China; U Arab Emirates: United Arab Emirates.

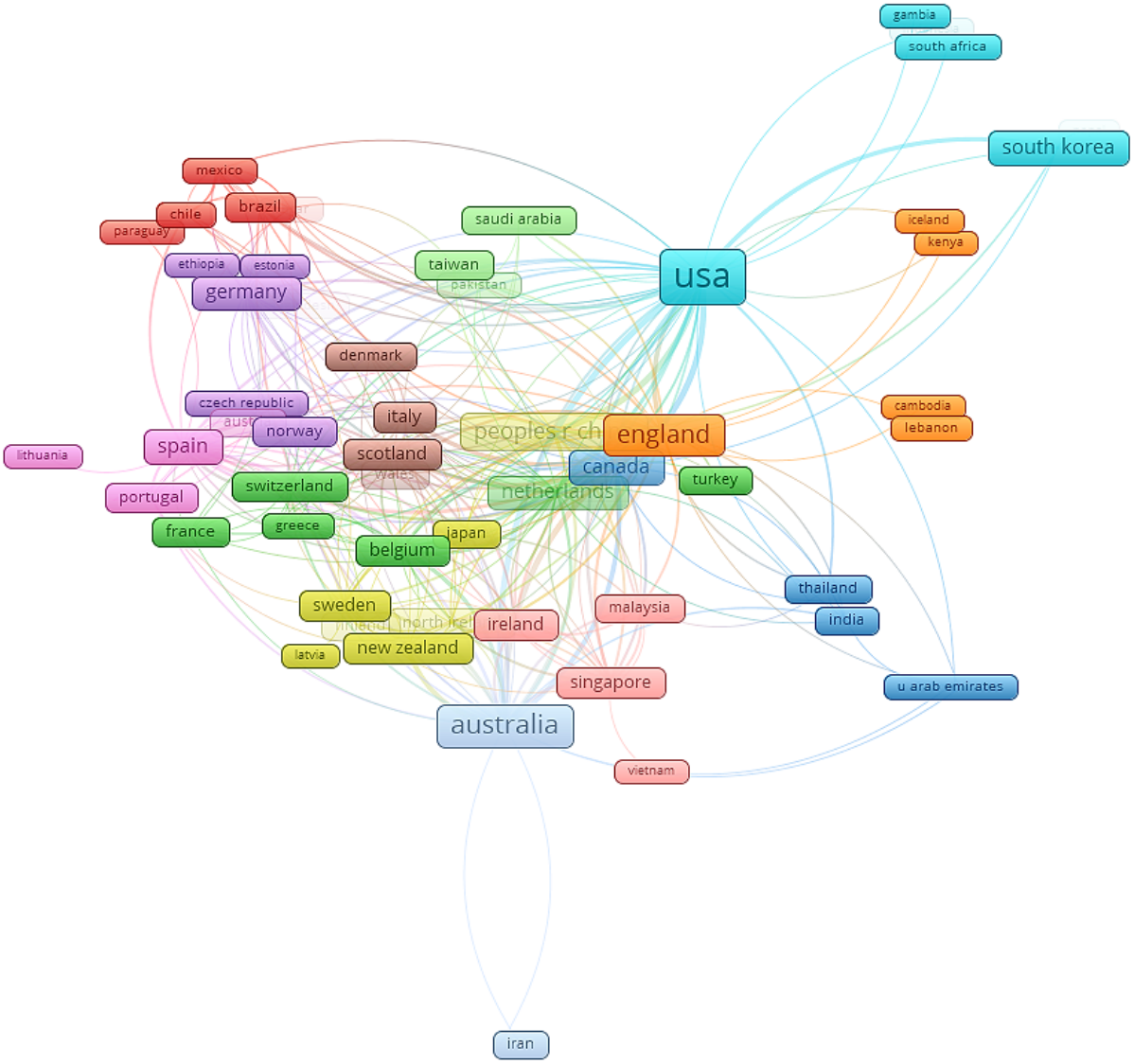


Table 3. Top 10 countries that published articles on mobile health technologies for these three domains, 2010-2021.

Rank	Country	Publications, n	Citations, n
1	United States	405	12,672
2	Australia	154	4301
3	England	125	4602
4	China	68	819
5	Spain	60	1115
6	Canada	52	876
7	South Korea	49	564
8	Netherlands	43	1383
9	Germany	36	519
10	Ireland	31	571

XSL•FO
RenderX

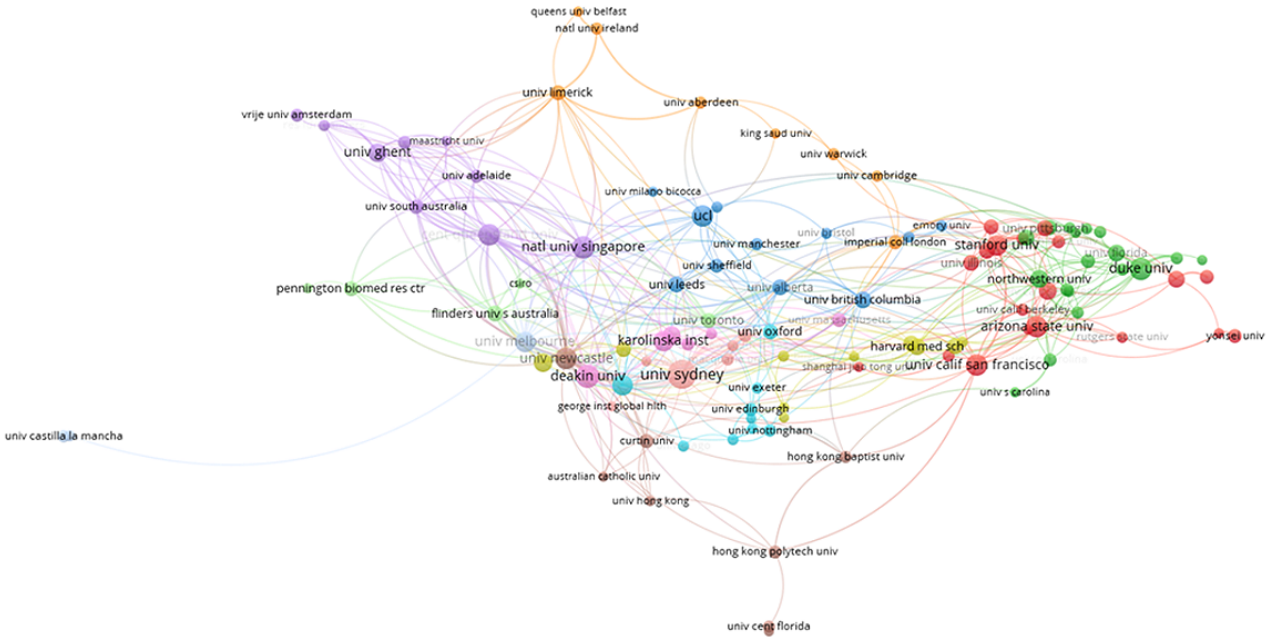
Distribution of Coauthorship of Institutions

According to our study findings, 1494 institutes contributed to at least one study. Table 4 shows the top 10 most productive research institutes that used mHealth technologies in these domains. The University of Sydney (36 articles) ranked first among all research institutions, followed by Deakin University (25 articles), the National University of Singapore (23 articles), and Duke University (22 articles). Figure 3 displays the coauthorship analysis of 117 institutions that published at least 5 articles. It forms a total of 12 clusters (cluster 1, red color, 19 institutions; cluster 2, blue color, 18 institutions; and cluster 12, ash color, 3 institutions), differentiated by various color.

Table 4. Top 10 institutions that published papers on mobile health technologies for these three domains, 2010-2021.

Rank	Institution	Country	Publications, n	Citations, n
1	University of Sydney	Australia	36	1009
2	Deakin University	Australia	25	413
3	National University of Singapore	Singapore	23	324
4	Duke University	United States	22	973
5	Central Queensland University	Australia	21	1004
6	University of California, San Francisco	United States	21	569
7	University of Newcastle	Australia	20	754
8	The University of California, Los Angeles	United States	20	1613
9	Arizona State University	United States	20	504
10	University of Auckland	New Zealand	19	882

Figure 3. The coauthorship network of institutions that contributed at least 5 articles in these domains, 2010-2021.



Distribution and Coauthorship of Authors

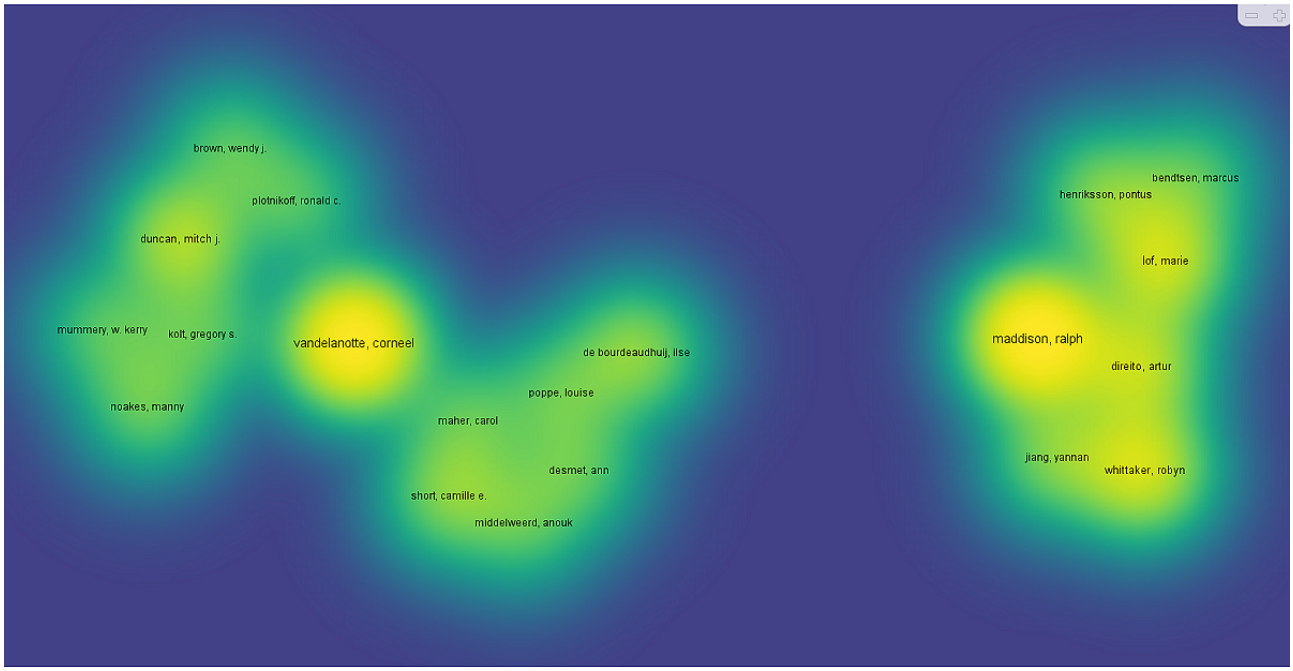
Based on our study, 1035 articles were published by 4976 authors with at least one article. Table 5 shows the top 10 most productive authors who conducted and published articles in these domains. Ralph M ranked highest among all authors (23 articles), followed by Corneel V (22 articles), Robyn W (13 articles), and Marie L (12 articles). Our analysis shows that 43 of 4976 authors had published at least 5 articles. The largest set of associated authors consisted of 20 authors in 3 clusters (Figure 4).

XSL•FO
RenderX

Table 5. Top 10 authors that published papers on mobile health technologies for these domains, 2010-2021.

Rank	Authors	Publications, n	Citations, n
1	Ralph M	23	775
2	Corneel V	22	1124
3	Robyn W	13	665
4	Marie L	12	68
5	Artur D	11	516
6	Mitch JD	10	551
7	Yoshimi F	9	315
8	Ilse DB	8	85
9	Pontus H	7	29
10	Yannan J	7	290

Figure 4. The coauthorship network of authors who contributed research regarding mobile health technologies for these domains, 2010-2021.

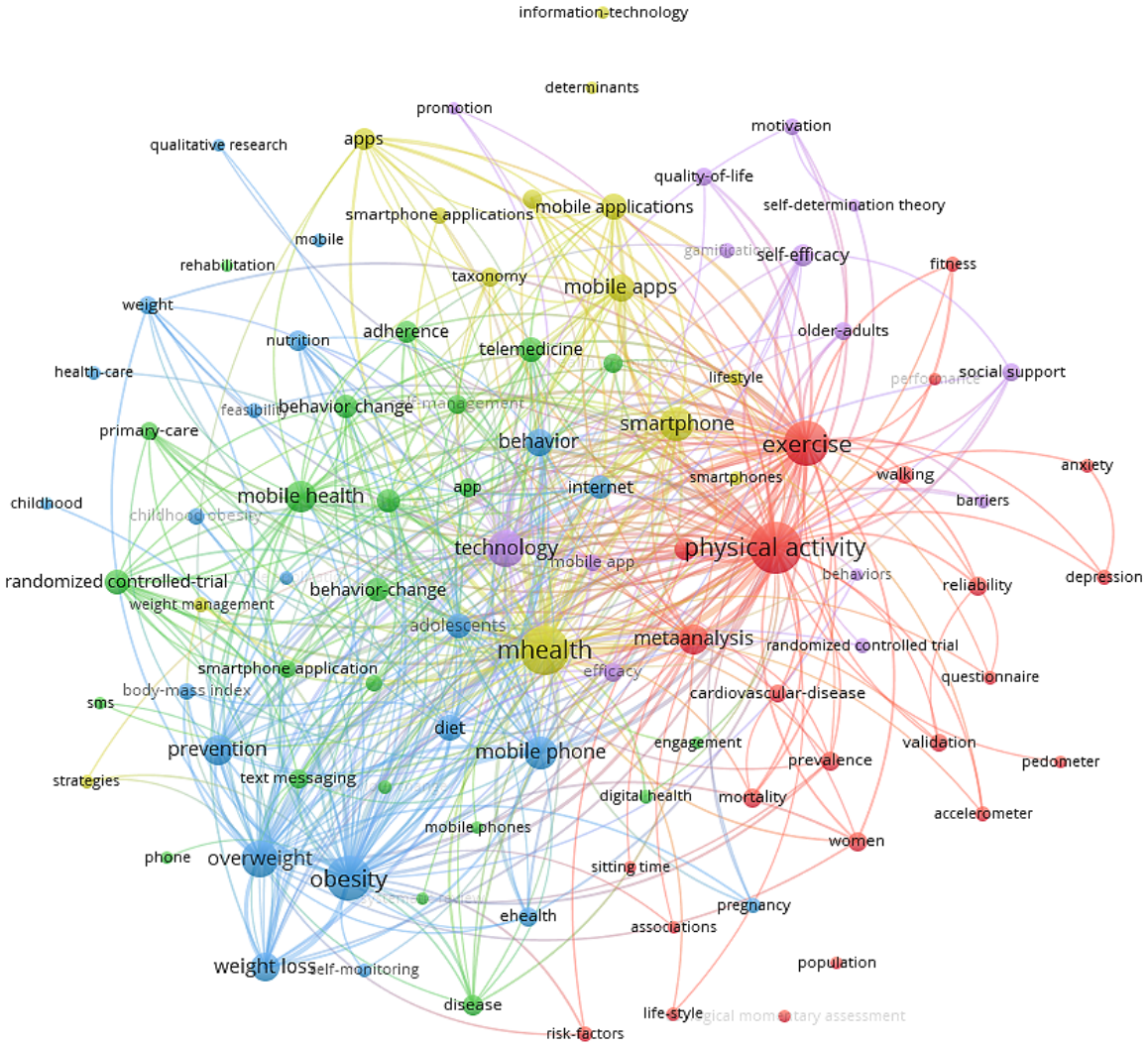


Co-occurrence Analysis of Top 100 Keywords

The top 100 keywords were classified into 5 clusters using keyword-clustering analysis (Figure 5). The five most common keywords for mHealth technologies were physical activity (n=282), mHealth (n=260), exercise (n=220), obesity (n=220), and health (n=220). The 5 clusters are represented by color: red (cluster 1), green (cluster 2), blue (cluster 3), yellow (cluster 4), and violet (cluster 5). The node labels are the keywords, and the node size depends on the number of keyword co-occurrences. The links connecting two nodes show a co-occurrence relationship between the keywords.

XSL•FO
RenderX

Figure 5. The co-occurrence network of the top 100 keywords in these domains, 2010-2021.



Discussion

Main Findings

The use of mHealth technologies has been rapidly increasing. mHealth technologies are virtually ubiquitous and provide great opportunities to deliver health care management and health promotion to people in low-resource settings and with limited access to care [24-26]. It also facilitates the tracking of activity, behavior, and weight, allowing for real-time recording, feedback, and accountability [27]. Previous studies have shown that mHealth technologies support a reduction in obesity, improving physical activity and reducing the risk of developing life-threatening complications [28-30].

This study explored the bibliometric analyses of mHealth intervention research in weight loss, physical activity, and sedentary behavior. By analyzing the results of the data set regarding the publication pattern on research related to mHealth technology with these areas, we found a rapid increase in the growth of interest in this subject in the last decade. The trend substantially increased in the most recent years, from 2018 to 2021. This increased trend of publication could reflect the improvements, functionality, and developments of mHealth research in parallel with comparison to other areas of use. Therefore, mHealth offers patients many high-quality choices in the self-management of their chronic diseases, provides new opportunities and convenience, and addresses potential benefits. However, an effective cooperation among clinical experts, mHealth technology developers, and providers must achieve the ultimate goal and objectives. The primary objective of this study was to highlight the most frequent subject categories, along with popular keywords and terms. These aspects reflected recent research trends and influential articles, authors, and collaboration networks among researchers and institutions. Our findings show that the United States was the most substantial contributor, and 4 of the 10 most productive institutions were from the United States. The *Journal of Medical Internet Research* was also the leading journal publishing research in these three areas as well as a bibliometric study that examined the overall mHealth literature [15].

Global Trends in mHealth Research

The emergence of mHealth technologies has opened up new opportunities to improve patient care in many ways. The obstacles of remote locations and accessibility to a wide range of populations, especially in areas with minimum medical

XSL•FO
RenderX

resources, have been reported. The publication growth trends on these domains between 2010 and 2021 had increased exponentially. This suggests that the acceptability of mHealth research on these domains has been increased, and research into mHealth progressed relatively higher. Since the number of mHealth users is gradually increasing, more and more researchers are focusing on these domains, while the number of mHealth-related publications are showing increased trends [31,32]. The number of research papers published in these areas since 2010 was more than 1000. For journal sources, the top 3 journals publishing mHealth in these domains belong to the area of medical informatics and public health. As mHealth technologies evaluated weight management, increasing physical activity, and changing sedentary behavior, the authors preferred top-ranking and reputed journals in medical informatics and public-related journals. *JMIR mHealth and uHealth*, the *Journal of Medical Internet Research*, and *BMC Public Health* are the most popular open-access journals and the most popular outlet for researchers in these fields. Moreover, journals that publish open access obtain a higher number of citations than non–open access journals [33,34].

International Trends

The findings of our study show that the United States remained the most productive country in these areas. The number of annual publications in the United States had steadily increased. This is followed by Australia, England, China, Spain, and Canada, which were experiencing the rapid growth of research in these areas using mHealth technologies. The vast majority (approximately 97%) of American adults own a mobile phone [35,36], which opens up the opportunity to conduct research that is particularly relevant to the management of chronic diseases [37,38]. A previous study reported that more than 40% of US adults have two or more chronic diseases [39] and 71% of all US health care spending is for chronic diseases [40]. The prevalence of obesity in the United States, England, and Australia are 37.7% [41], 28% [42], and 26% [43], respectively. Obesity and sedentary behavior are the major risk factors for several chronic conditions such as dyslipidemia [44], cardiovascular disease [45], and cancer [46]. These conditions are also associated with detrimental psychological, social, and economic consequences [47]. Therefore, mHealth technologies have become more attractive for conducting research in these domains. The five most productive countries had potential research collaboration in these domains. A coauthorship network always reflects the collaborative relationship among researchers and provides possible opportunities for other researchers to collaborate. The most productive European countries (England, Netherlands, Spain, Ireland, and Norway) had close collaboration. However, they had a strong collaboration with the United States and Australia. Moreover, Asian countries such as China, Korea, and India were the most productive countries and established strong research collaboration with the United States, England, Netherland, Canada, and Australia. However, international research collaboration always depends on several key factors such as international relations, geography, and political and economic alliances [25].

Limitations

Even though our study provided a comprehensive picture of mHealth research on weight loss, physical activity, and sedentary behavior, there are still several limitations that need to be addressed. First, we collected data from a single database. Although, WoS is an extensive database that offers a wide variety of publications needed for the comprehensive analysis of any topic [25,48]. However, future studies might include other popular databases such as Scopus or PubMed to include more potential studies. Second, we included only studies published in English; however, we might have missed some publications due to this language restriction. Third, we did not consider gray articles and published material such as meeting abstracts, letters, and editorials. Finally, we were unable to conduct individual analyses of weight loss, physical activity, and sedentary behavior due to the limited number of studies.

Conclusion

We aimed to present a clear picture of mHealth-related research in weight loss, physical activity, and sedentary behavior. Likely, the findings of this study show that the growth rate of mHealth research regarding these three domains has been rapid in the past decade and the annual growth rate is continuously growing; high-income countries like the United States, England, and Australia are the main force behind mHealth-related research on these topics; and medical informatics journals such as *JMIR mHealth and uHealth* and the *Journal of Medical Internet Research* are the top contributors to this topic based on the amount of articles published. Since mHealth research into health care, including these three topics, is accelerating rapidly, these findings can assist researchers and health care policy makers in taking proper directions in future research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary table and figure.
[DOCX File , 23 KB - jmir_v24i6e35747_app1.docx]

References

1. Islam MM, Poly TN, Walther BA, Jack Li YC. Use of mobile phone app interventions to promote weight loss: meta-analysis. JMIR Mhealth Uhealth 2020 Jul 22;8(7):e17039 [FREE Full text] [doi: 10.2196/17039] [Medline: 32706724]

XSL•FO
RenderX

2. Silva BM, Rodrigues JJ, de la Torre Díez I, López-Coronado M, Saleem K. Mobile-health: a review of current state in 2015. J Biomed Inform 2015 Aug;56:265-272 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]
3. Kahn JG, Yang JS, Kahn JS. 'Mobile' health needs and opportunities in developing countries. Health Aff (Millwood) 2010 Feb;29(2):252-258. [doi: 10.1377/hlthaff.2009.0965] [Medline: 20348069]
4. Gilmore LA, Klempel MC, Martin CK, Myers CA, Burton JH, Sutton EF, et al. Personalized mobile health intervention for health and weight loss in postpartum women receiving women, infants, and children benefit: a randomized controlled pilot study. J Womens Health (Larchmt) 2017 Jul;26(7):719-727 [FREE Full text] [doi: 10.1089/jwh.2016.5947] [Medline: 28338403]
5. Jatobá LC, Grossmann U, Kunze C, Ottenbacher J, Stork W. Context-aware mobile health monitoring: evaluation of different pattern recognition methods for classification of physical activity. Annu Int Conf IEEE Eng Med Biol Soc 2008;2008:5250-5253. [doi: 10.1109/IEMBS.2008.4650398] [Medline: 19163901]
6. Lee AM, Chavez S, Bian J, Thompson LA, Gurka MJ, Williamson VG, et al. Efficacy and effectiveness of mobile health technologies for facilitating physical activity in adolescents: scoping review. JMIR Mhealth Uhealth 2019 Feb 12;7(2):e11847 [FREE Full text] [doi: 10.2196/11847] [Medline: 30747716]
7. Buckingham SA, Williams AJ, Morrissey K, Price L, Harrison J. Mobile health interventions to promote physical activity and reduce sedentary behaviour in the workplace: a systematic review. Digit Health 2019;5:2055207619839883 [FREE Full text] [doi: 10.1177/2055207619839883] [Medline: 30944728]
8. Shaw RJ, Bosworth HB, Silva SS, Lipkus IM, Davis LL, Sha RS, et al. Mobile health messages help sustain recent weight loss. Am J Med 2013 Nov;126(11):1002-1009 [FREE Full text] [doi: 10.1016/j.amjmed.2013.07.001] [Medline: 24050486]
9. Jung J, Cho I. Promoting physical activity and weight loss with mHealth interventions among workers: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2022 Jan 21;10(1):e30682 [FREE Full text] [doi: 10.2196/30682] [Medline: 35060913]
10. Flores Mateo G, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile phone apps to promote weight loss and increase physical activity: a systematic review and meta-analysis. J Med Internet Res 2015 Nov 10;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
11. Yerrakalva D, Yerrakalva D, Hajna S, Griffin S. Effects of mobile health app interventions on sedentary time, physical activity, and fitness in older adults: systematic review and meta-analysis. J Med Internet Res 2019 Nov 28;21(11):e14343 [FREE Full text] [doi: 10.2196/14343] [Medline: 31778121]
12. Compernolle S, DeSmet A, Poppe L, Crombez G, De Bourdeaudhuij I, Cardon G, et al. Effectiveness of interventions using self-monitoring to reduce sedentary behavior in adults: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2019 Aug 13;16(1):63 [FREE Full text] [doi: 10.1186/s12966-019-0824-3] [Medline: 31409357]
13. Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: an overview and guidelines. J Business Res 2021 Sep;133:285-296. [doi: 10.1016/j.jbusres.2021.04.070]
14. Guo Y, Hao Z, Zhao S, Gong J, Yang F. Artificial intelligence in health care: bibliometric analysis. J Med Internet Res 2020 Jul 29;22(7):e18228 [FREE Full text] [doi: 10.2196/18228] [Medline: 32723713]
15. Sweileh WM, Al-Jabi SW, AbuTaha AS, Zyoud SH, Anayah FMA, Sawalha AF. Bibliometric analysis of worldwide scientific literature in mobile - health: 2006-2016. BMC Med Inform Decis Mak 2017 May 30;17(1):72 [FREE Full text] [doi: 10.1186/s12911-017-0476-7] [Medline: 28558687]
16. López-Robles JR, Otegi-Olaso JR, Porto-Gómez I. Bibliometric analysis of worldwide scientific literature in Project Management Techniques and Tools over the past 50 years: 1967-2017. 2018 Presented at: Research and Education in Project Management 2018; February 22-24, 2018; Bilbao, Spain.
17. Adunlin G, Diaby V, Xiao H. Application of multicriteria decision analysis in health care: a systematic review and bibliometric analysis. Health Expect 2015 Dec;18(6):1894-1905. [doi: 10.1111/hex.12287] [Medline: 25327341]
18. Diaby V, Campbell K, Goeree R. Multi-criteria decision analysis (MCDA) in health care: a bibliometric analysis. Operations Res Health Care 2013 Mar;2(1-2):20-24. [doi: 10.1016/j.orhc.2013.03.001]
19. Liao H, Tang M, Luo L, Li C, Chiclana F, Zeng X. A bibliometric analysis and visualization of medical big data research. Sustainability 2018 Jan 11;10(2):166. [doi: 10.3390/su10010166]
20. Kan W, Chou W, Chien T, Yeh Y, Chou P. The most-cited authors who published papers in JMIR mHealth and uHealth using the authorship-weighted scheme: bibliometric analysis. JMIR Mhealth Uhealth 2020 May 07;8(5):e11567 [FREE Full text] [doi: 10.2196/11567] [Medline: 32379053]
21. 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]
22. 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]
23. Islam M, Poly T, Alsinglawi B, Lin LF, Chien SC, Liu JC, et al. Application of artificial intelligence in COVID-19 pandemic: bibliometric analysis. Healthcare (Basel) 2021 Apr 09;9(4):441 [FREE Full text] [doi: 10.3390/healthcare9040441] [Medline: 33918686]

XSL•FO
RenderX

24. Quintiliani LM, Mann DM, Puputti M, Quinn E, Bowen DJ. Pilot and feasibility test of a mobile health-supported behavioral counseling intervention for weight management among breast cancer survivors. JMIR Cancer 2016 May 09;2(1):e4 [FREE Full text] [doi: 10.2196/cancer.5305] [Medline: 28410174]
25. Cao J, Lim Y, Sengoku S, Guo X, Kodama K. Exploring the shift in international trends in mobile health research from 2000 to 2020: bibliometric analysis. JMIR Mhealth Uhealth 2021 Sep 08;9(9):e31097 [FREE Full text] [doi: 10.2196/31097] [Medline: 34494968]
26. Chen M, Wu T, Lv M, Chen C, Fang Z, Zeng Z, et al. Efficacy of mobile health in patients with low back pain: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2021 Jun 11;9(6):e26095 [FREE Full text] [doi: 10.2196/26095] [Medline: 34114965]
27. Nicklas JM, Leiferman JA, Lockhart S, Daly KM, Bull SS, Barbour LA. Development and modification of a mobile health program to promote postpartum weight loss in women at elevated risk for cardiometabolic disease: single-arm pilot study. JMIR Form Res 2020 Apr 09;4(4):e16151 [FREE Full text] [doi: 10.2196/16151] [Medline: 32271149]
28. Stork MJ, Bell EG, Jung ME. Examining the impact of a mobile health app on functional movement and physical fitness: pilot pragmatic randomized controlled trial. JMIR Mhealth Uhealth 2021 May 28;9(5):e24076 [FREE Full text] [doi: 10.2196/24076] [Medline: 34047704]
29. Qian J, Wu T, Lv M, Fang Z, Chen M, Zeng Z, et al. The value of mobile health in improving breastfeeding outcomes among perinatal or postpartum women: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2021 Jul 16;9(7):e26098 [FREE Full text] [doi: 10.2196/26098] [Medline: 34269681]
30. Kim M, Kim C, Kim E, Choi M. Effectiveness of mobile health-based exercise interventions for patients with peripheral artery disease: systematic review and meta-analysis. JMIR Mhealth Uhealth 2021 Feb 15;9(2):e24080 [FREE Full text] [doi: 10.2196/24080] [Medline: 33587042]
31. Robbins R, Krebs P, Jagannathan R, Jean-Louis G, Duncan DT. Health app use among US mobile phone users: analysis of trends by chronic disease status. JMIR Mhealth Uhealth 2017 Dec 19;5(12):e197 [FREE Full text] [doi: 10.2196/mhealth.7832] [Medline: 29258981]
32. Luxton DD, McCann RA, Bush NE, Mishkind MC, Reger GM. mHealth for mental health: integrating smartphone technology in behavioral healthcare. Professional Psychol Res Pract 2011 Dec;42(6):505-512. [doi: 10.1037/a0024485]
33. Ottaviani J. The post-embargo open access citation advantage: it exists (probably), its modest (usually), and the rich get richer (of course). PLoS One 2016;11(8):e0159614 [FREE Full text] [doi: 10.1371/journal.pone.0159614] [Medline: 27548723]
34. Eysenbach G. Citation advantage of open access articles. PLoS Biol 2006 May;4(5):e157 [FREE Full text] [doi: 10.1371/journal.pbio.0040157] [Medline: 16683865]
35. Iliescu R, Kumaravel A, Smurawska L, Torous J, Keshavan M. Smartphone ownership and use of mental health applications by psychiatric inpatients. Psychiatry Res 2021 May;299:113806. [doi: 10.1016/j.psychres.2021.113806] [Medline: 33667947]
36. Gmunder KN, Ruiz JW, Franceschi D, Suarez MM. Demographics associated with US healthcare disparities are exacerbated by the telemedicine surge during the COVID-19 pandemic. J Telemed Telecare 2021 Jun 23:1357633X211025939. [doi: 10.1177/1357633X211025939] [Medline: 34160328]
37. Moore SL, Fischer HH, Steele AW, Joshua Durfee M, Ginosar D, Rice-Peterson C, et al. A mobile health infrastructure to support underserved patients with chronic disease. Healthc (Amst) 2014 Mar;2(1):63-68. [doi: 10.1016/j.hjdsi.2013.12.016] [Medline: 26250090]
38. Yu SWY, Hill C, Ricks ML, Bennet J, Oriol NE. The scope and impact of mobile health clinics in the United States: a literature review. Int J Equity Health 2017 Oct 05;16(1):178 [FREE Full text] [doi: 10.1186/s12939-017-0671-2] [Medline: 28982362]
39. Buttorff C, Ruder T, Bauman M. Multiple Chronic Conditions in the United States. Santa Monica, CA: Rand; 2017:1-28.
40. Leroy L, Bayliss E, Domino M, Miller BF, Rust G, Gerteis J, AHRQ MCC Research Network. The Agency for Healthcare Research and Quality Multiple Chronic Conditions Research Network: overview of research contributions and future priorities. Med Care 2014 Mar;52 Suppl 3:S15-S22. [doi: 10.1097/MLR.0000000000000095] [Medline: 24561753]
41. Flegal KM, Kruszon-Moran D, Carroll MD, Fryar CD, Ogden CL. Trends in obesity among adults in the United States, 2005 to 2014. JAMA 2016 Jun 07;315(21):2284-2291. [doi: 10.1001/jama.2016.6458] [Medline: 27272580]
42. Gurka MJ, Filipp SL, DeBoer MD. Geographical variation in the prevalence of obesity, metabolic syndrome, and diabetes among US adults. Nutr Diabetes 2018 Mar 13;8(1):14. [doi: 10.1038/s41387-018-0024-2] [Medline: 29549249]
43. Keramat SA, Alam K, Al-Hanawi MK, Gow J, Biddle SJH, Hashmi R. Trends in the prevalence of adult overweight and obesity in Australia, and its association with geographic remoteness. Sci Rep 2021 May 31;11(1):11320. [doi: 10.1038/s41598-021-90750-1] [Medline: 34059752]
44. Bays HE, Toth PP, Kris-Etherton PM, Abate N, Aronne LJ, Brown WV, et al. Obesity, adiposity, and dyslipidemia: a consensus statement from the National Lipid Association. J Clin Lipidol 2013;7(4):304-383. [doi: 10.1016/j.jacl.2013.04.001] [Medline: 23890517]
45. Piché ME, Poirier P, Lemieux I, Després JP. Overview of epidemiology and contribution of obesity and body fat distribution to cardiovascular disease: an update. Prog Cardiovasc Dis 2018;61(2):103-113. [doi: 10.1016/j.pcad.2018.06.004] [Medline: 29964067]

XSL•FO
RenderX

46. De Pergola G, Silvestris F. Obesity as a major risk factor for cancer. J Obes 2013;2013:291546. [doi: 10.1155/2013/291546] [Medline: 24073332]
47. Apovian C. Obesity: definition, comorbidities, causes, and burden. Am J Manag Care 2016 Jun;22(7 Suppl):s176-s185 [FREE Full text] [Medline: 27356115]
48. Ghanbari M, Behzadifar M, Doshmangir L, Martini M, Bakhtiari A, Alikhani M, et al. Mapping research trends of universal health coverage from 1990 to 2019: bibliometric analysis. JMIR Public Health Surveill 2021 Jan 11;7(1):e24569 [FREE Full text] [doi: 10.2196/24569] [Medline: 33427687]

Abbreviations

mHealth: mobile health
WoS: Web of Science

Edited by A Mavragani; submitted 16.12.21; peer-reviewed by R Newton, L Guo; comments to author 10.01.22; revised version received 15.02.22; accepted 10.05.22; published 08.06.22.

Please cite as:
Wu CC, Huang CW, Wang YC, Islam M, Kung WM, Weng YC, Su CH
mHealth Research for Weight Loss, Physical Activity, and Sedentary Behavior: Bibliometric Analysis
J Med Internet Res 2022;24(6):e35747
URL: https://www.jmir.org/2022/6/e35747
doi:10.2196/35747
PMID:35675126

©Chieh-Chen Wu, Chih-Wei Huang, Yao-Chin Wang, Md.Mohaimenul Islam, Woon-Man Kung, Yung-Ching Weng, Chun-Hsien Su. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 08.06.2022. 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

Emerging Trends and Research Foci in Artificial Intelligence for Retinal Diseases: Bibliometric and Visualization Study

Junqiang Zhao[1,2*], MD; Yi Lu[2*], MSCN; Yong Qian[3], BSc; Yuxin Luo[4], BSc; Weihua Yang[4], MD

[1]Department of Medical Engineering, Xinxiang Medical University, Xinxiang, Henan, China

[2]Department of Nursing, Xinxiang Medical University, Xinxiang, Henan, China

[3]Jiangsu Testing and Inspection Institute for Medical Devices, Nanjing, Jiangsu, China

[4]The Laboratory of Artificial Intelligence and Bigdata in Ophthalmology, Affiliated Eye Hospital of Nanjing Medical University, Nanjing, Jiangsu, China

[*]these authors contributed equally

Corresponding Author:
Weihua Yang, MD
The Laboratory of Artificial Intelligence and Bigdata in Ophthalmology
Affiliated Eye Hospital of Nanjing Medical University
No.138 Hanzhong Road, Gulou District
Nanjing, Jiangsu, 210029
China
Phone: 86 13867252557
Fax: 86 2586677722
Email: benben0606@139.com

Abstract

Background: Patients with retinal diseases may exhibit serious complications that cause severe visual impairment owing to a lack of awareness of retinal diseases and limited medical resources. Understanding how artificial intelligence (AI) is used to make predictions and perform relevant analyses is a very active area of research on retinal diseases. In this study, the relevant Science Citation Index (SCI) literature on the AI of retinal diseases published from 2012 to 2021 was integrated and analyzed.

Objective: The aim of this study was to gain insights into the overall application of AI technology to the research of retinal diseases from set time and space dimensions.

Methods: Citation data downloaded from the Web of Science Core Collection database for AI in retinal disease publications from January 1, 2012, to December 31, 2021, were considered for this analysis. Information retrieval was analyzed using the online analysis platforms of literature metrology: Bibliometrc, CiteSpace V, and VOSviewer.

Results: A total of 197 institutions from 86 countries contributed to relevant publications; China had the largest number and researchers from University College London had the highest H-index. The reference clusters of SCI papers were clustered into 12 categories. “Deep learning” was the cluster with the widest range of cocited references. The burst keywords represented the research frontiers in 2018-2021, which were “eye disease” and “enhancement.”

Conclusions: This study provides a systematic analysis method on the literature regarding AI in retinal diseases. Bibliometric analysis enabled obtaining results that were objective and comprehensive. In the future, high-quality retinal image–forming AI technology with strong stability and clinical applicability will continue to be encouraged.

(J Med Internet Res 2022;24(6):e37532) doi:10.2196/37532

KEYWORDS

artificial intelligence; retinal disease; data visualization; bibliometric; citespace, VOSviewer; retinal; eye; visual impairment

Introduction

Retinal diseases are the main afflictions affecting human vision. Diabetic retinopathy (DR) is an eye vascular disease caused by diabetes [1]. Following DR, retinal vein occlusion is the most frequent retinal vascular disorder [2]. Drusen, long-spaced collagen, and phospholipid vesicles are all linked to age-related macular degeneration (AMD). These structures exist between the retinal pigment epithelium’s basement membrane and the rest of the Bruch membrane [3]. Glaucoma is a disease that

XSL•FO
RenderX

leads to the death of retinal ganglion cells as well as the loss of axons that make up the optic nerve [4]. Early detection of the disease is challenging; however, the condition may be improved with appropriate treatment [5]. These lesions are the major cause of vision loss or impairment in working-age and elderly adults worldwide [6,7]. The identification of retinopathy and maculopathy retinopathy may be time-intensive and requires expert training.

Artificial intelligence (AI), in which training data are used to develop a system, has become increasingly popular regarding clinical image analysis and disease diagnosis [8-13]. The US Food and Drug Administration has approved a device based on AI to diagnose DR, despite the fact that the application and development of AI in medicine are still in an infancy stage [14]. To address the current limitations of auxiliary examination processes, computer algorithms determine the optimal decision boundary in a multidimensional feature space [15]. At present, such systems are still being improved by researchers.

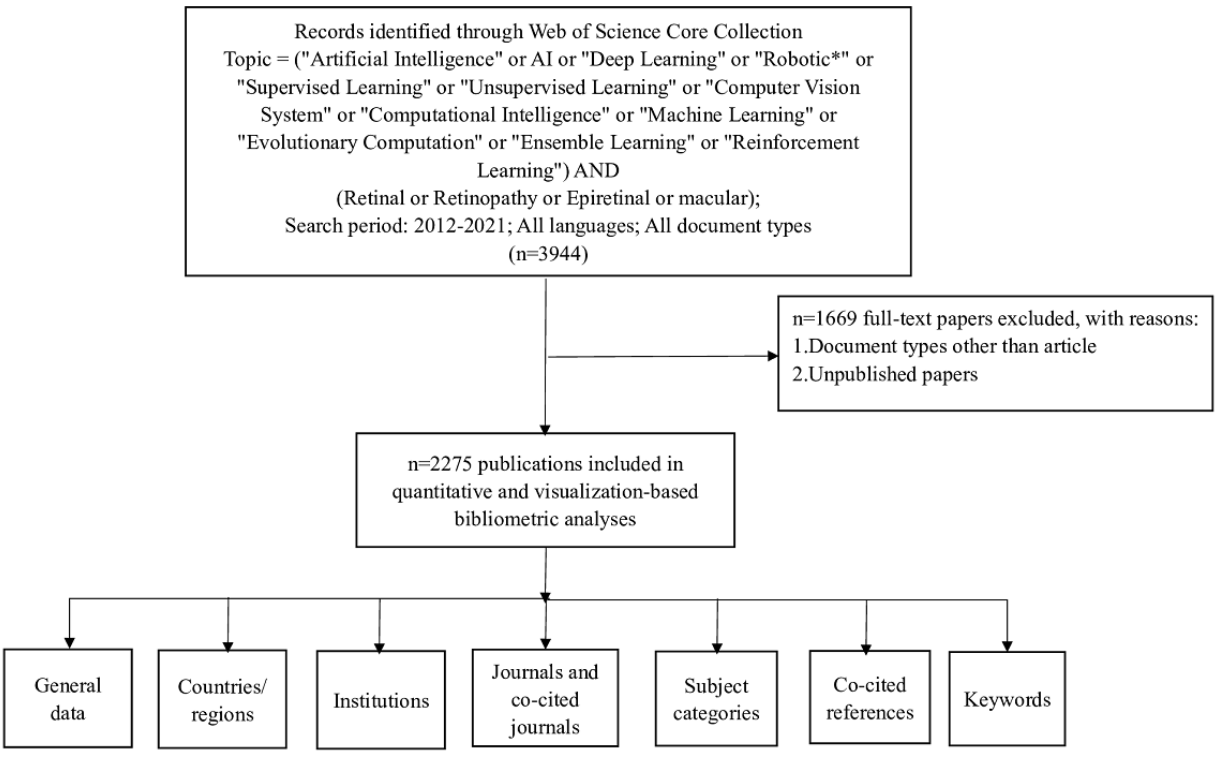
The aim of this study was to gain insights into the overall application of AI technology in the research of retinal diseases from specific time and space dimensions. We used bibliometric methods to analyze papers in the Science Citation Index (SCI) reporting studies performed from January 1, 2012, to December 31, 2021, on AI in retinal disease research. The citations of countries, regions, institutions, periodicals, study categories, keywords, and references were included in the data. In addition, we established a visual and unbiased approach to explore hotspot knowledge frontiers in a research area. This study thus provides a useful reference for algorithm researchers, ophthalmologists, and experts in the field of medical engineering.

Methods

Paper Selection

On February 15, 2022, all citation data published between January 1, 2012, and December 31, 2021, were retrieved from the Web of Science Core Collection (WoSCC). The data were independently verified by two authors (YL and JZ). The detailed search string is listed in Figure 1. The document type was article. From each publication, we gathered the following basic data: title, abstract, authors, institution, country or region, journal, keywords, and references. The detailed search and analysis processes are depicted in Figure 1.

Figure 1. Frame flow diagram for the detailed selection criteria and bibliometric analysis steps of applying artificial intelligence (AI) to the study of retinal diseases in the Web of Science Core Collection database.



Data Exclusion

Unpublished and document types other than articles were excluded. The citation data were downloaded on February 15, 2022, and some 2021 documents included by WoSCC were not published and were thus not included in this study. Some data were excluded because their document types were not articles, such as procedures, papers, review articles, meeting abstracts, early access, editorial materials, book chapters, letters, corrections, data papers, books, and retracted publications.

Statistical Analysis

Collaborative networks of countries, institutions, journals, keywords, and research categories were analyzed and visualized using the bibliometrics online analysis platform Bibliometrc [16], CiteSpace V, and VOSviewer. We collected detailed

XSL•FO
RenderX

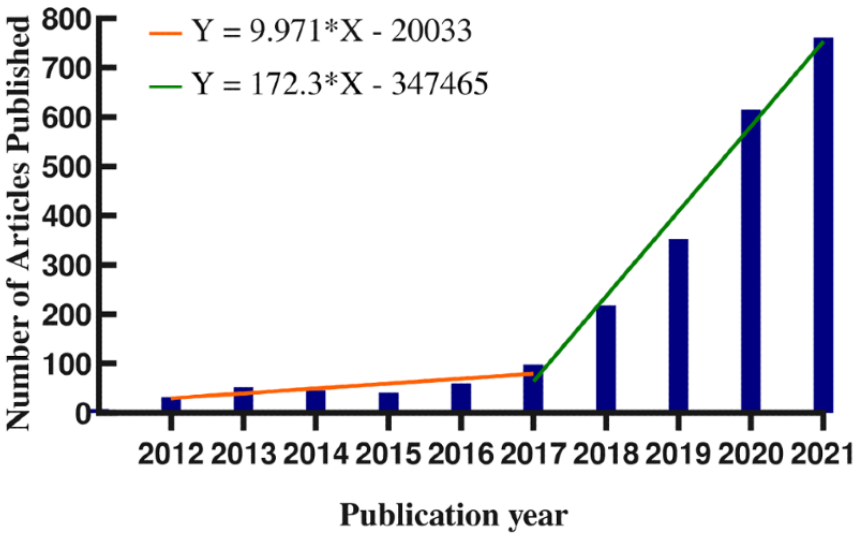
citation features for analysis, including the number of annual publications, countries, institutes, journals, subject categories, cocited references, and keywords. The H-index represents an estimate of the importance and general impact of the research contributions [17].

Results

Distribution of Articles by Publication Year

We analyzed a total of 2275 papers published between 2012 and 2021. The numbers of published studies on the application of AI technology to retinal illnesses over time are summarized in Figure 2. Since 2017, the number of studies on the use of AI in the treatment of retinal illnesses has skyrocketed.

Figure 2. Trends in the number of publications on applying artificial intelligence to the study of retinal diseases from 2012 to 2021.



Countries or Regions and Institutes

These citations mentioned a total of 86 nations or territories. In Figure 3, the publications from different countries or regions are represented by blocks of different colors. The size of the colored block area represents the number of citations and the size of the different colored coverage areas represents the intensity of the cooperation. In Figure 4, a larger label area for a given country represents a greater contribution to the related literature. The purple node area indicates the strength of the centrality; the higher the centrality value, the more cooperative the relations it establishes in the country where the node is located. As per Figures 3 and 4, China and the United States have contributed the largest number of documents to this field. The United States, the United Kingdom, and Singapore cooperated more with other countries.

Table 1 lists the top 10 countries cited. China had the largest number of publications, followed by the United States, India, and the United Kingdom. Britain had the strongest centrality, followed by the United States and Singapore.

A total of 197 institutions have published relevant papers, and the clustering of their cooperative relationships is shown in Figure 5. The top 10 institutions regarding the frequency of cited institutions are listed in Table 1, including three Chinese institutions (Sun Yat Sen University, Chinese Academy of Sciences, and Shanghai Jiao Tong University), three US institutions (Johns Hopkins University, Oregon Health and Science University, and Stanford University), two Singapore institutions (Singapore National Eye Centre and National University Singapore), one Austrian institution (Medical University of Vienna), and one UK institution (University College London). Among them, the number of citations with authors from University College London ranked in 10th position; however, their H-index was the highest. In addition, Johns Hopkins University and University College London, which were the two highest-ranked institutions in the center, appeared in the same cluster.

XSL•FO
RenderX

Figure 3. The cooperation of countries or regions that contributed to publications on applying artificial intelligence to the study of retinal diseases from 2012 to 2021.

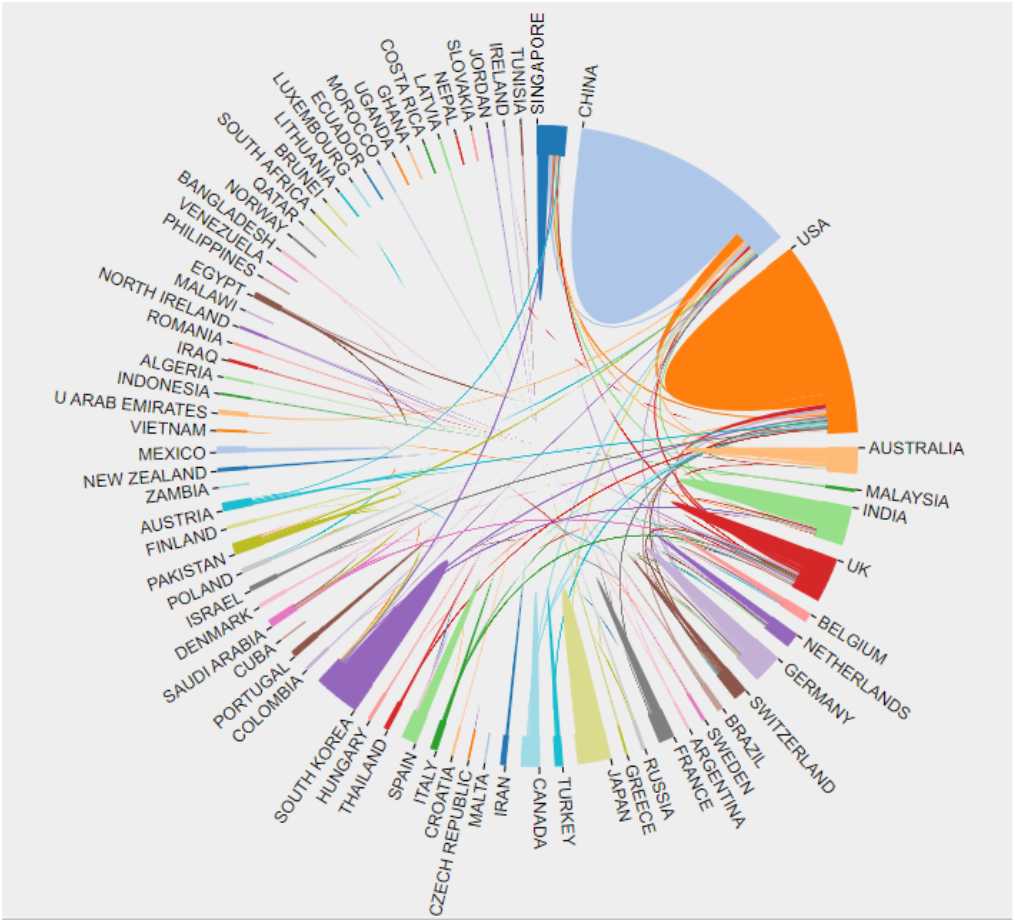
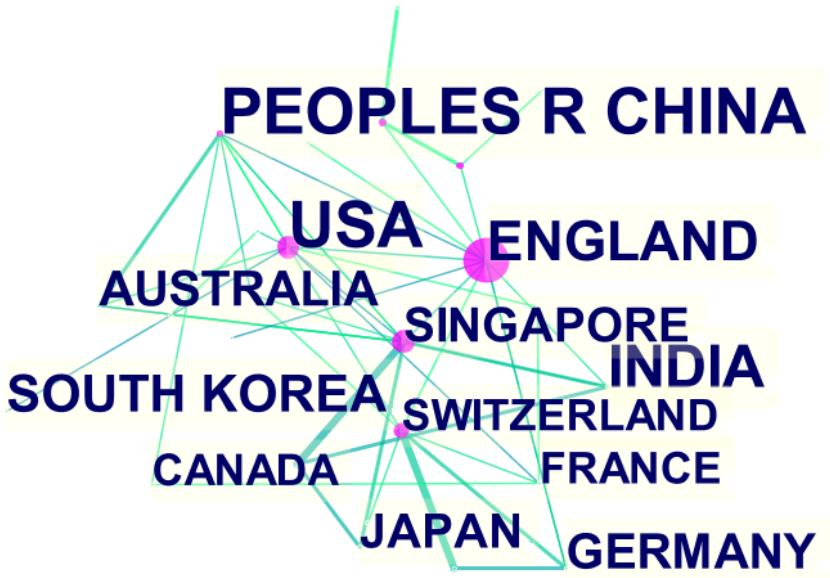


Figure 4. The cooperation of countries or regions that contributed to publications on applying artificial intelligence to the study of retinal diseases from 2012 to 2021.

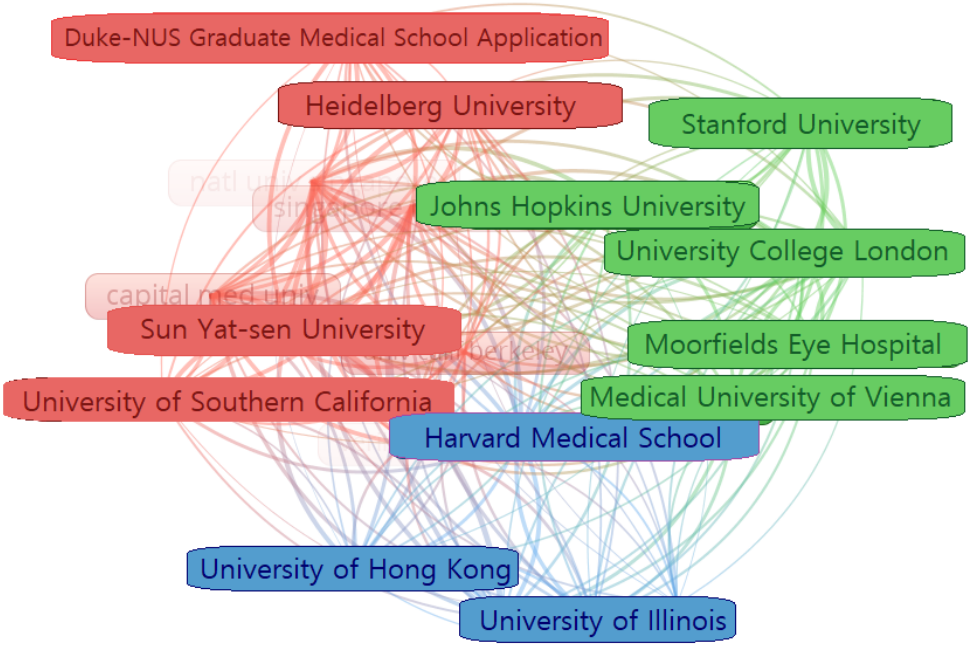


XSL•FO
RenderX

Table 1. The top 10 countries or regions and institutions with publications on the application of artificial intelligence in retinal diseases from 2012 to 2021.

Rank	Count	Centrality	H-index
Country or region			
1. People's Republic of China	634	0.16	42
2. United States	620	0.57	56
3. India	309	0.04	33
4. England	205	1.00	33
5. South Korea	150	0.00	24
6. Germany	132	0.04	24
7. Australia	120	0.00	30
8. Japan	98	0.00	19
9. Singapore	98	0.49	20
10. Canada	74	0.06	19
Institution			
1. Sun Yat Sen University	62	0.05	16
2. Chinese Academy of Science	51	008	14
3. Johns Hopkins University	49	0.13	16
4. Oregon Health and Science University	48	0.01	16
5. Stanford University	47	0.03	18
6. Medical University of Vienna	42	0.06	19
7. Singapore National Eye Centre	39	0.11	18
8. National University of Singapore	38	0.07	20
9. Shanghai Jiao Tong University	38	0.00	10
10. University College London	37	0.12	21

Figure 5. Network map of institutions that contributed to publications on the application of artificial intelligence in retinal diseases from 2012 to 2021.



XSL•FO
RenderX

Journals

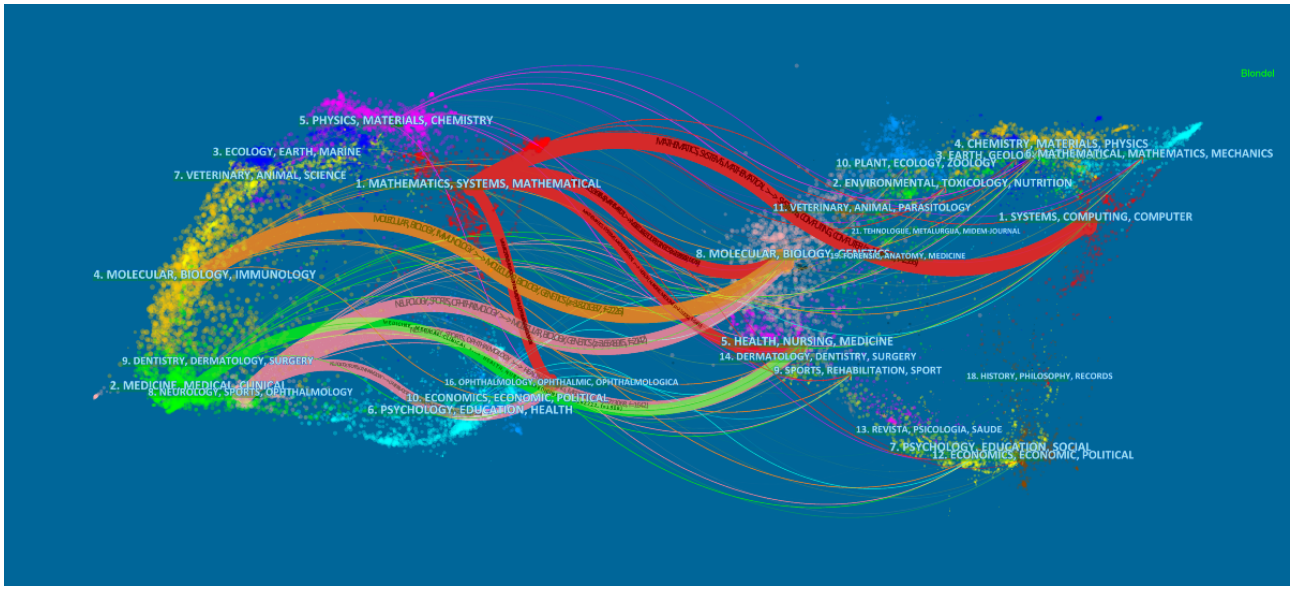
Citing journals represents the research frontier and cited journals represents the research foundation. The top 10 citing journals and cited journals are shown in Table 2. The most frequently cited journals in the included citations were *Translational Vision Science Technology*, *Scientific Reports*, and *IEEE Access*. The journals that appeared most frequently among the cited journals were *Ophthalmology*, *British Journal of Ophthalmology*, and *IEEE Transactions on Medical Imaging*. *PLoS One* appeared in both the top-ranked citing and cited journals lists. The dual map of the journals is shown in Figure 6. Red represents the discipline field with the greatest influence. The research influence in the field of mathematics/systems/mathematical subject ranked first among the citing journals.

Table 2. The top 10 citing journals and cited journals of publications on the application of artificial intelligence in retinal diseases from 2012 to 2021.

Rank	Count
Citing journals	
1. Translational Vision Science Technology	100
2. Scientific Reports	86
3. IEEE Access	85
4. Biomedical Optics Express	67
5. PLoS One	53
6. IEEE Transactions on Medical Imaging	48
7. American Journal of Ophthalmology	41
8. Computer Methods and Programs in Biomedicine	40
9. British Journal of Ophthalmology	36
10. Eye	31
Cited journals	
1. Ophthalmology	1140
2. Investigative Ophthalmology & Visual Science	1083
3. IEEE Transactions on Medical Imaging	974
4. Lecture Notes in Computer Science	855
5. British Journal of Ophthalmology	778
6. PLoS One	775
7. Medical Image Analysis	714
8. JAMA (Journal of the American Medical Association)	681
9. American Journal of Ophthalmology	673
10. JAMA Ophthalmology	647

XSL•FO
RenderX

Figure 6. The dual-map overlay of journals that contributed to publications on the application of artificial intelligence in retinal diseases from 2012 to 2021. Red represents the greatest influence.



Research Category

Figure 7 and Table 3 present the research areas of the citations. The most involved research areas were Ophthalmology and Engineering Electrical Electronic. The highest H-index score areas were Engineering Biomedical and Radiology Nuclear Medicine Medical Imaging. This indicates that research on AI in retinal diseases is primarily focused within the fields of computer engineering and medical imaging.

Figure 7. Network map of the research categories of publications on the application of artificial intelligence in retinal diseases from 2012 to 2021.

XSL•FO
RenderX

Table 3. The top 10 research categories of publications on the application of artificial intelligence in retinal diseases from 2012 to 2021.

Rank	Research category	Count	H-index
1	Ophthalmology	489	40
2	Engineering Electrical Electronic	332	33
3	Engineering Biomedical	318	42
4	Computer Science Artificial Intelligence	265	34
5	Computer Science Interdisciplinary Applications	250	40
6	Radiology Nuclear Medicine Medical Imaging	246	42
7	Computer Science Information Systems	206	23
8	Multidisciplinary Sciences	185	25
9	Medical Informatics	173	30
10	Mathematical Computational Biology	139	24

Keywords

Keywords were retrieved and examined from the relevant literature. Table 4 lists the top 20 keywords used. Among them, the keywords cited over 200 times were “diabetic retinopathy,” “classification,” “validation,” and “imaging.” The keywords of the 2267 articles were analyzed and divided into four categories (deep learning, DR, optical coherence tomography, and classification), as shown in Figure 8. The time trend was examined using the hotspot transfer method, which was applied to the first 15 keywords with the highest citation outbreak. As shown in Figure 9, the key words with the greatest outburst intensity were “pattern” and “retinal ganglion cell.” The red grid indicates the emergence of keywords. “Information” and “neuron” were the keywords with the longest use (2012-2019), and “eye disease” and “enhancement” were the most popular keywords from 2018 to 2021.

Table 4. The top 20 keywords on the application of artificial intelligence in retinal diseases from 2012 to 2021.

Rank	Keyword	Count
1	Diabetic retinopathy	380
2	Classification	271
3	Image	270
4	Validation	224
5	Segmentation	213
6	Optical coherence tomography	154
7	Diagnosis	152
8	System	133
9	Prevalence	124
10	Macular degeneration	118
11	Algorithm	111
12	Retinal image	110
13	Disease	106
14	Model	105
15	Neural network	104
16	Blood vessel	103
17	Retinopathy	101
18	Eye	99
19	Progression	82
20	Automated detection	77

XSL•FO
RenderX

Figure 8. Network map of the 50 top-ranking keywords divided into four clusters.

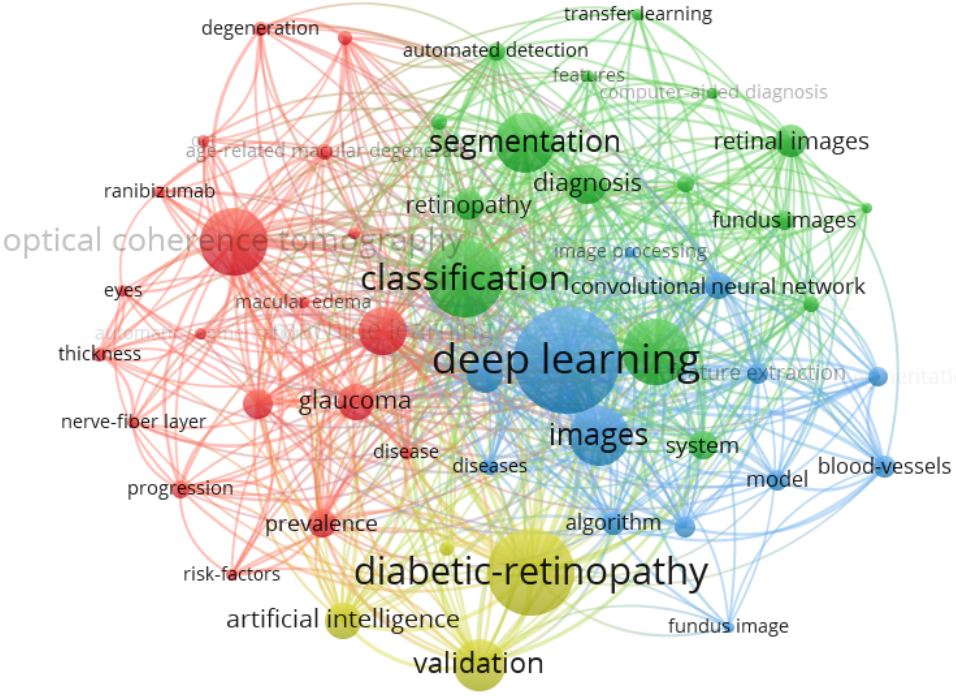


Figure 9. Top 15 keywords with the strongest citation bursts of publications on the application of artificial intelligence in retinal diseases from 2012 to 2021. Red indicates the emergence of keywords.

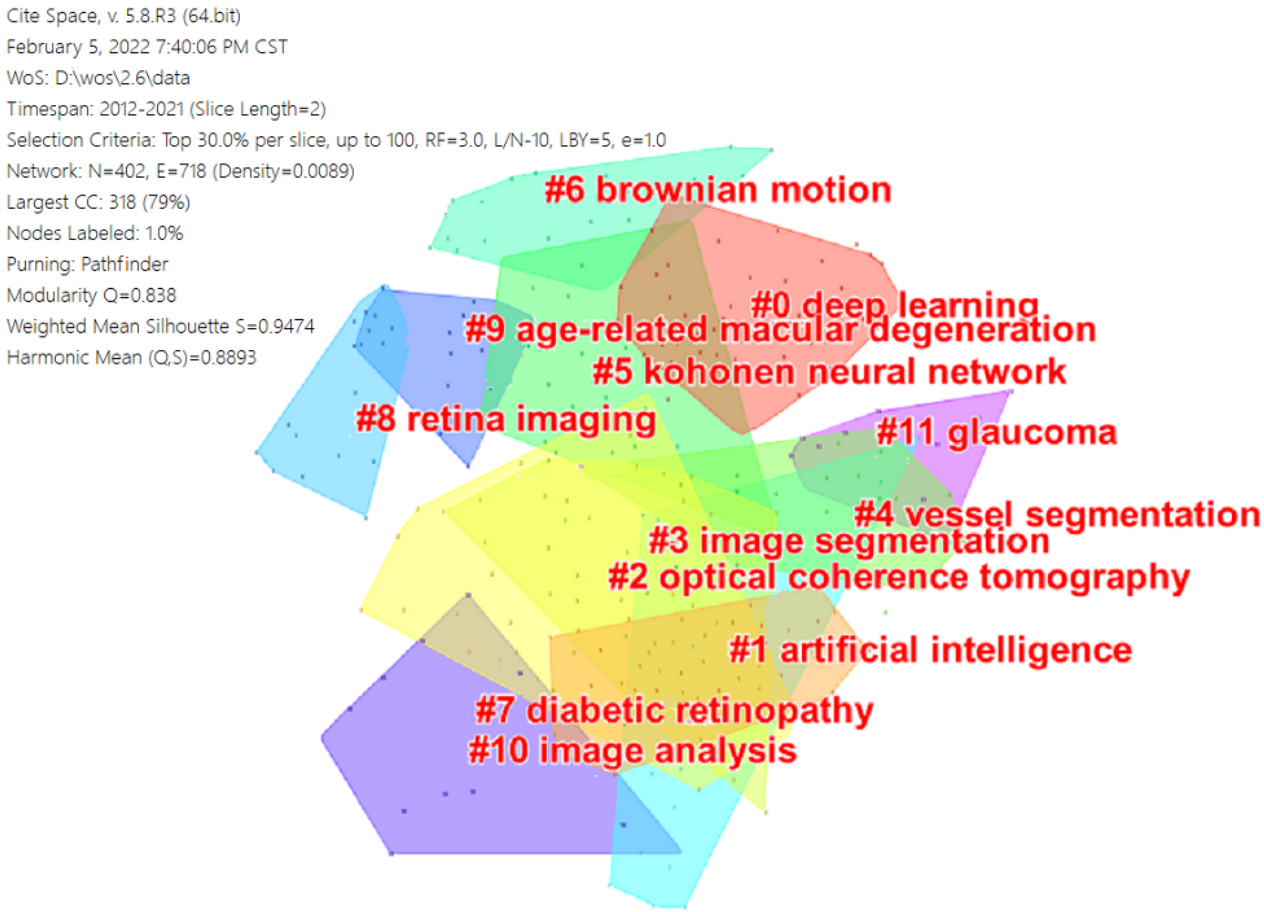
Keywords	Year	Strength	Begin	End	2012 - 2021
pattern	2012	4.24	**2012**	2015	
retinal ganglion cell	2012	4.22	**2012**	2013	
information	2012	3.98	**2012**	2019	
machine learning classifier	2012	3.7	**2012**	2015	
neuron	2012	3.66	**2012**	2019	
cortex	2012	3.98	**2014**	2019	
drusen	2012	3.83	**2014**	2019	
tracking	2012	3.13	**2014**	2019	
optimization	2012	3.16	**2016**	2019	
segmentation	2012	3.95	**2018**	2019	
retinal image	2012	3.71	**2018**	2019	
retinal layer	2012	3.68	**2018**	2019	
mechanism	2012	3.09	**2018**	2019	
eye disease	2012	3.45	**2020**	2021	
enhancement	2012	3.13	**2020**	2021	

Reference Network

A reference's citation frequency represents its influence. We constructed a cocited literature network and analyzed the scientific relevance of the publications. The modularity of the network was measured using the modularity index; the higher the modularity Q score, the better the network's clustering. Q>0.3 indicates that the network community structure obtained is noteworthy. The better the network's homogeneity, the closer the Silhouette S value is to 1. The cited documents were grouped into 12 clusters, as shown in Figure 10. The homogeneity of these clusters was characterized by modularity Q=0.838 and weighted mean Silhouette S=0.9474. Cluster markers were created using the index items taken from the literature. The modularity of the network was measured using the modularity index. "Deep learning" was the cluster with the widest range.

XSL•FO
RenderX

Figure 10. Reference cocitation map of publications on the application of artificial intelligence in retinal diseases from 2012 to 2021.



Overall Results

There were 2275 SCI papers reviewed in this study of AI and retinal diseases, which were published from 2012 to 2021, with a major increase in these publications since 2017. This is due to one study [18] being cited 506 times in 2016, which provided an important scientific and technical reference for researchers in this field. China and the United States had the maximum number of publications. The United Kingdom had the most cooperative relations. In addition, the three outcome indicators of the institution indicated that the research conducted by University College London and Johns Hopkins University has had a major influence on this research field. The number of citations in this field published by citing journals and cited journals indicated that the application of AI in retinal diseases has mainly been in the fields of computer engineering and medical imaging based on digital technology. This same pattern also applied to the citation categories.

The important nodes in the clustering demonstrated this research field's knowledge bases after clustering the cited references. They were labeled #0 deep learning, #1 artificial intelligence, #2 optical coherence tomography, #3 image segmentation, #4 vessel segmentation, #5 Kohonen neural network, #6 Brownian motion, #7 diabetic retinopathy, #8 retina imaging, #9 age-related macular degeneration, #10 image analysis, and #11 glaucoma. The top 10 cocited references in these clusters are listed in Table 5. These studies' conclusions require a substantial amount of basic work to be performed by ophthalmologists. For example, from the top-ranking article, 54 ophthalmologists or ophthalmic trainees participated in the identification of 5-year image data of three eye hospital patients presenting for image-based DR screening. From the second-ranked article, 494,661 retinal photographs were used to assess the diagnosis performance of a deep-learning system for DR and related eye illnesses. Each retinal image was analyzed by two trained senior, certified nonmedical professional graders. To achieve progress in the application of AI in retinal diseases, more cooperative and collaborative relationships among ophthalmologists, imaging technicians, and computer technology researchers are required in the future.

XSL•FO
RenderX

Table 5. The top 10 publications on the application of artificial intelligence (AI) in retinal diseases from 2012 to 2021.

Rank	Reference	Title of cocited reference	Count	Interpretation of the findings
1	Gulshan et al [18]	Development and validation of a deep learning algorithm for the detection of DR[a] in retinal fundus photographs	506	For diagnosing referable DR, a deep machine learning method had good sensitivity and specificity
2	Ting et al [19]	Development and validation of a deep learning system for DR and related eye diseases using retinal images from multiethnic populations with diabetes	270	The deep learning system demonstrated high sensitivity and specificity for detecting DR and related eye disorders
3	Lam et al [20]	Automated Identification of DR using deep learning	214	A high-reliability data-driven AI-based grading technique for screening and identifying fundus pictures taken from patients with diabetes. For further assessment and therapy, these patients should be referred to an ophthalmologist
4	Ronneberger et al [21]	U-Net: convolutional networks for biomedical image segmentation	208	This research provides a network and training technique that heavily depends on data augmentation to make better use of existing annotated samples
5	He et al [22]	Deep residual learning for image recognition	173	This research proposes a residual learning paradigm for network training
6	Kermany et al [23]	Identifying medical diagnoses and treatable diseases by image-based deep learning	163	This paper describes the development of a diagnostic tool for screening patients with common treatable blinding retinal disorders based on a deep-learning architecture
7	LeCun et al [24]	Deep learning	162	AI will advance as a result of systems that combine representation learning and complicated reasoning
8	De Fauw et al [25]	Clinically applicable deep learning for diagnosis and referral in retinal disease	160	When using tissue segmentations from a different type of device, a unique deep learning architecture was used to a clinically diverse data set to retain referral accuracy
9	Abràmoff et al [26]	Improved automated detection of DR on a publicly available dataset through the integration of deep learning	159	Deep learning–enhanced algorithms have the potential to improve the effectiveness of DR screening, thereby preventing vision loss and blindness from this dreadful disease
10	Esteva et al [27]	Dermatologist-level classification of skin cancer with deep neural networks	145	This research shows how deep learning works in dermatology and how it can be applied to other fields, including ophthalmology, otolaryngology, radiography, and pathology

[a]DR: diabetic retinopathy.

Discussion

Research Hotspots and Frontiers

Overview

Keywords provide a quick summary of the most important aspects and points in a collection of articles [28]. Current research hotspots and frontiers can be identified using burst keyword analysis. Following the capture of the burst keywords, two study fields were identified: eye disease (2018-2021) and enhancement (2018-2021).

Eye Disease

The application of AI technology to eye diseases is comparable to the best clinical systems and has achieved competitive results in solving issues related to the diagnosis and monitoring of complex ophthalmic diseases. AMD is usually asymptomatic, and an intermediate stage may not be identified. Moreover, AMD affects several people worldwide and thus identifying it can be time-consuming and difficult without the assistance of experts. Fortunately, applying deep learning–based automated algorithms may solve this challenge. This could also address the expenses of screening or monitoring, health care access, and the evaluation of innovative treatments for AMD development or progression [29]. DR also causes challenges for many people. It is the leading cause of vision loss and preventable blindness in adults aged 20-74 years in middle- and high-income countries [30]. Using a combination of digital retinal image analysis and telemedicine assessment to help identify people at risk of cardiovascular disease and cognitive impairment may have benefits beyond sight-threatening diseases prevention [11,31]. Aamir et al [32] built a hierarchical deep convolutional neural

XSL•FO
RenderX

network (CNN) for glaucoma recognition and prevention using an advanced deep-learning technique. This project was then used to extract multilayer features from 1338 images to verify the performance of the algorithm, achieving nearly 100% specificity, sensitivity, accuracy, and precision. These studies highlight the current research findings based on the use of AI technology in clinical applications for the management of ophthalmic diseases.

In addition, AI technology has been applied to the screening, referral, diagnosis, health care, and follow-up visits of patients with a variety retinal illness. Wang et al [33] developed a two-step semiautomatic deep learning algorithm–assisted technique to identify fundus pictures and aid in the detection of DR with vision-threatening complications. Optical coherence tomographic (OCT) angiography is a noninvasive imaging technology that may generate angiograms at precise depths within the retina as well as visualize the microvasculature in real time [34]. A British study published in 2021 assessed an AI decision support system with the use of OCT scanning of retinal pictures to identify the digital referral path, providing evidence of the contributing reasons and difficulties of adopting the digital path in real life, with the goal of helping to eliminate unnecessary referrals [35]. To boost doctors' faith in AI systems to make accurate diagnoses, some AI systems must be written as interpretable programs [36]. According to a follow-up poll conducted in 2017, radiologists used certain touch-environment solutions forced by the clinical setting at the time, demonstrating that they are still opposed to the transfer from traditional to updated interfaces [37].

Enhancement

Ophthalmologists are confused by the quality differences among fundus diagnostic images [38]. Enhancing the analysis of retinal image structures requires the development of a computer-assisted algorithm to correct the low fundus image quality [39]. Wan et al [40] proposed a deep learning–based technique that overcomes the limitations of current imaging algorithms and improves the low retinal image quality. CNN models may be developed under strong and mostly correct assumptions regarding the nature of macular disease images [41]. El-Hag et al [42]established the importance of the proposed blurry image improvement phase. Additionally, using CNN as a classification technique with hazy logic augmentation was shown to improve the classification of normal and abnormal outcomes. In the testing phase, this resulted in a classification accuracy of 100% [42]. The blood vessels in the neural network must be divided into arteries and veins to diagnose hypertensive retinopathy using retinal diagnostic images. According to this demand, Hussein and Faheem [43] proposed the use of an AI method to improve vascular contrast. Zhou et al [44] proposed the learning of discriminative CNN features and enhanced thin vessels in color fundus images to further improve the segmentation performance. This algorithm improves the contrast of the retinal vessels and was verified by three pediatric ophthalmologists [45]. Goel et al [46] showed that using a development learning model to transfer learning can improve the accuracy of correct classification of different aneurysms in the retina area caused by DR.

AI technology can improve diagnosis accuracy and can also save time for both doctors and patients by increasing the contrast between image structures, such as segmenting distinct blood arteries or calculating normal and pathological structures. The image quality of retinal disease examination needs to be unified with high precision. This is currently a research hotspot to provide more high-quality research images based on research and development in AI technology, improvements in image acquisition technology, and the standardization of acquisition steps.

Limitations

This bibliometric analysis only included the literature data in WoSCC. Some other databases were not included, such as PubMed, Medline, and Cochrane. In addition, the citation data analyzed were only from the literature published from 2012 to 2021, rather than collecting all articles published in this research field to date. Some 2022 studies are still ongoing and have not yet been published. These criteria may result in publication bias.

Conclusions

This study provides a systematic literature analysis on the use of AI in retinal diseases. Bibliometric analysis enabled obtaining objective and comprehensive results. Judging by the volume of published papers and research subjects, this study area is still popular and a noteworthy topic with major interdisciplinary exploration space. Ophthalmologists, imaging experts, and computer algorithm researchers in developing and developed countries need to make full use of population advantages or core technologies in different regions to strengthen collaboration. This idea has become a research hotspot that uses the existing basic clinical research results and a more advanced algorithm mode to develop a high-quality ophthalmic examination image system and further verify its clinical applicability. At present, an algorithm program with 100% diagnostic accuracy for retinal disease has been developed [42]. In the future, high-quality retinal image–forming AI technology with strong stability and clinical applicability will continue to be encouraged.

Acknowledgments

This research was funded by the Nanjing Enterprise Expert Team Project, Medical Science and Technology Development Project Fund of Nanjing (YKK21262), and Medical Big Data Clinical Research Project of Nanjing Medical University.

XSL•FO
RenderX

Authors' Contributions

JZ, YL, and YQ acquired, analyzed, and discussed the data, and drafted the manuscript. YL drafted the manuscript. WY designed the research, acquired the clinical information, and revised the manuscript. All authors have contributed to the manuscript and approved the submitted version.

Conflicts of Interest

None declared.

References

1. Zhang P, Xue W, Huang X, Xu Y, Lu L, Zheng K, et al. Prevalence and risk factors of diabetic retinopathy in patients with type 2 diabetes in Shanghai. Int J Ophthalmol 2021;14(7):1066-1072 [FREE Full text] [doi: 10.18240/ijo.2021.07.16] [Medline: 34282393]
2. Song P, Xu Y, Zha M, Zhang Y, Rudan I. Global epidemiology of retinal vein occlusion: a systematic review and meta-analysis of prevalence, incidence, and risk factors. J Glob Health 2019 Jun;9(1):010427. [doi: 10.7189/jogh.09.010427] [Medline: 31131101]
3. Age-Related Eye Disease Study Research Group. The Age-Related Eye Disease Study system for classifying age-related macular degeneration from stereoscopic color fundus photographs: the Age-Related Eye Disease Study Report Number 6. Am J Ophthalmol 2001 Nov;132(5):668-681. [doi: 10.1016/s0002-9394(01)01218-1] [Medline: 11704028]
4. Cesareo M, Ciuffoletti E, Ricci F, Missiroli F, Giuliano MA, Mancino R, et al. Visual disability and quality of life in glaucoma patients. Prog Brain Res 2015;221:359-374. [doi: 10.1016/bs.pbr.2015.07.003] [Medline: 26518087]
5. Ireka OJ, Ogbonnaya CE, Arinze OC, Ogbu N, Chuka-Okosa CM. Comparing posture induced intraocular pressure variations in normal subjects and glaucoma patients. Int J Ophthalmol 2021;14(3):399-404 [FREE Full text] [doi: 10.18240/ijo.2021.03.11] [Medline: 33747816]
6. King H, Aubert RE, Herman WH. Global burden of diabetes, 1995-2025: prevalence, numerical estimates, and projections. Diabetes Care 1998 Sep;21(9):1414-1431. [doi: 10.2337/diacare.21.9.1414] [Medline: 9727886]
7. Nucci C, Russo R, Martucci A, Giannini C, Garaci F, Floris R, et al. New strategies for neuroprotection in glaucoma, a disease that affects the central nervous system. Eur J Pharmacol 2016 Sep 15;787:119-126. [doi: 10.1016/j.ejphar.2016.04.030] [Medline: 27089818]
8. Chen Q, Yu W, Lin S, Liu B, Wang Y, Wei Q, et al. Artificial intelligence can assist with diagnosing retinal vein occlusion. Int J Ophthalmol 2021;14(12):1895-1902 [FREE Full text] [doi: 10.18240/ijo.2021.12.13] [Medline: 34926205]
9. Wan C, Li H, Cao G, Jiang Q, Yang W. An artificial intelligent risk classification method of high myopia based on fundus images. J Clin Med 2021 Sep 29;10(19):4488 [FREE Full text] [doi: 10.3390/jcm10194488] [Medline: 34640509]
10. Xu J, Shen J, Jiang Q, Wan C, Yan Z, Yang W. Research on the segmentation of biomarker for chronic central serous chorioretinopathy based on multimodal fundus image. Dis Markers 2021;2021:1040675. [doi: 10.1155/2021/1040675] [Medline: 34527086]
11. Wan C, Chen Y, Li H, Zheng B, Chen N, Yang W, et al. EAD-Net: a novel lesion segmentation method in diabetic retinopathy using neural networks. Dis Markers 2021;2021:6482665. [doi: 10.1155/2021/6482665] [Medline: 34512815]
12. Calisto FM, Santiago C, Nunes N, Nascimento JC. Introduction of human-centric AI assistant to aid radiologists for multimodal breast image classification. Int J Hum Comput Stud 2021 Jun;150:102607. [doi: 10.1016/j.ijhcs.2021.102607]
13. Calisto F, Nunes N, Nascimento J. BreastScreening. 2020 Sep Presented at: International Conference on Advanced Visual Interfaces; September 20-October 2, 2020; Salerno, Italy p. 1-5. [doi: 10.1145/3399715.3399744]
14. FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. US Food and Drug Administration. URL: https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye [accessed 2018-04-11]
15. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017 Dec;42:60-88. [doi: 10.1016/j.media.2017.07.005] [Medline: 28778026]
16. Bibliomtrc. URL: https://bibliometric.com/ [accessed 2022-04-25]
17. Hirsch JE. An index to quantify an individual's scientific research output. Proc Natl Acad Sci U S A 2005 Nov 15;102(46):16569-16572 [FREE Full text] [doi: 10.1073/pnas.0507655102] [Medline: 16275915]
18. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 2016 Dec 13;316(22):2402-2410. [doi: 10.1001/jama.2016.17216] [Medline: 27898976]
19. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA 2017 Dec 12;318(22):2211-2223 [FREE Full text] [doi: 10.1001/jama.2017.18152] [Medline: 29234807]
20. Lam C, Yi D, Guo M, Lindsey T. Automated detection of diabetic retinopathy using deep learning. AMIA Jt Summits Transl Sci Proc 2018;2017:147-155 [FREE Full text] [Medline: 29888061]

XSL•FO
RenderX

21. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Lecture Notes in Computer Science. 2015 Jan 01 Presented at: 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI); October 5-9, 2015; Munich, Germany p. 234-241 URL: https://arxiv.org/pdf/1505.04597.pdf
22. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 Jan 01 Presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 27-30, 2016; Seattle, WA p. 770-778 URL: https://openaccess.thecvf.com/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf
23. Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 2018 Feb 22;172(5):1122-1131 [FREE Full text] [doi: 10.1016/j.cell.2018.02.010] [Medline: 29474911]
24. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]
25. De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med 2018 Sep;24(9):1342-1350. [doi: 10.1038/s41591-018-0107-6] [Medline: 30104768]
26. Abràmoff MD, Lou Y, Erginay A, Clarida W, Amelon R, Folk JC, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Invest Ophthalmol Vis Sci 2016 Oct 01;57(13):5200-5206. [doi: 10.1167/iovs.16-19964] [Medline: 27701631]
27. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Feb 02;542(7639):115-118 [FREE Full text] [doi: 10.1038/nature21056] [Medline: 28117445]
28. Li H, An H, Wang Y, Huang J, Gao X. Evolutionary features of academic articles co-keyword network and keywords co-occurrence network: Based on two-mode affiliation network. Phys A Stat Mech Appl 2016 May;450:657-669. [doi: 10.1016/j.physa.2016.01.017]
29. Shen J, Wang L, Wang LL, Lyu CF, Liu S, Xie GT, et al. Image enhancement of color fundus photographs for age-related macular degeneration: the Shanghai Changfeng Study. Int J Ophthalmol 2022 Feb 18;15(2):268-275 [FREE Full text] [doi: 10.18240/ijo.2022.02.12] [Medline: 35186687]
30. Cheung N, Tikellis G, Wang JJ. Diabetic retinopathy. Ophthalmology 2007 Nov;114(11):2098-9; author reply 2099. [doi: 10.1016/j.ophtha.2007.07.010] [Medline: 17980748]
31. Vujosevic S, Aldington SJ, Silva P, Hernández C, Scanlon P, Peto T, et al. Screening for diabetic retinopathy: new perspectives and challenges. Lancet Diabetes Endocrinol 2020 Apr;8(4):337-347. [doi: 10.1016/S2213-8587(19)30411-5] [Medline: 32113513]
32. Aamir M, Irfan M, Ali T, Ali G, Shaf A, Al-Beshri A, et al. An adoptive threshold-based multi-level deep convolutional neural network for glaucoma eye disease detection and classification. Diagnostics 2020 Aug 18;10(8):602 [FREE Full text] [doi: 10.3390/diagnostics10080602] [Medline: 32824682]
33. Wang Y, Shi D, Tan Z, Niu Y, Jiang Y, Xiong R, et al. Screening referable diabetic retinopathy using a semi-automated deep learning algorithm assisted approach. Front Med 2021 Nov 25;8:740987. [doi: 10.3389/fmed.2021.740987] [Medline: 34901058]
34. Jia Y, Tan O, Tokayer J, Potsaid B, Wang Y, Liu JJ, et al. Split-spectrum amplitude-decorrelation angiography with optical coherence tomography. Opt Express 2012 Feb 09;20(4):4710. [doi: 10.1364/oe.20.004710]
35. Han JED, Liu X, Bunce C, Douiri A, Vale L, Blandford A, et al. Teleophthalmology-enabled and artificial intelligence-ready referral pathway for community optometry referrals of retinal disease (HERMES): a Cluster Randomised Superiority Trial with a linked Diagnostic Accuracy Study-HERMES study report 1-study protocol. BMJ Open 2022 Feb 01;12(2):e055845 [FREE Full text] [doi: 10.1136/bmjopen-2021-055845] [Medline: 35105593]
36. Ahuja AS, Halperin LS. Understanding the advent of artificial intelligence in ophthalmology. J Curr Ophthalmol 2019 Jun;31(2):115-117 [FREE Full text] [doi: 10.1016/j.joco.2019.05.001] [Medline: 31317087]
37. Calisto F, Ferreira A, Nascimento J, Gonçalves D. Towards touch-based medical image diagnosis annotation. 2017 Jan 01 Presented at: 12th Association of Computing Machinery International Conference on Interactive Surfaces and Spaces (ACM ISS); October 17-20, 2017; Brighton, England p. 390-395.
38. Philip S, Cowie LM, Olson JA. The impact of the Health Technology Board for Scotland's grading model on referrals to ophthalmology services. Br J Ophthalmol 2005 Jul;89(7):891-896 [FREE Full text] [doi: 10.1136/bjo.2004.051334] [Medline: 15965173]
39. Wan C, Wu J, Li H, Yan Z, Wang C, Jiang Q, et al. Optimized-Unet: novel algorithm for parapapillary atrophy segmentation. Front Neurosci 2021;15:758887. [doi: 10.3389/fnins.2021.758887] [Medline: 34720868]
40. Wan C, Zhou X, You Q, Sun J, Shen J, Zhu S, et al. Retinal image enhancement using cycle-constraint adversarial network. Front Med 2021;8:793726. [doi: 10.3389/fmed.2021.793726] [Medline: 35096883]
41. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM 2017 May 24;60(6):84-90. [doi: 10.1145/3065386]

XSL•FO
RenderX

42. El-Hag NA, Sedik A, El-Shafai W, El-Hoseny HM, Khalaf AAM, El-Fishawy AS, et al. Classification of retinal images based on convolutional neural network. Microsc Res Tech 2021 Mar;84(3):394-414. [doi: 10.1002/jemt.23596] [Medline: 33350559]
43. Hussain S, Faheem MR. Separation of veins and arteries for estimating hypertensive retinopathy in fundus images. Biomed Res Ther 2016 Jun 26;3(6):673-678. [doi: 10.7603/s40730-016-0027-3]
44. Zhou L, Yu Q, Xu X, Gu Y, Yang J. Improving dense conditional random field for retinal vessel segmentation by discriminative feature learning and thin-vessel enhancement. Comput Methods Programs Biomed 2017 Sep;148:13-25. [doi: 10.1016/j.cmpb.2017.06.016] [Medline: 28774435]
45. Intriago-Pazmino M, Ibarra-Fiallo J, Crespo J, Alonso-Calvo R. Enhancing vessel visibility in fundus images to aid the diagnosis of retinopathy of prematurity. Health Informatics J 2020 Dec;26(4):2722-2736 [FREE Full text] [doi: 10.1177/1460458220935369] [Medline: 32674723]
46. Goel S, Gupta S, Panwar A, Kumar S, Verma M, Bourouis S, et al. Deep learning approach for stages of severity classification in diabetic retinopathy using color fundus retinal images. Math Probl Eng 2021 Nov 24;2021:1-8. [doi: 10.1155/2021/7627566]

Abbreviations

AI: artificial intelligence
AMD: age-related macular degeneration
CNN: convolutional neural network
DR: diabetic retinopathy
OCT: optical coherence tomography
SCI: Science Citation Index
WoSCC: Web of Science Core Collection

Edited by R Kukafka; submitted 24.02.22; peer-reviewed by FM Calisto, G Nneji, H Monday; comments to author 19.03.22; revised version received 25.03.22; accepted 04.04.22; published 14.06.22.

Please cite as:
Zhao J, Lu Y, Qian Y, Luo Y, Yang W
Emerging Trends and Research Foci in Artificial Intelligence for Retinal Diseases: Bibliometric and Visualization Study
J Med Internet Res 2022;24(6):e37532
URL: https://www.jmir.org/2022/6/e37532
doi:10.2196/37532
PMID:35700021

©Junqiang Zhao, Yi Lu, Yong Qian, Yuxin Luo, Weihua Yang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.06.2022. 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

Research Trends in Immune Checkpoint Blockade for Melanoma: Visualization and Bibliometric Analysis

Yantao Xu[1,2,3,4,5], MD; Zixi Jiang[1,2,3,4,5], MD; Xinwei Kuang[1,3,4,6], MD; Xiang Chen[1,2,3,4,6*], MD, PhD; Hong Liu[1,2,3,4,6*], MD, PhD

[1]Department of Dermatology, Xiangya Hospital, Central South University, Changsha, China

[2]National Engineering Research Center of Personalized Diagnostic and Therapeutic Technology, Xiangya Hospital, Central South University, Changsha, China

[3]Hunan Engineering Research Center of Skin Health and Disease, Xiangya Hospital, Central South University, Changsha, China

[4]Hunan Key Laboratory of Skin Cancer and Psoriasis, Xiangya Hospital, Central South University, Changsha, China

[5]Xiangya School of Medicine, Central South University, Changsha, China

[6]National Clinical Research Center of Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China

[*]these authors contributed equally

Corresponding Author:
Hong Liu, MD, PhD
Department of Dermatology
Xiangya Hospital
Central South University
87 Xiangya Road, Kaifu District
Changsha, 410005
China
Phone: 86 0731 89753999
Email: hongliu1014@csu.edu.cn

Abstract

Background: Melanoma is one of the most life-threatening skin cancers; immune checkpoint blockade is widely used in the treatment of melanoma because of its remarkable efficacy.

Objective: This study aimed to conduct a comprehensive bibliometric analysis of research conducted in recent decades on immune checkpoint blockade for melanoma, while exploring research trends and public interest in this topic.

Methods: We summarized the articles in the Web of Science Core Collection on immune checkpoint blockade for melanoma in each year from 1999 to 2020. The R package bibliometrix was used for data extraction and visualization of the distribution of publication year and the top 10 core authors. Keyword citation burst analysis and cocitation networks were calculated with CiteSpace. A Gunn online world map was used to evaluate distribution by country and region. Ranking was performed using the Standard Competition Ranking method. Coauthorship analysis and co-occurrence were analyzed and visualized with VOSviewer.

Results: After removing duplicates, a total of 9169 publications were included. The distribution of publications by year showed that the number of publications rose sharply from 2015 onwards and either reached a peak in 2020 or has yet to reach a peak. The geographical distribution indicated that there was a large gap between the number of publications in the United States and other countries. The coauthorship analysis showed that the 149 top institutions were grouped into 8 clusters, each covering approximately a single country, suggesting that international cooperation among institutions should be strengthened. The core author extraction revealed changes in the most prolific authors. The keyword analysis revealed clustering and top citation bursts. The cocitation analysis of references from 2010 to 2020 revealed the number of citations and the centrality of the top articles.

Conclusions: This study revealed trends in research and public interest in immune checkpoint blockade for melanoma. Our findings suggest that the field is growing rapidly, has several core authors, and that the United States is taking the lead position. Moreover, cooperation between countries should be strengthened, and future research hot spots might focus on deeper exploration of drug mechanisms, prediction of treatment efficacy, prediction of adverse events, and new modes of administration, such as combination therapy, which may pave the way for further research.

(J Med Internet Res 2022;24(6):e32728) doi:10.2196/32728

XSL•FO
RenderX

KEYWORDS

melanoma; immune checkpoint blockade; bibliometric; research trends; dermatology; cancer

Introduction

In the past 10 years, although the frequency of melanoma has continued to increase, the lethality of advanced melanoma has decreased. Nevertheless, melanoma is still one of the most life-threatening skin cancers [1]. Globally, melanoma cases have grown rapidly. The incidence of new melanomas exceeded 350,000 in 2015 [2]; moreover, the rate of new melanomas has continually increased for the last 20 years [3]. Immune checkpoint blockade (ICB) is widely used in the treatment of melanomas, especially those with negative regulators, such as T-lymphocyte associated protein-4 (CTLA-4), programmed death receptor 1 (PD-1), and the ligands of PD-1 (PD-L1). ICBs are often chosen as the subject of clinical trials. The reason for the remarkable efficacy of ICBs in melanoma is the biological features of the cancer. Melanoma is often described as an archetypal immunogenic cancer, which guarantees the efficacy of immunotherapy; this is supported by many studies that have observed tumor progression and increased rates of melanoma in immunosuppressed individuals [4,5]. Moreover, melanoma always carries a large tumor mutation burden [6], which increases the probability of driving out a stronger immune response in the host. Because of these biological and clinical features, ICBs for melanoma have been thoroughly researched in the past two decades and have gradually became a research hotspot.

Bibliometric analysis is a quantitative science approach using methods such as co-occurrence analysis and citation analysis to evaluate research performance [7,8]. In the health care field, bibliometrics are mostly used to measure the influence or impact of research articles. Bibliometric methods estimate how much influence or impact a selected research article may have on future research, and the results are especially valuable for those topics that are gradually becoming more intriguing. However, there has been no bibliometric analysis of ICB for melanoma. In this study, we conducted a comprehensive bibliometric analysis covering recent decades while exploring research trends and the public interest in ICB for melanoma. We determined the research landscape of ICB for melanoma in terms of chronological distribution, geographical distribution, publication sources, author publications, and cocitations. We also identified the co-occurrence of authors, organizations, and keywords. The purpose of this study is to provide a systematic summary of research trends and the public interest in ICB for melanoma from an evaluative bibliometric perspective.

Methods

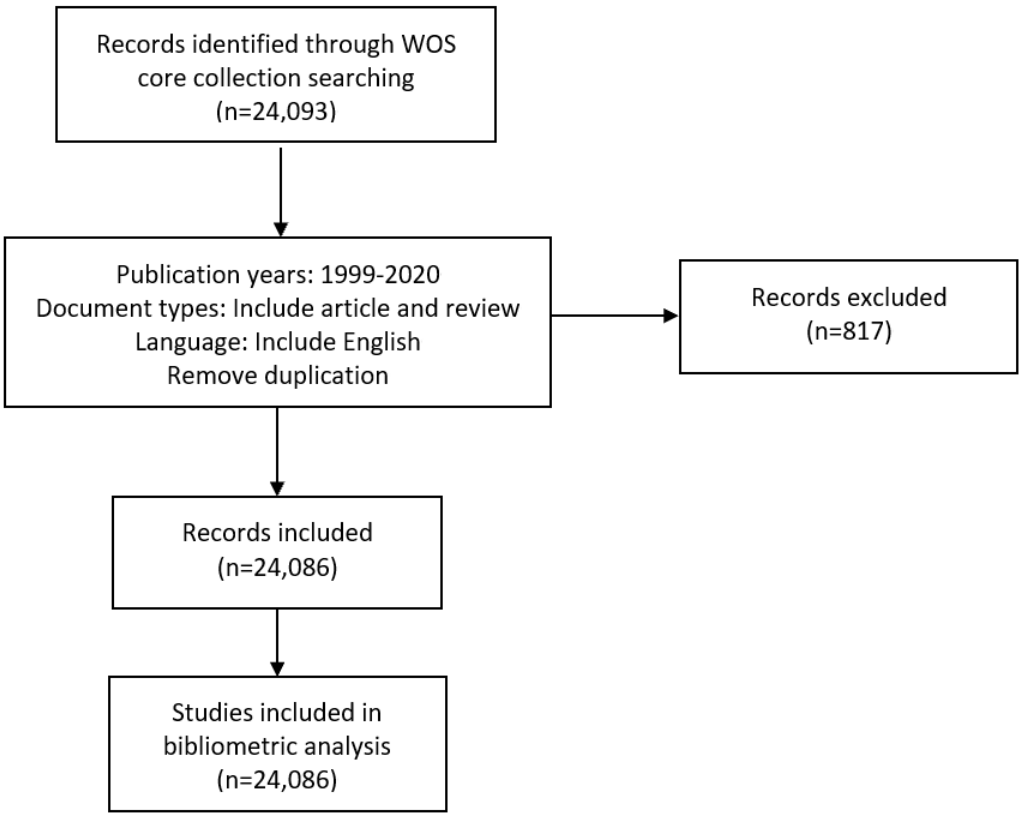
Data Sources and Search Strategy

Bibliographic data for the analysis were all acquired from the Web of Science Core Collection, which includes the Science Citation Index Expanded, Social Science Citation Index, and Emerging Source Citation Index [9]. To perform a comprehensive literature search of immune checkpoint inhibitors in melanoma, we designed a systematic search strategy. Generally, the search strategy was as follows: (TS=(ipilimumab OR pembrolizumab OR nivolumab OR immunotherapy OR “immune checkpoint blockade” OR PD-1 OR PD-L1 OR CTLA-4) OR TI=(ipilimumab OR pembrolizumab OR nivolumab OR immunotherapy OR “immune checkpoint blockade” OR PD-1 OR PD-L1 OR CTLA-4 OR yervoy OR Keytruda OR opdivo) OR AB=(ipilimumab OR pembrolizumab OR nivolumab OR immunotherapy OR “immune checkpoint blockade” OR PD-1 OR PD-L1 OR CTLA-4 OR yervoy OR Keytruda OR opdivo)) AND (TS=(melanoma OR melanocarcinoma) OR TI=(melano* OR melanoma OR melanocarcinoma) OR AB=(melano* OR melanoma OR melanocarcinoma)). The language was restricted to English and the document type was limited to articles and reviews. The time span of the search excluded the year 2022 for clearer annual results.

A total of 24,093 documents were retrieved from the Web of Science Core Collection. After excluding documents that were published as preprints in 2021 and then published as final versions in 2022 and documents with an unknown publication date, 24,086 documents remained in the bibliometric analysis and visualization. The search details are presented as a flowchart (Figure 1). The search was completed on December 6, 2021.

XSL•FO
RenderX

Figure 1. Detailed search flowchart, showing steps in the identification and screening of papers. Publication years spanned 1999 to 2021. Only documents published in English were included. Endnote was used to remove duplicates. The R package Bibliometrix was used to remove documents that were published as preprints in 2021 by extracting the publication date.



Data Extraction and Analysis

The retrieval characteristics used for publications on ICB for melanoma included the distribution of publication year, country and region, organization, journal, core authors, keywords, and key references. The detailed search strategy is shown in Multimedia Appendix 1. Bibliometric analysis and network visualization were performed with VOSviewer (version 1.6.14, Leiden University), CiteSpace (version 5.7.R5W, Drexel University), and the bibliometrix package in R (version 3.6, R Foundation). The bibliometrix package was used for data extraction and visualization of the distribution of publication year and the top 10 core authors. The keywords citation burst analysis and cocitation network analysis were performed with CiteSpace. The Gunn online world map was used to evaluate the distribution of countries and regions. Ranking was performed using the standard competition ranking method. Other analyses, including the coauthorship analysis, co-occurrence analysis, and visualization, were conducted with VOSviewer.

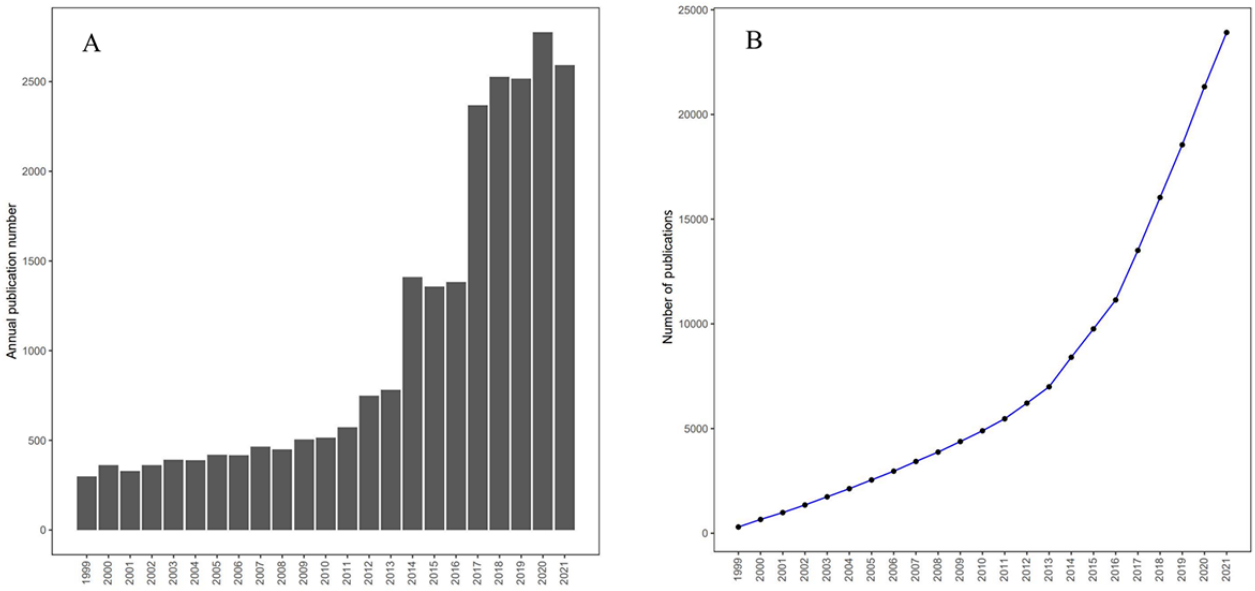
Results

Distribution of Publications

Figure 2 shows the chronological distribution of publications by year as a bar chart (Figure 2A). From 1999 to 2013, the annual number of publications steadily grew, with no obvious research trends, and remained relatively stable. The annual number of publications then rose sharply from 2015 to 2020. Figure 2B shows the total, cumulative number of publications as a plot. There was a relatively slow increase in the cumulative number of publications from 1999 to 2015, with the number of publications growing sharply from 2015 to 2017 onwards; the peak either occurred in 2020 or has yet to occur.

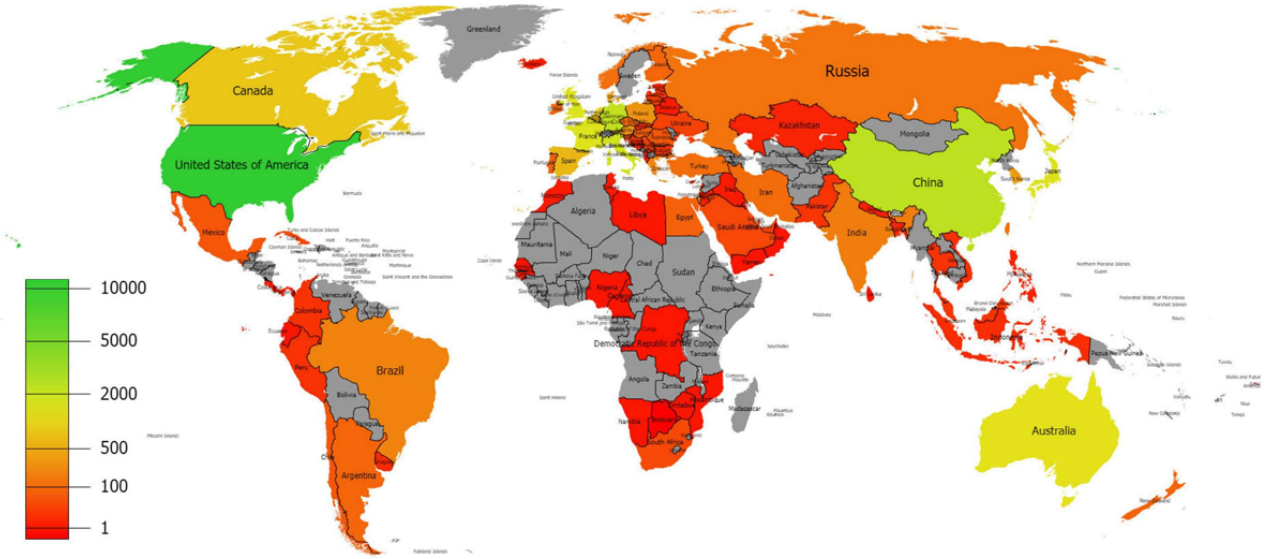
XSL•FO
RenderX

Figure 2. Distribution of publications by year. (A) The cumulative number of publications and (B) the annual number of publications on immune checkpoint blockade for melanoma. The peak of cumulative publications occurred in 2020. The annual number of publications increased relatively slowly from 1999 to 2021 and sharply from 2014 to 2017 and onwards. The peak of annual publication either occurred in 2020 or has yet to occur. The publication data for 2021 does not include data for December.



As for geographical distribution, 24,086 documents were published from 117 different countries and regions. Studies involving multiple countries were included in the analysis, with each country being counted individually. We classified documents by country and visualized the spatial distribution as a heatmap (Figure 3). Table 1 lists the top 12 most prolific countries. In total, the country with the largest number of publications was the United States (11,113/24,086 publications, 46.1%), far surpassing China (2345/24,086 publications, 9.7%) and Germany (2223/24,086 publications, 9.2%). As for citations, America was also far ahead. It is interesting that although China had the second largest number of publications, the number of citations lagged far behind other countries, which made the number of citations per publication rather small.

Figure 3. Geographical distribution of global publications. The green-to-red gradient represents a decreasing number of publications. Gray represents countries with no publications.



XSL•FO
RenderX

Table 1. Top 12 most productive countries and regions.

Rank	Country/region	Publications	Citations	Citations per publication
1	United States	11,113	642,788	57.84
2	China (mainland)	2345	45,215	19.28
3	Germany	2223	129,248	58.14
4	Italy	1847	87,903	47.59
5	France	1601	128,827	80.47
6	Japan	1598	53,323	33.37
7	England	1464	97,981	66.93
8	Australia	1381	81,059	58.70
9	Netherland	1087	75,699	69.64
10	Switzerland	945	50,695	53.64
11	Canada	862	77,164	89.52
12	Spain	658	58,543	88.97

Analysis of Leading Organizations and Public Sources

The information on leading organizations was analyzed with VOSviewer. Generally, 24,086 documents were published by 13,359 different organizations. After merging duplicates and excluding disjointed organizations, a final total of 243 organizations met the inclusion threshold and are shown in the visualization. The top 10 most productive organizations are listed in Table 2. The most prolific organization was the Memorial Sloan Kettering Cancer Center (903/24,086 publications, 3.7%), followed by the University of Texas MD Anderson Cancer Center (859/24,086 publications, 3.6%) and the National Cancer Institute (645/24,086 publications, 2.7%). Among the top 10 institutions, 9 of 10 were from the United States, which corresponded to the distribution by country and region. We also conducted a coauthorship analysis of the organizations (Figure 4). We found that all 243 top published institutions were grouped into clusters, with each cluster representing approximately one country, except for the United States, which dominated 2 clusters and had wide correlations. The red and lower yellow clusters mainly represent American institutions, including Memorial Sloan Kettering Cancer Center, the MD Anderson Cancer Center, and the National Cancer Institution. The blue cluster represents Johannes Gutenberg University, in Germany. The upper yellow cluster represents universities and institutions in France. The green cluster includes other European countries and Australia. Two rather further away clusters mainly represent Japan and Korea, in light blue, and China, in purple, with few links with other clusters. This result suggests that the United States apparently led this topic and that international cooperation among various institutions from different countries should be strengthened, especially for institutions in China, Japan, and Korea.

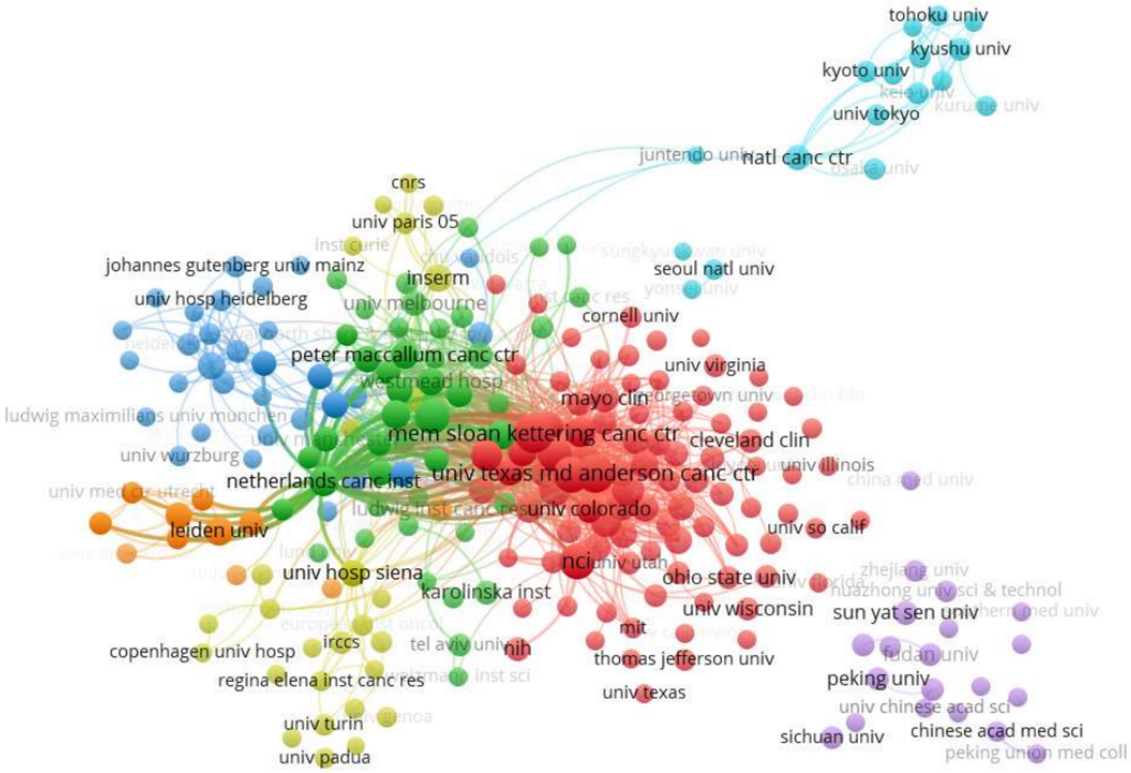
Table 2. Top 10 most productive organizations.

Rank	Organization	Country	Articles	Citations	Total link strength[a]
1	Memorial Sloan Kettering Cancer Center	United States	903	120,565	3183
2	University of Texas MD Anderson Cancer Center	United States	859	54,089	2381
3	National Cancer Institute	United States	645	71,055	759
4	Dana-Farber Cancer Institute	United States	617	82,093	3079
5	University of Sydney	Australia	537	33,668	2749
6	University of Pittsburgh	United States	505	25,886	1310
7	Harvard Medical School	United States	480	21,477	1317
8	University of California Los Angeles	United States	476	50,391	1926
9	Massachusetts General Hospital	United States	440	32,663	1634
10	Mayo Clinic	United States	367	27,207	823

[a]Total link strength in VOSviewer represents all links between a given node and other nodes, which indicates how the entry interacts with other entries. The strength of a link is given by a nonnegative number. If one node has no links with other nodes, the total strength of the link equals zero.

XSL•FO
RenderX

Figure 4. Coauthorship analysis of organizations. Plot showing a coauthorship analysis of organizations. The normalization method was fractionalization. The weight was the number of publications. The thickness of the lines indicates the strength of coauthorship relationships. Different colors indicate clusters.



Core journals were identified by analysis of publication sources. After analyzing bibliographies, we extracted the top 10 most prolific journals along with their impact factor (IF) in 2020 and 2021 in the field of ICB for melanoma (Table 3). We found that the most prolific journal was the *Journal of Clinical Oncology* (impact factor 44.544), which ranked first, with 1051 documents published. The second and the third most-prolific journals were *Cancer Immunology,Immunotherapy* (677 publications, IF 6.958) and *Cancer Research* (627 publications, IF 12.701). IF for the top 10 most prolific journals in 2019 ranged from 4.456 for the *Journal of Immunology* to 44.544 for the *Journal of Oncology*. Our findings for the total number of publications and IF suggest that *Cancer Research* might be the most influential journal in the field of ICB for melanoma.

Table 3. Top 10 most prolific journals.

Rank	Journal	Publications	Impact factor (2020)	Impact factor (2021)
1	*Journal of Clinical Oncology*	1051	32.956	44.544
2	*Cancer Immunology, Immunotherapy*	677	5.442	6.958
3	*Cancer Research*	627	9.727	12.701
4	*Clinical Cancer Research*	602	10.107	12.531
5	*Journal for Immunotherapy of Cancer*	589	9.913	13.751
6	*Oncoimmunology*	508	5.869	8.11
7	*Annals of Oncology*	498	18.274	32.976
8	*Journal of Immunotherapy*	465	4.11	4.456
9	*Journal of Immunology*	359	4.886	5.422
10	*Frontiers in Immunology*	336	5.085	7.561

Analysis of Coauthorship and Core Authors

Information on authors and co-authors was also analyzed with VOSviewer. A total of 24,086 publications were produced by a total of 93,587 authors. The 3 most important evaluation criteria for core authors included the number of published documents, total citations, and the H index. Therefore, we extracted and visualized the top 10 most prolific authors

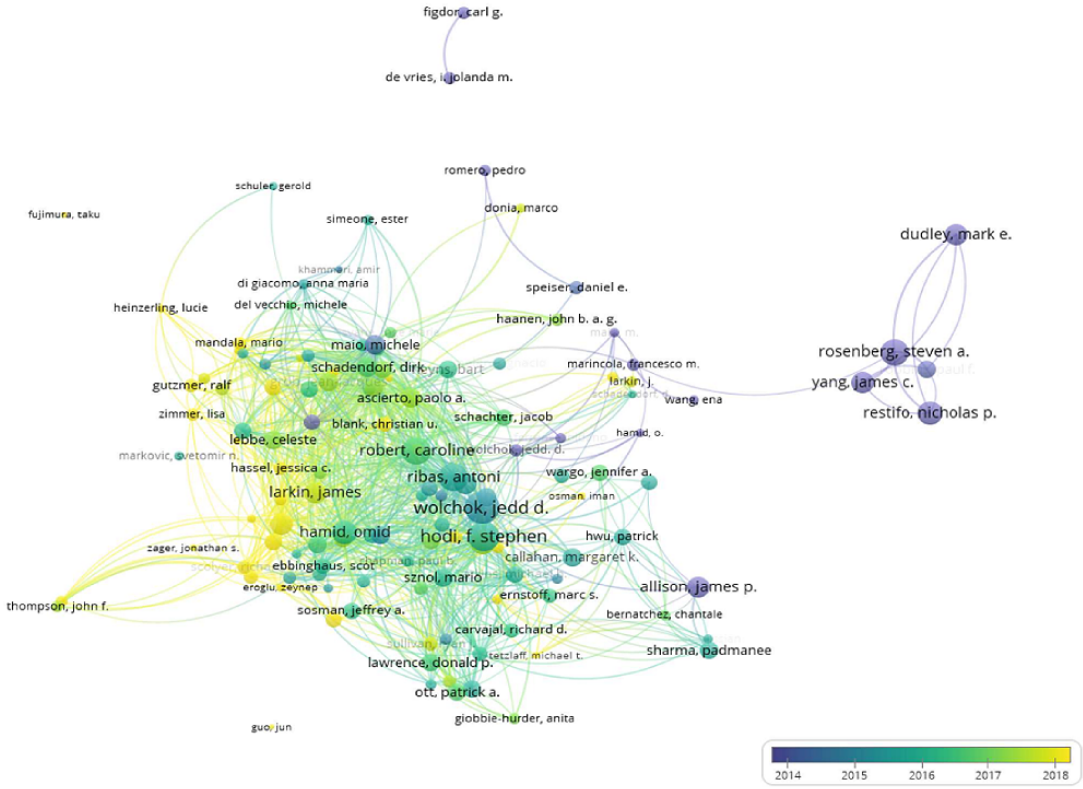
XSL•FO
RenderX

according to these criteria (Table 4 and Figure 5). Paolo A Ascierto, the director of the Unit of Melanoma, Cancer Immunotherapy and Innovative Therapy at the National Tumour Institute, Fondazione G. Pascale, was the most productive author in this field, publishing 312 articles and being cited 10,756 times in general, followed by F Stephen Hodi, the director of the Melanoma Center and Center for Immuno-Oncology of the Dana-Farber/Brigham and Women's Cancer Center, with 301 publications and 35,302 total citations, and Caroline Robert, from the Institut de Cancérologie Gustave Roussy. Five of the top 10 cited authors were from the United States, and interestingly, the top 10 most productive authors together accounted for 9.1% of the total literature, showing the dominance of the United States in this research field. It is to be noted that Georgina V Long of Australia (190 publications), Dirk Schadendorf of Germany (184 publications), and Reinhard Dummer of Switzerland (152 publications) occupied places 6, 7 and 9 in the table, which, combined with the previous findings on geographical distribution, indicate that these countries also have important research roles in the field of ICB for melanoma.

Table 4. Top 10 core authors by number of publications.

Rank	Authors	Organizations	Publications	Citations	H index
1	Paolo A Ascierto	National Tumour Institute, Fondazione G. Pascale (Italy)	312	10,756	64
2	F Stephen Hodi	Dana-Farber/Brigham and Women's Cancer Center (US)	301	35,302	93
3	Caroline Robert	Institut de Cancérologie Gustave Roussy (France)	265	27,523	78
4	Jedd D Wolchok	Memorial Sloan Kettering Cancer Center (US)	252	47,787	99
5	Antoni Ribas	University of California Los Angeles (US)	233	30,109	88
6	Georgina V Long	University of Sydney (Australia)	190	13,657	67
7	Dirk Schadendorf	University Hospital Essen (German)	184	11,733	68
8	John M Kirkwood	University of Pittsburgh Medical Center (US)	159	8493	56
9	Reinhard Dummer	University of Zurich (Switzerland)	152	7641	46
10	Steven A Rosenberg	National Cancer Institute (US)	139	27,236	99

Figure 5. Overlay visualization of coauthorship relationships between authors. The analysis method was Linlog/modularity. The weight was citations. Scores are the average year of publication. The thickness of the lines indicates the strength of the relationships. The colors of the circles represent the average year of publication.



XSL•FO
RenderX

Figure 5 is an overlay visualization of the coauthorship relationships between authors. A total of 93,587 authors were analyzed, of whom 135 met the inclusion threshold. The visualization suggested that most of the top influential authors, such as Jedd D Wolchok, F Stephen Hodi, and Antoni Ribas, had close collaborations. In addition, we found the average year of publication for these authors was between 2015 and 2016, while other authors, such as Georgina V Long, published more actively after 2018, which may suggest that they have the potential to take the lead in the future.

Analysis of Keywords and Burst Terms

In total, 30,486 keywords were extracted from 24,086 documents after removing duplicates. We used a network and overlay visualization of author-given keywords to analyze the co-occurrence of keywords. A total of 30,486 keywords were analyzed, of which the 180 most frequently occurring that met the inclusion threshold were grouped into 2 clusters (Figure 6A) and grouped by date of publication (between 2016 and 2018) (Figure 6B). The date range for publication was chosen to correspond to the high-growth phase of publication seen in Figure 2. For the network map, the keywords were mainly distinguished into 2 clusters. One mainly represented keywords related to tumor biology and the early, discovery stages of research into immunotherapy as a potentially promising modality for melanoma. These keywords are shown in red and include "antigen," "dendritic cells," "T cells," "tumor microenvironment," "immunotherapy," "immune checkpoint," "melanoma," and "metastatic melanoma." The other cluster, shown in green, generally included terms related to clinical oncology issues, including the names of US Food and Drug Administration (FDA)-proven antibodies, such as "nivolumab," "pembrolizumab," "atezolizumab," and "ipilimumab"; words related to outcomes, such as "efficacy," "safety," and "adverse event"; and keywords that appear frequently in phase II and phase III trials, such as "survival," "safety," "double-blind," "open-label," and "multi-center." Interestingly, this cluster included nearly all the keywords related to targeted therapies and radiotherapy, such as "BRAF," "MEK," "vemurafenib," and "dabrafenib," as well as the names of other types of cancer, the novel usages of immunotherapy words, such as "combination" and "adjuvant therapy," and multiple indicators for evaluating efficacy and safety. It might be the case that these interventions are now establishing more links with immunotherapy and will potentially become research hot spots in the future. The overlay map of keywords grouped by date of publication showed that research hot spots changed over time, starting with "target therapy" and moving on to "CTLA-4 inhibitor," "PD-1/PD-L1 inhibitor," and "FDA-proven ICB," which is similar to the order in which these technologies developed.

Figure 6. Co-occurrence analysis of keywords. These two plots show the co-occurrence of keywords. The normalization method we chose was Linlog/modularity. The weight was occurrence for each plot. (A) shows the 180 top-occurring items among 30,486 keywords, grouped into 2 clusters, with the colors of the circles representing each cluster. (B) shows the keywords grouped by year of publication, with the colors of the circles representing the average year.

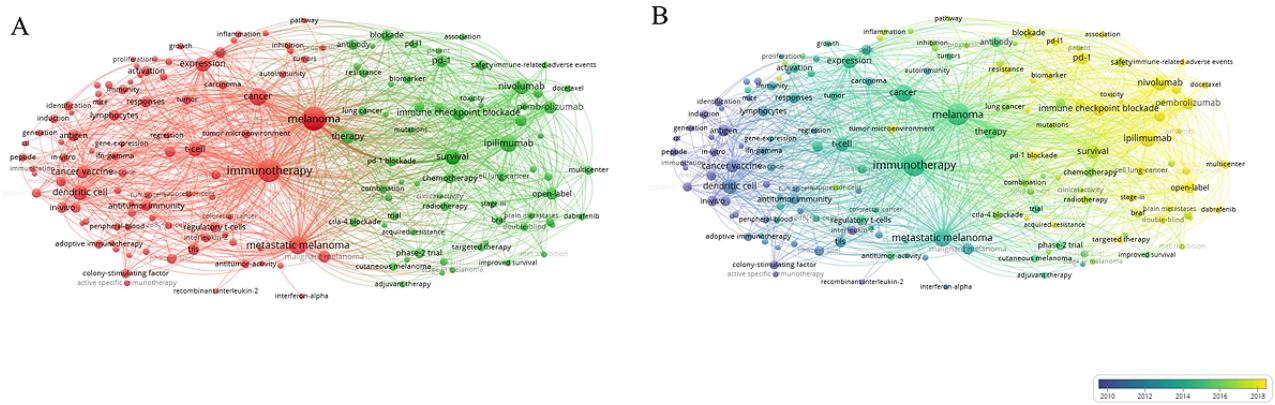
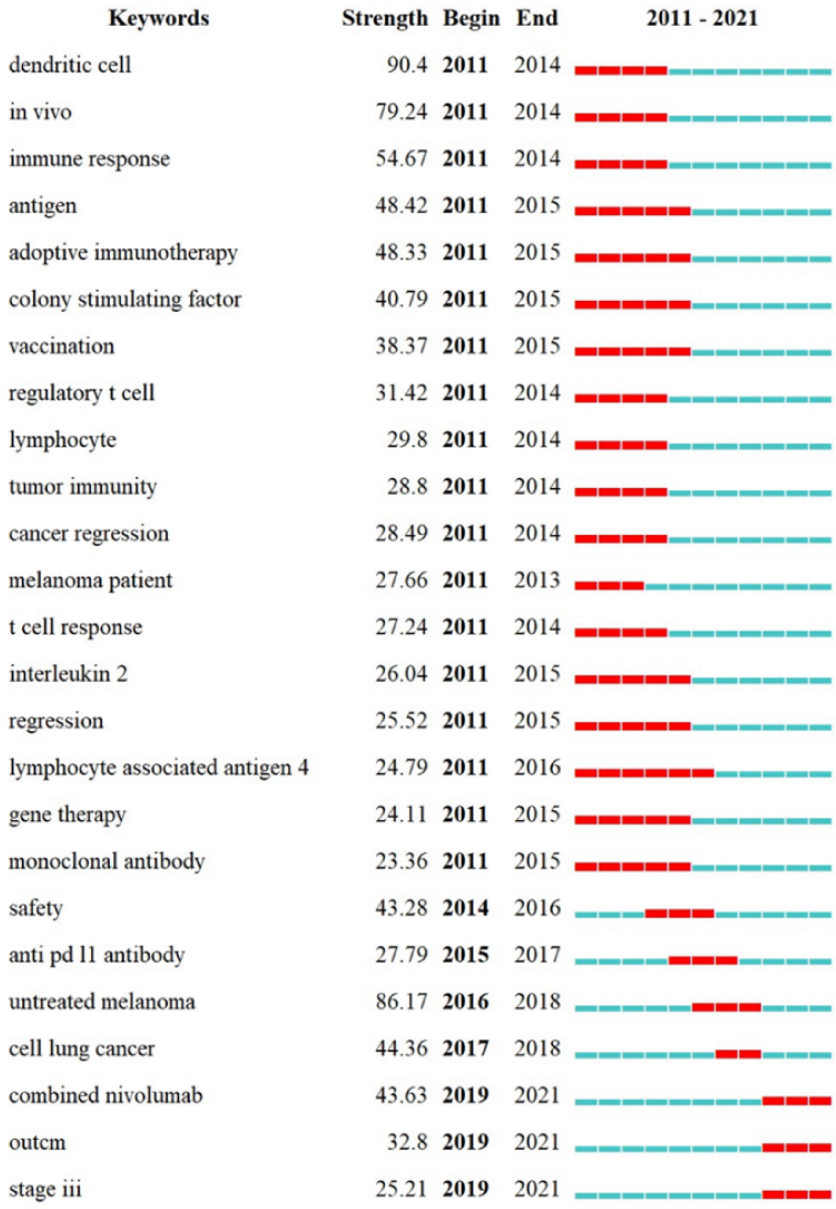


Figure 7 shows the top 25 keywords with the strongest frequency burst, which suggests a keyword that has undergone a great change in a short period of time (as analyzed with CiteSpace). "Dendritic cell," "in vivo" and other mechanism- and trial-related keywords had rather stronger strength, which continued from 2011 to 2015, suggesting that possible mechanisms and clinical applications were a sustained research hot spot for these years. However, in the most recent 5 months, the burst keywords changed to include terms such as "safety," "combined therapy," and "stage III trial", suggesting that the interests of the researchers changed.

XSL•FO
RenderX

Figure 7. Top 25 keywords with the strongest frequency bursts. A strong frequency burst indicates that a variable has undergone a great change in a short period of time. The red bars indicate the durations of the bursts.



Citation Analysis

The number of citations of the publications was mainly extracted with bibliometrix. The top 10 most highly cited documents were extracted and are listed in Table 5. Generally, the number of citations ranged from 3037 to 9113. An article published in the *New England Journal of Medicine* titled “Improved survival with ipilimumab in patients with metastatic melanoma” ranked first, with 9113 total citations. Only one of the top 10 articles was a review (“The Blockade of Immune Checkpoints in Cancer Immunotherapy,” published in *Nature Reviews Cancer* in 2012). It is likely that this review was cited so often because it was the first comprehensive review of this topic. A total of 7 of the 9 articles were published in the *New England Journal of Medicine*. This shows the dominant position of this journal in the publication of research in the medical category, especially in the publication of high-quality research.

XSL•FO
RenderX

Table 5. Top 10 most highly cited publications.

Rank	Title	DOI[a]	Source	Publication date	Total citations[b]
1	Improved Survival with Ipilimumab in Patients with Metastatic Melanoma [10]	10.1056/NEJMoa1003466	*New Engl J Med*	Aug 2010	9549
2	Safety, Activity, and Immune Correlates of Anti-Pd-1 Antibody in Cancer [11]	10.1056/NEJMoa1200690	*New Engl J Med*	Jun 2012	7926
3	The Blockade of Immune Checkpoints in Cancer Immunotherapy [12]	10.1038/nrc3239	*Nat Rev Cancer*	Apr 2012	7160
4	Safety and Activity of Anti-Pd-L1 Antibody in Patients with Advanced Cancer [13]	10.1056/NEJMoa1200694	*New Engl J Med*	Jun 2012	5026
5	Pembrolizumab Versus Chemotherapy for Pd-L1-Positive Non-Small-Cell Lung Cancer [14]	10.1056/NEJMoa1606774	*New Engl J Med*	Nov 2016	4794
6	Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma [15]	10.1056/NEJMoa1504030	*New Engl J Med*	Sep 2015	4751
7	PD-1 Blockade Induces Responses by Inhibiting Adaptive Immune Resistance [16]	10.1038/nature13954	*Nature*	Nov 2014	3514
8	Nivolumab in Previously Untreated Melanoma Without Braf Mutation [17]	10.1056/NEJMoa1412082	*New Engl J Med*	Jan 2015	3421
9	Pembrolizumab Versus Ipilimumab in Advanced Melanoma [18]	10.1056/NEJMoa1503093	*New Engl J Med*	Jun 2015	3376
10	Predictive Correlates of Response to the Anti-Pd-L1 Antibody Mpdl3280a in Cancer Patients	10.1038/nature14011	*Nature*	Nov 2014	3087

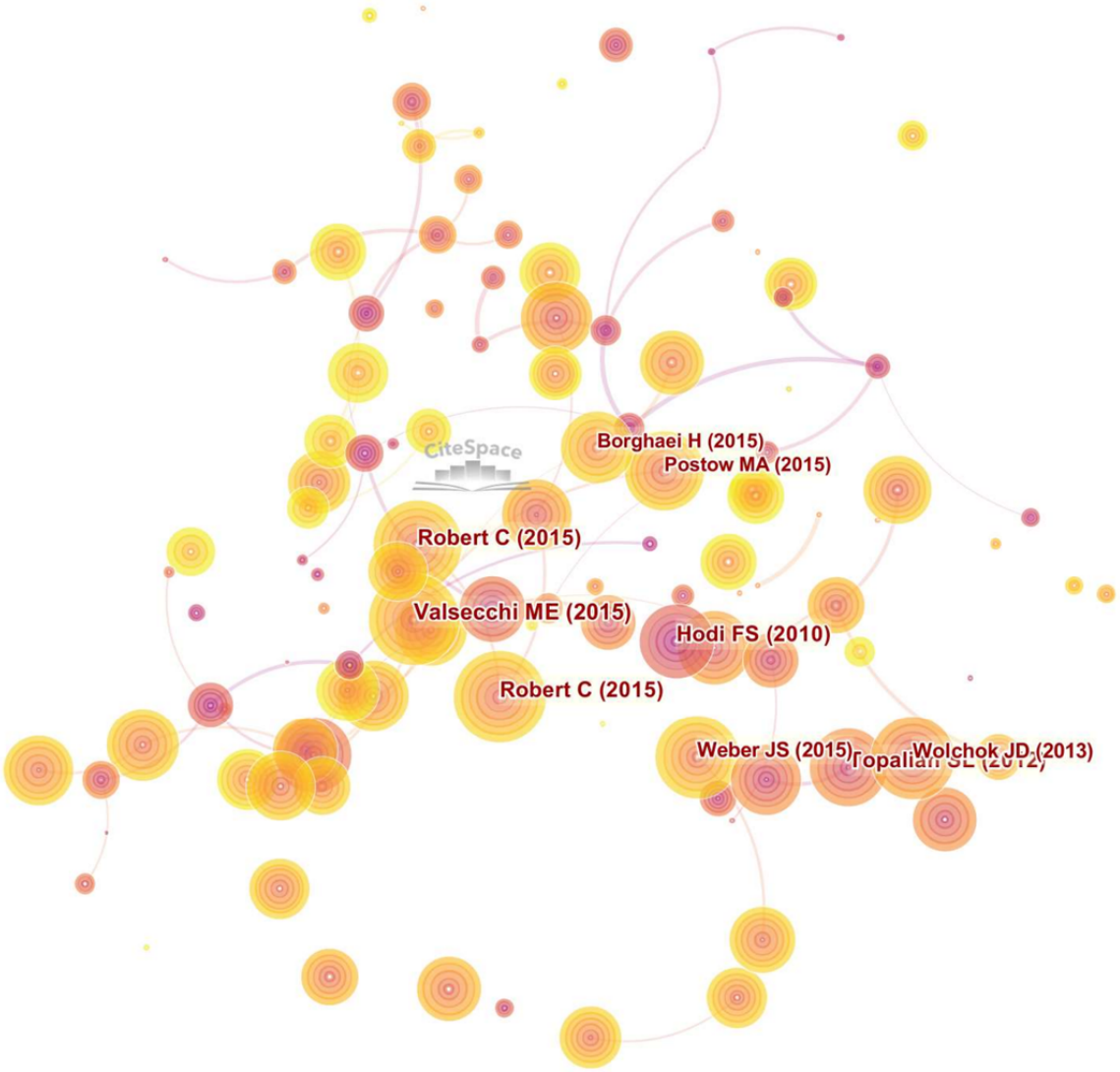
[a]DOI: Digital Object Identifier.

[b]Total citations were until the end of December 2021.

For a comprehensive analysis of citations, we used CiteSpace (version 5.8R3) to evaluate cocitation references (Figure 8). we performed a cocitation analysis of references from 2010 to 2020. In CiteSpace, the size of a circle indicates the number of documents cited. The purple area of the circle indicates the centrality of a document. The analysis revealed no significant centrality in the documents, indicating that the literature was largely scattered.

XSL•FO
RenderX

Figure 8. Cocitation analysis of references. Using CiteSpace, we performed a cocitation analysis of references from 2010 to 2020. In CiteSpace, the size of a circle indicates the number of documents cited. The purple area of the circle indicates the centrality of a document.



Discussion

Principal Findings

This study updates current knowledge on research interests related to ICB for melanoma, providing researchers and physicians an overview of the landscape of the field and potential future research hot spots. We conducted a comprehensive search of literature published on this topic before December 2021 in the Web of Science Core Collection. We retrieved 24,086 bibliographies and performed a bibliometric analysis.

First, using the bibliometric method, we analyzed chronological trends in the publications. The results show that from 1999 to 2013, the annual number of publications was rather small, with a small, linear slope of growth. The number of newly published papers from 1999 to 2013 remained under 200, with rather small growth every year. The next period was from 2014 to 2016, when publications related to ICB grew rapidly. The annual number of publications grew to over 1000, but did not reach 2000, which is consistent with previous research [19]. The third period was from 2017 to 2021, when the annual number of publications grew to over 2000, representing maturity in the theoretical aspect of this field. The number of publications has continued to increase, and the topic has gradually become a hot spot.

The chronological trend was reflected in several critical articles and specific time points, which enables us to reveal the roadmap for this field. The very first research on ICB was on CTLA-4 blockade, which was conduced beginning in 1987 and was first proved in 2011 [20,21]. The recombinant human immunoglobulin G1 and G2 monoclonal antibodies of CTLA-4, known as ipilimumab and tremelimumab, respectively, were trialed in melanoma and other advanced cancers, and both showed good efficacy [10,22-31]. As for PD-1 and PD-L1, since the very first paper on PD-1 was published in 1991, the annual number of publications remained less than 5 for a long period, from 1999 to 2003. The annual number of publications did not

XSL•FO
RenderX

reach 300 until a breakthrough on cancer immunotherapy in 2013 [32]. After that, the FDA approved anti-PD-1 antibodies (pembrolizumab and nivolumab) for advanced metastatic melanoma in 2014 and an anti-PD-L1 antibody (atezolizumab) in 2016. After that, the number of publications rapidly increased, reaching the thousands [33-35]. Our results for chronological distribution, unsurprisingly, suggest that the topic of ICB for melanoma has gradually become a research hot spot, and that it is currently in a major development period. Theory developed rapidly during this period, and the number of papers increased rapidly. Moreover, the growth curve was sharp and the gradient of the curve did not slow down during 2020. We predict that this field will continue to develop in the next few years.

As for geographical distribution, among the 80 different countries and regions involved in this bibliometric analysis, the most prolific are listed in Table 2. The United States published most of the documents in this field, and Table 5 indicates that nearly all core documents were produced in partnership with American institutions [10-19]. China is the second most prolific country, but considering the number of citations, China is still far from taking the lead in this area. Other countries with a high number of publications, such as Japan, Australia, and various European counties, have a good scientific base in this field. The situation is also reflected in the most productive organizations and the most productive authors; 9 of the 10 most productive organizations and 6 of the 10 most productive authors are from the United States, which has proven the dominance of the United States in this research field.

For the most productive journals, considering the evidence for the total number of publications and IF, the *Journal of Clinical Oncology* might be the most influential journal in the topic of immune checkpoint inhibitors in melanoma. Papers published by the *Journal of Clinical Oncology* in 2021 included several that mainly concentrated on the long-term outcomes of ICB [36], the combination of ICB and other therapies, and expansion of the indications for ICB [37,38]. We also compared the IF in 2020 and 2021 in core journals in the field of ICB for melanoma and found that the IF of all top 10 core journals grew. However, considering that oncology journals generally had increased IFs in these years, we cannot make the conclusion that all journals increased their impact.

In the aspect of cooperation between authors, the network analysis showed that most cooperation took place within countries, and that there was little cooperation between countries. The same phenomenon was also revealed by a coauthorship analysis of organizations, in which clusters showed intricate connections within countries and lesser connections between clusters. These findings suggest that cooperation between states represents an area that should be strengthened.

We also performed analysis of keywords and burst terms to investigate research trends, finding that the change in focus was remarkable. Generally, research trends and the public interest changed in two major aspects: from the laboratory to translational medicine and clinical research, as well as from early ICB developments, such as CTLA-4, to later ones, such as PD-1 and PD-L1 blockades. The focus of research gradually changed from mechanisms to efficacy and adverse events. This indicates that the theory was becoming mature and that the application of ICB therapy was being explored, including enhancing its efficacy, reducing its adverse effects, and expanding its use to other, more specific cancer types [39].

From the initial research on the mechanisms of immunotherapy, including the alteration of immune cells and immune molecules under ICB treatment to subsequent translational, clinical research into the interactions of immune checkpoints with costimulatory molecules, cancer drive genes, and cancer hallmarks, studies investigating the mechanism of ICB have been maturing. In the next several years, screening of biomarkers to predict treatment efficacy and adverse events, improve the efficacy of ICB and reduce adverse events, explore drug combinations, and extend the indications for ICB might become hot spots in this clearly evolving field of research.

Strengths and Limitations

As far as we know, this is the first study to use a bibliometric analysis to investigate research trends and public interest in ICB for melanoma. Our bibliometric analysis was much more comprehensive and intuitive than a literature review would have been, because of our use of systematic searching and quantitative statistical analysis. Moreover, we used not only CiteSpace, but also VOSviewer and the R package bibliometrix for better data extraction, bibliometric analysis, and visualization. However, this study still has some limitations. We only extracted literature from the Web of Science Core Collection database, and although this approach left little possibility for ignoring some of the documents, this type of literature might have had fewer citations. Furthermore, the bibliometric analysis methods we used can only be applied to general information, rather than full texts. Thus, we might have lost important information that only existed in the full text of the articles, such as the authors' points of view and their prospective opinions of the field.

Conclusion

Our bibliometric analysis should help researchers to understand the trends and public interest in ICB for melanoma. The annual number of publications was rather small, without obvious research trends at the beginning of this century, but has gradually matured in the past 6 years. In the past 2 decades, the United States has contributed the most to this field, followed by China and Germany. The top 3 most productive journals were the *Journal of Clinical Oncology*, *Cancer Immunology*, and *Immunotherapy and Cancer Research*. Cooperation between authors and organizations from different countries needs to be strengthened. In summary, ICB for melanoma is a prolific, fast-growing, and high-profile topic and more research is expected to refine knowledge in this field.

XSL•FO
RenderX

Acknowledgments

This study was supported by grants from the National Natural Science Foundation of China (31800979 and 82102891), the Natural Science Foundation of China for Outstanding Young Scholars (82022060), the National Key Research and Development Program of China (2019YFA0111600 and 2019YFE0120800), the Major International (Regional) Joint Research Program of China (81620108024), the Natural Science Foundation of Hunan Province for outstanding Young Scholars (2019JJ30040), the Natural Science Foundation of Hunan Province of China (2018SK2082 and 2020JJ4884), the Scientific Research Project of Hunan Health and Family Planning Commission (B20180855), and the Youth Science Foundation of Xiangya Hospital (2020Q02).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Comprehensive searching strategy.
[DOCX File , 13 KB - jmir_v24i6e32728_app1.docx]

References

1. Henley SJ, Ward EM, Scott S, Ma J, Anderson RN, Firth AU, et al. Annual report to the nation on the status of cancer, part I: National cancer statistics. Cancer 2020 May 15;126(10):2225-2249 [FREE Full text] [doi: 10.1002/cncr.32802] [Medline: 32162336]
2. Fitzmaurice C. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 2006 to 2016: A systematic analysis for the Global Burden of Disease study. JCO 2018 May 20;36(15_suppl):1568-1568. [doi: 10.1200/jco.2018.36.15_suppl.1568]
3. Skin cancer statistics. World Cancer Research Fund International. URL: https://www.wcrf.org/dietandcancer/cancer-trends/skin-cancer-statistics [accessed 2022-06-14]
4. Robbins HA, Clarke CA, Arron ST, Tatalovich Z, Kahn AR, Hernandez BY, et al. Melanoma Risk and Survival among Organ Transplant Recipients. J Invest Dermatol 2015 Nov;135(11):2657-2665 [FREE Full text] [doi: 10.1038/jid.2015.312] [Medline: 26270022]
5. Shiels MS, Copeland G, Goodman MT, Harrell J, Lynch CF, Pawlish K, et al. Cancer stage at diagnosis in patients infected with the human immunodeficiency virus and transplant recipients. Cancer 2015 Jun 15;121(12):2063-2071 [FREE Full text] [doi: 10.1002/cncr.29324] [Medline: 25739496]
6. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, Australian Pancreatic Cancer Genome Initiative, ICGC Breast Cancer Consortium, ICGC MMML-Seq Consortium, ICGC PedBrain, et al. Signatures of mutational processes in human cancer. Nature 2013 Aug 22;500(7463):415-421 [FREE Full text] [doi: 10.1038/nature12477] [Medline: 23945592]
7. Cooper ID. Bibliometrics basics. J Med Libr Assoc 2015 Oct;103(4):217-218 [FREE Full text] [doi: 10.3163/1536-5050.103.4.013] [Medline: 26512226]
8. Pendlebury DA. White paper: Using bibliometrics in evaluating research. Research Department, Thomson Reuters. URL: https://services.anu.edu.au/files/system/Pendlebury_White_Paper.pdf [accessed 2022-06-13]
9. Taylor BE, McClave SA, Martindale RG, Warren MM, Johnson DR, Braunschweig C, et al. Guidelines for the Provision and Assessment of Nutrition Support Therapy in the Adult Critically Ill Patient. Crit Care Med 2016;44(2):390-438 [FREE Full text] [doi: 10.1097/ccm.0000000000001525]
10. Hodi FS, O'Day SJ, McDermott DF, Weber RW, Sosman JA, Haanen JB, et al. Improved survival with ipilimumab in patients with metastatic melanoma. N Engl J Med 2010 Aug 19;363(8):711-723 [FREE Full text] [doi: 10.1056/NEJMoa1003466] [Medline: 20525992]
11. Topalian SL, Hodi FS, Brahmer JR, Gettinger SN, Smith DC, McDermott DF, et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. N Engl J Med 2012 Jun 28;366(26):2443-2454 [FREE Full text] [doi: 10.1056/NEJMoa1200690] [Medline: 22658127]
12. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. Nat Rev Cancer 2012 Mar 22;12(4):252-264 [FREE Full text] [doi: 10.1038/nrc3239] [Medline: 22437870]
13. Brahmer JR, Tykodi SS, Chow LQ, Hwu W, Topalian SL, Hwu P, et al. Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. N Engl J Med 2012 Jun 28;366(26):2455-2465 [FREE Full text] [doi: 10.1056/NEJMoa1200694] [Medline: 22658128]
14. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, KEYNOTE-024 Investigators. Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. N Engl J Med 2016 Nov 10;375(19):1823-1833. [doi: 10.1056/NEJMoa1606774] [Medline: 27718847]
15. Larkin J, Hodi FS, Wolchok JD. Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma. N Engl J Med 2015 Jul 02;373(1):23-34 [FREE Full text] [doi: 10.1056/NEJMoa1504030] [Medline: 26027431]

XSL•FO
RenderX

16. Tumeh PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJM, Robert L, et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. Nature 2014 Nov 27;515(7528):568-571 [FREE Full text] [doi: 10.1038/nature13954] [Medline: 25428505]
17. Robert C, Long GV, Brady B, Dutriaux C, Maio M, Mortier L, et al. Nivolumab in Previously Untreated Melanoma without Mutation. N Engl J Med 2015 Jan 22;372(4):320-330 [FREE Full text] [doi: 10.1056/nejmoa1412082] [Medline: 25399552]
18. Robert C, Schachter J, Long GV, Arance A, Grob JJ, Mortier L, KEYNOTE-006 investigators. Pembrolizumab versus Ipilimumab in Advanced Melanoma. N Engl J Med 2015 Jun 25;372(26):2521-2532. [doi: 10.1056/NEJMoa1503093] [Medline: 25891173]
19. Gao Y, Shi S, Ma W, Chen J, Cai Y, Ge L, et al. Bibliometric analysis of global research on PD-1 and PD-L1 in the field of cancer. Int Immunopharmacol 2019 Jul;72:374-384. [doi: 10.1016/j.intimp.2019.03.045] [Medline: 31030093]
20. Brunet J, Denizot F, Luciani M, Roux-Dosseto M, Suzan M, Mattei M, et al. A new member of the immunoglobulin superfamily--CTLA-4. Nature 1987 Jul;328(6127):267-270. [doi: 10.1038/328267a0] [Medline: 3496540]
21. Sharma P, Allison JP. The future of immune checkpoint therapy. Science 2015 Apr 03;348(6230):56-61. [doi: 10.1126/science.aaa8172] [Medline: 25838373]
22. Hodi F, Mihm MC, Soiffer RJ, Haluska FG, Butler M, Seiden MV, et al. Biologic activity of cytotoxic T lymphocyte-associated antigen 4 antibody blockade in previously vaccinated metastatic melanoma and ovarian carcinoma patients. Proc Natl Acad Sci U S A 2003 Apr 15;100(8):4712-4717. [doi: 10.1073/pnas.0830997100] [Medline: 12682289]
23. Phan GQ, Yang JC, Sherry RM, Hwu P, Topalian SL, Schwartzentruber DJ, et al. Cancer regression and autoimmunity induced by cytotoxic T lymphocyte-associated antigen 4 blockade in patients with metastatic melanoma. Proc Natl Acad Sci U S A 2003 Jul 08;100(14):8372-8377 [FREE Full text] [doi: 10.1073/pnas.1533209100] [Medline: 12826605]
24. Attia P, Phan GQ, Maker AV, Robinson MR, Quezado MM, Yang JC, et al. Autoimmunity Correlates With Tumor Regression in Patients With Metastatic Melanoma Treated With Anti–Cytotoxic T-Lymphocyte Antigen-4. JCO 2005 Sep 01;23(25):6043-6053 [FREE Full text] [doi: 10.1200/jco.2005.06.205]
25. Sanderson K, Scotland R, Lee P, Liu D, Groshen S, Snively J, et al. Autoimmunity in a phase I trial of a fully human anti-cytotoxic T-lymphocyte antigen-4 monoclonal antibody with multiple melanoma peptides and Montanide ISA 51 for patients with resected stages III and IV melanoma. J Clin Oncol 2005 Feb 01;23(4):741-750. [doi: 10.1200/JCO.2005.01.128] [Medline: 15613700]
26. Comin-Anduix B, Lee Y, Jalil J, Algazi A, de la Rocha P, Camacho LH, et al. Detailed analysis of immunologic effects of the cytotoxic T lymphocyte-associated antigen 4-blocking monoclonal antibody tremelimumab in peripheral blood of patients with melanoma. J Transl Med 2008 May 01;6(1):22 [FREE Full text] [doi: 10.1186/1479-5876-6-22] [Medline: 18452610]
27. Ménard C, Ghiringhelli F, Roux S, Chaput N, Mateus C, Grohmann U, et al. Ctla-4 blockade confers lymphocyte resistance to regulatory T-cells in advanced melanoma: surrogate marker of efficacy of tremelimumab? Clin Cancer Res 2008 Aug 15;14(16):5242-5249. [doi: 10.1158/1078-0432.CCR-07-4797] [Medline: 18698043]
28. Ribas A, Comin-Anduix B, Economou JS, Donahue TR, de la Rocha P, Morris LF, et al. Intratumoral immune cell infiltrates, FoxP3, and indoleamine 2,3-dioxygenase in patients with melanoma undergoing CTLA4 blockade. Clin Cancer Res 2009 Jan 01;15(1):390-399. [doi: 10.1158/1078-0432.CCR-08-0783] [Medline: 19118070]
29. Vonderheide R, LoRusso PM, Khalil M, Gartner EM, Khaira D, Soulieres D, et al. Tremelimumab in combination with exemestane in patients with advanced breast cancer and treatment-associated modulation of inducible costimulator expression on patient T cells. Clin Cancer Res 2010 Jul 01;16(13):3485-3494. [doi: 10.1158/1078-0432.CCR-10-0505] [Medline: 20479064]
30. Lynch TJ, Bondarenko I, Luft A, Serwatowski P, Barlesi F, Chacko R, et al. Ipilimumab in combination with paclitaxel and carboplatin as first-line treatment in stage IIIB/IV non-small-cell lung cancer: results from a randomized, double-blind, multicenter phase II study. J Clin Oncol 2012 Jun 10;30(17):2046-2054. [doi: 10.1200/JCO.2011.38.4032] [Medline: 22547592]
31. Slovin S, Higano C, Hamid O, Tejwani S, Harzstark A, Alumkal J, et al. Ipilimumab alone or in combination with radiotherapy in metastatic castration-resistant prostate cancer: results from an open-label, multicenter phase I/II study. Ann Oncol 2013 Jul;24(7):1813-1821 [FREE Full text] [doi: 10.1093/annonc/mdt107] [Medline: 23535954]
32. Couzin-Frankel J. Breakthrough of the year 2013. Cancer immunotherapy. Science 2013 Dec 20;342(6165):1432-1433. [doi: 10.1126/science.342.6165.1432] [Medline: 24357284]
33. Poole RM. Pembrolizumab: first global approval. Drugs 2014 Oct 21;74(16):1973-1981. [doi: 10.1007/s40265-014-0314-5] [Medline: 25331768]
34. Wang J, Yuan R, Song W, Sun J, Liu D, Li Z. PD-1, PD-L1 (B7-H1) and Tumor-Site Immune Modulation Therapy: The Historical Perspective. J Hematol Oncol 2017 Jan 25;10(1):34 [FREE Full text] [doi: 10.1186/s13045-017-0403-5] [Medline: 28122590]
35. Markham A. Atezolizumab: First Global Approval. Drugs 2016 Aug 13;76(12):1227-1232. [doi: 10.1007/s40265-016-0618-8] [Medline: 27412122]

XSL•FO
RenderX

36. Wolchok JD, Chiarion-Sileni V, Gonzalez R, Grob JJ, Rutkowski P, Lao CD, et al. Long-Term Outcomes With Nivolumab Plus Ipilimumab or Nivolumab Alone Versus Ipilimumab in Patients With Advanced Melanoma. J Clin Oncol 2022 Jan 10;40(2):127-137 [FREE Full text] [doi: 10.1200/JCO.21.02229] [Medline: 34818112]
37. Olson DJ, Eroglu Z, Brockstein B, Poklepovic AS, Bajaj M, Babu S, et al. Pembrolizumab Plus Ipilimumab Following Anti-PD-1/L1 Failure in Melanoma. J Clin Oncol 2021 Aug 20;39(24):2647-2655 [FREE Full text] [doi: 10.1200/jco.21.00079]
38. Pelster MS, Gruschkus SK, Bassett R, Gombos DS, Shephard M, Posada L, et al. Nivolumab and Ipilimumab in Metastatic Uveal Melanoma: Results From a Single-Arm Phase II Study. J Clin Oncol 2021 Feb 20;39(6):599-607 [FREE Full text] [doi: 10.1200/jco.20.00605]
39. Waldman AD, Fritz JM, Lenardo MJ. A guide to cancer immunotherapy: from T cell basic science to clinical practice. Nat Rev Immunol 2020 Nov 20;20(11):651-668 [FREE Full text] [doi: 10.1038/s41577-020-0306-5] [Medline: 32433532]

Abbreviations

CTLA-4: T-lymphocyte associated protein-4
FDA: US Food and Drug Administration
ICB: immune checkpoint blockades
IF: impact factor
PD-1: programmed death receptor 1
PD-L1: ligands of programmed death receptor 1

Edited by A Mavragani; submitted 07.08.21; peer-reviewed by I Zachary, S Azadnajafabad, P Zarogoulidis; comments to author 23.10.21; revised version received 13.12.21; accepted 26.03.22; published 27.06.22.

Please cite as:
Xu Y, Jiang Z, Kuang X, Chen X, Liu H
Research Trends in Immune Checkpoint Blockade for Melanoma: Visualization and Bibliometric Analysis
J Med Internet Res 2022;24(6):e32728
URL: https://www.jmir.org/2022/6/e32728
doi:10.2196/32728
PMID:35759331

©Yantao Xu, Zixi Jiang, Xinwei Kuang, Xiang Chen, Hong Liu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.06.2022. 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

Establishing Institutional Scores With the Rigor and Transparency Index: Large-scale Analysis of Scientific Reporting Quality

Joe Menke[1,2], BSc; Peter Eckmann[2,3]; Ibrahim Burak Ozyurt[2,3], PhD; Martijn Roelandse[4], PhD; Nathan Anderson[2], BSc; Jeffrey Grethe[2,3], PhD; Anthony Gamst[5], PhD; Anita Bandrowski[2,3], PhD

[1]Center for Research in Biological Systems, University of California, San Diego, La Jolla, CA, United States

[2]SciCrunch Inc., San Diego, CA, United States

[3]Department of Neuroscience, University of California, San Diego, La Jolla, CA, United States

[4]Martijnroelandse.dev, Ouderkerk aan de Amstel, Netherlands

[5]Department of Mathematics, University of California, San Diego, CA, United States

Corresponding Author:
Anita Bandrowski, PhD
Department of Neuroscience
University of California, San Diego
9500 Gilman Drive
La Jolla, CA, 92093
United States
Phone: 1 858 822 3629
Email: abandrowski@health.ucsd.edu

Abstract

Background: Improving rigor and transparency measures should lead to improvements in reproducibility across the scientific literature; however, the assessment of measures of transparency tends to be very difficult if performed manually.

Objective: This study addresses the enhancement of the Rigor and Transparency Index (RTI, version 2.0), which attempts to automatically assess the rigor and transparency of journals, institutions, and countries using manuscripts scored on criteria found in reproducibility guidelines (eg, Materials Design, Analysis, and Reporting checklist criteria).

Methods: The RTI tracks 27 entity types using natural language processing techniques such as Bidirectional Long Short-term Memory Conditional Random Field–based models and regular expressions; this allowed us to assess over 2 million papers accessed through PubMed Central.

Results: Between 1997 and 2020 (where data were readily available in our data set), rigor and transparency measures showed general improvement (RTI 2.29 to 4.13), suggesting that authors are taking the need for improved reporting seriously. The top-scoring journals in 2020 were the *Journal of Neurochemistry* (6.23), *British Journal of Pharmacology* (6.07), and *Nature Neuroscience* (5.93). We extracted the institution and country of origin from the author affiliations to expand our analysis beyond journals. Among institutions publishing >1000 papers in 2020 (in the PubMed Central open access set), Capital Medical University (4.75), Yonsei University (4.58), and University of Copenhagen (4.53) were the top performers in terms of RTI. In country-level performance, we found that Ethiopia and Norway consistently topped the RTI charts of countries with 100 or more papers per year. In addition, we tested our assumption that the RTI may serve as a reliable proxy for scientific replicability (ie, a high RTI represents papers containing sufficient information for replication efforts). Using work by the Reproducibility Project: Cancer Biology, we determined that replication papers (RTI 7.61, SD 0.78) scored significantly higher (P<.001) than the original papers (RTI 3.39, SD 1.12), which according to the project required additional information from authors to begin replication efforts.

Conclusions: These results align with our view that RTI may serve as a reliable proxy for scientific replicability. Unfortunately, RTI measures for journals, institutions, and countries fall short of the replicated paper average. If we consider the RTI of these replication studies as a target for future manuscripts, more work will be needed to ensure that the average manuscript contains sufficient information for replication attempts.

(J Med Internet Res 2022;24(6):e37324) doi:10.2196/37324

XSL•FO
RenderX

KEYWORDS

research reproducibility; rigor and transparency; reproducibility crisis; reporting transparency; science of science; research metric; data and code availability; cell line authentication; university ranking

Introduction

Background

Research reproducibility is necessary for scientific progress. However, over the last decade, numerous reports on research irreproducibility have shed light on a lingering problem, one that is proving to be both troublesome and costly [1-5]. Ioannidis [1] and the Open Science Psychology collaboration examined the issue from a statistical point of view, arguing that multiple comparisons that are not necessarily reported affect the published literature. Begley and Ellis [2] described an account in which their teams attempted to reproduce key cancer studies and were largely unable to do so; however, they did not share their data. The Center for Open Science recently published a series of papers summarized by Errington et al [6], which describe an open replication attempt that had similar findings to the work by Begely and Ellis [2]. Vasilevsky et al [4] clearly showed that approximately half of the reagents in papers cannot be tracked down, whereas Freedman et al [7] attempted to visualize the economic impact of irreproducibility.

Fortunately, many stakeholders responded to address these issues. Funders such as the National Institutes of Health (NIH), the largest public source of health research funding worldwide [8], have made significant efforts across multiple fronts. The NIH advanced open publication efforts with the creation of PubMed Central. In terms of guidelines, the NIH gathered copious stakeholder feedback and designed and implemented rigor and reproducibility guidelines (adapted from the study by Landis et al [9]). The NIH also rewrote their instructions to grantees, released numerous training modules and webinars, and implemented a data sharing policy to improve the reproducibility of funded research [10,11]. Even some private funders such as the Gates Foundation have begun requiring their funded research (both the manuscript and its data) to immediately become open access once published [7].

Journals and publishers have also responded to this. In an effort to encourage reproducibility, numerous scientific organizations and journals have adopted the Transparency and Openness Promotion guidelines, which focus on establishing best practices at the level of individual journals [12]. Similarly, the publisher-driven Materials Design, Analysis, and Reporting framework is a multidisciplinary research framework designed to improve reporting transparency across life science research at the level of individual manuscripts [13]. This framework provides a consistent, minimum reporting checklist whose criteria were used in part to create the first Rigor and Transparency Index (RTI), a journal quality metric focusing on research methodologies and transparency in reporting [14]. Because of the RTI, journals can be compared using a range of criteria that impact reproducibility, providing a proxy for research quality and a strong incentive for improvement.

Unfortunately, these types of indicators and incentives do not exist for all stakeholders. Research institutions, in particular, have few options for determining whether investigators will follow the guidelines. In fact, there is no simple way to see a university's corpus, let alone to estimate its quality. Despite previously contributing to the *Reproducibility Crisis* [15], institutional output is still difficult to track and measure. Various systems for ranking faculty are in place at institutions, including counting publications, counting citations, and counting *high impact* publications; however, issues have been reported when using the impact factor for these purposes [16,17]. Some institutions started leaning more heavily on assessments of open science [18], which reduced the reliance on paper counting or on the impact of particular journals. Indeed, tying researcher assessment to any single factor, even if that happens implicitly by reviewers looking for recognizable journal names, may place inappropriate pressure on scientists to focus on strategies that increase research notoriety rather than quality, which can have wider implications [19,20].

After receiving feedback from several stakeholders [21,22], we developed a new version of SciScore, an automated natural language processing tool suite that detects transparency criteria and research resources within individual papers. In conjunction with this, we linked published manuscripts with their disambiguated research institutions. Here, we introduce the latest version of the RTI, version 2.0, which represents the mean SciScore over a subset of papers and demonstrates how it can be used to assess reporting transparency within research institutions. The fact that the MacLeod laboratory is endeavoring to register a report assessing institutions on similar metrics (MacLeod personal communication) suggests the importance of assessing based on quality rather than citations alone.

Objectives

The overall aim of this study was to establish a scientific reporting quality metric across institutions and countries and to highlight the need for high-quality reporting to ensure replicability within biomedicine, using manuscripts from the PubMed Central Open Access Initiative and the Reproducibility Project: Cancer Biology [6].

Methods

Individual Manuscript Processing

Overview

Individual manuscripts were processed using the latest version of SciScore (research resource identifier [RRID]: SCR_016251). SciScore uses multiple conditional random field (CRF)-based models [23] in combination with regular expression patterns for named entity recognition. For more information on the core features used within CRF models, please see our previous work on the Resource Disambiguator for the Web, which used the same framework [24]. SciScore classifiers currently recognize 27 entity types. New entity types include field sample permits, general euthanasia statements, inclusion and exclusion criteria, attrition, general replication statements, number of replications,

XSL•FO
RenderX

type of replication, age, weight, code availability, data availability, and statistical tests. Table S1 in Multimedia Appendix 1 provides a full list of entity types and their descriptions.

Classifiers were validated using precision, recall, and their harmonic mean (F_1). Their initial performances were calculated using 10 random splits of the human-curated data, where 90% was used for training and 10% for testing; each performance score was the average of all 10 training trials. Classifier performances are listed by entity type in Table S2 in Multimedia Appendix 2. The study by Menke et al [14] provides the full description of how the data sets were labeled and how the classifiers were trained and tested. In addition to its CRF-based classifiers, SciScore has begun to implement regular expressions for detecting protocols, data, and code identifiers. Regular expression pattern sets were initially adapted from the identifier patterns listed by [25] (RRID: SCR_003735). These sets were then adjusted and supplemented accordingly. These patterns are listed in Multimedia Appendix 3.

In addition, enhanced table detection and tabular data extraction within SciScore were performed using neural network models. More specifically, table and section header boundary detection and subsequent table row detection in the provided free text were performed with feedforward neural networks using a sliding context window approach.

New Criteria and Scoring Framework

Of the new criteria added (ie, field sample permits, general euthanasia statements, euthanasia agents, inclusion and exclusion criteria, attrition, general replication statements, number of replications, type of replication, age, weight, protocol identifiers, code availability, code identifiers, data availability, data identifiers, and statistical tests), the vast majority have been implemented in RTI, version 2.0. When creating the manually checked data sets, we grouped euthanasia and euthanasia agents to align with the output of the automated pipeline. Some criteria presented in SciScore's output, namely oligonucleotides and statistical tests, were also omitted in terms of scoring, where we continued to refine their natural language processing algorithms.

The scoring framework was previously described in our study using RTI, version 1.0 [14]. To summarize the key findings, research papers were scored on a 10-point scale, where a maximum of 5 points was derived from the manuscript's rigor adherence and another 5 points from its key resource identification performance. A comparison of the total number of identified criteria with the total number of expected criteria provided the rigor adherence score. Please note that currently, code availability, data availability, and the various identifiers (protocol, code, and data) do not yet affect scoring (ie, they do not contribute to found or expected tallies). This will be addressed in future studies.

Following a similar found-to-expected scoring system, key resource identification performance is calculated by comparing the number of uniquely identifiable resources found (ie, those with RRIDs or RRID suggestions) to the total number of resources detected. If no resources or criteria were found or if the only criteria found does not impact scoring (code availability, data availability, protocol identifiers, code identifiers, data identifiers, statistical tests, and oligonucleotides), then the paper was scored as a 0 and was considered *not applicable*. Papers with a score of 0 were excluded from the data set because there was no way to determine if scoring was appropriate.

Other than the addition of new criteria, the only key scoring change between RTI, version 1.0, and RTI, version 2.0, was the inclusion of more conditional scoring logic within the rigor adherence section. In RTI, version 1.0, the only conditional scoring logic being implemented involved cell line authentication, which was only expected when a cell line was detected in the manuscript. In RTI, version 2.0, an additional scoring logic was included. This logic is outlined in Table 1. As an example, if a criterion was found in the ethics-1 grouping (IACUC, IRB, or consent), the model would expect at least one of the group selection criteria (inclusion and exclusion criteria or attrition), sex, at least one of the demographic criteria (age or weight), randomization, blinding, and power analysis. If a manuscript contained an IACUC and age but no other criteria, the model would detect 2 out of 6 expected criteria, which translates roughly to a 2 out of a maximum 5 points for this section. As another example, if euthanasia was detected, we would expect Institutional Animal Care and Use Committee, Institutional Review Board, or consent.

XSL•FO
RenderX

Table 1. Conditional scoring groupings and logic for rigor adherence section.

Grouping	Criteria included	If this grouping is detected, what is expected?	This grouping is expected when what is detected?
Ethics-1	Institutional Animal Care and Use Committee, Institutional Review Board, and consent	Group selection, sex, demographics, random, blinding, and power	Euthanasia
Ethics-2	Field sample permit	Random, blinding, and power	Never expected
Euthanasia	Euthanasia statement and euthanasia agent	Ethics-1, group selection, sex, demographics, random, blinding, and power	Never expected
Group selection	Inclusion and exclusion criteria and attrition	Random, blinding, and power	Ethics-1 and euthanasia
Sex	Sex	Random, blinding, and power	Ethics-1, euthanasia, and demographics
Demographics	Age and weight	Sex, random, blinding, and power	Ethics-1 and euthanasia
Random	Random	Blinding and power	Always expected
Blinding	Blinding	Random and power	Always expected
Power	Power analysis	Random and blinding	Always expected
Replication	Replication statement, number of replications, and type of replication	Random, blinding, and power	Never expected
Cell line authentication	Cell line authentication and cell line contamination	Sex, random, blinding, and power	Cell lines
Methods and materials availability[a]	Data availability, data identifiers, code availability, code identifiers, and protocol identifiers	Never expected; do not affect score	Never expected; do not affect score
Cell lines	Cell lines	Cell line authentication	Never expected
Other resources[b]	Antibodies, organisms, plasmids, and tools	Never expected; only affects resource transparency score	Never expected; only affects resource transparency score
Miscellaneous[a]	Oligonucleotides, statistical tests, and incorrect research resource identifiers	Never expected; does not affect either score	Never expected; does not affect either score

[a]Row indicates criteria that do not affect any score.

[b]Row indicates criteria that do not affect the rigor adherence score, only the resource transparency score.

Validation

Although some entity types have been previously tested (cell lines in the study by Babic et al [26] and multiple types in the study by Menke et al [14]), other entity types and regular expression patterns have not yet been thoroughly validated on complete articles outside of training sets. To remedy this, we tested the performance of our models using 423 papers that were previously selected at random for manual curation during testing using RTI, version 1.0. Originally, 2 sets of 250 papers were randomly selected based on their score during the first run in November 2019 (SciScore>0: 250 papers; SciScore=0: 250 papers). We used these hand-curated papers as the gold standard to retest performance during testing by RTI, version 2.0, to ensure that *not applicable* papers were out of scope and to analyze performance on scored papers. Consistent with our previous methods, if both the curator and the classifier agreed regarding the presence or absence of an entity type, then we assumed that the answer was correct and looked no further. Disagreements, in contrast, were classified as false negatives or false positives, with the assumption that the curator is always correct. False negatives occurred when the classifier noted an entity type as missing when it was really present. False positives occurred when the classifier incorrectly noted an entity type as being present when it was missing.

For testing *not applicable* papers (SciScore=0), a curator (NA) went through 232 the previously *not applicable* papers to determine whether each paper was still expected to be scored as a 0 even after the addition of new entity types. From the original 250 papers, 18 (7.2%) papers were removed because they were previously determined to have either no clear methods section (highly theoretical papers, editorials, etc) or contained only supplemental PDFs, which are effectively invisible to our models [14]. Of these 232 papers, 173 (74.6%) were hand scored as 0 and represented papers we expected to still be *not applicable*. We compared each classifier's output against our curator's for these 173 papers. A total of 87.9% (152/173) of the papers scored as expected (SciScore=0), and 12.1% (21/173) of the papers contained false positives across the various entity types. Entity types with multiple false positives included attrition (7/21, 33%), randomization (4/21, 19%), field sample permit (3/21, 14%), software tools (3/21, 14%), weight (3/21, 14%), and age (2/21, 10%).

For testing the scored papers, another set containing 250 papers (SciScore >0) was hand curated without exception. Hand-curated

XSL•FO
RenderX

data from our first run were supplemented with data for our new criteria, except for statistical tests, which were not tracked (similar to oligonucleotides and plasmids). In all, 2 curators (NA and JM) went through these papers and were blinded to our models' outputs (50 papers for NA and 200 papers for JM). This information was again compared with our classifiers' performances; the results of this analysis are shown in Table 2. All entity types had curator-classifier agreement rates >80%; many were >90%. As in our previous analysis, the overall agreement represents the additive probability for instances where multiple resources were mentioned. In all the cases, the agreement rate was measured above the raw classifier F_1 rate.

Overall, there was no significant decline in performance across the criteria featured in either version; any difference in scoring resulted from the addition of new training data or enhanced conditional scoring. As a result of these analyses, we did not seek to further tune the parameters.

Table 2. Rates of false negatives, false positives, and overall agreement based on manual analysis of 250 scored papers (SciScore >0) from our data set.

Entity type	False positives	False negatives	Overall agreement
	Size and rate, n (%)	Size and rate, n (%)	Size and rate, (agreed, n) (%)
Rigor criteria			
Institutional review board statement	14 (5.6)	11 (4.4)	225 (90)
Consent statement	1 (0.4)	11 (4.4)	238 (95.2)
Institutional animal care and use committee statement	2 (0.8)	17 (6.8)	231 (92.4)
Field sample permit	19 (7.6)	0 (0)	231 (92.4)
Euthanasia	6 (2.4)	7 (2.8)	237 (94.8)
Inclusion and exclusion criteria	10 (4)	17 (6.8)	223 (89.2)
Attrition	35 (14)	7 (2.8)	208 (83.2)
Type of replication	0 (0)	3 (1.2)	247 (98.8)
Number of replications	17 (6.8)	16 (6.4)	217 (86.8)
General replication	13 (5.2)	16 (6.4)	221 (88.4)
Randomization of participants into groups	20 (8)	4 (1.6)	226 (90.4)
Blinding of investigator or analysis	5 (2)	5 (2)	240 (96)
Power analysis for group size	12 (4.8)	4 (1.6)	234 (93.6)
Sex as a biological variable	6 (2.4)	21 (8.4)	223 (89.2)
Age	5 (2)	44 (17.6)	201 (80.4)
Weight	6 (2.4)	22 (8.8)	222 (88.8)
Cell line authentication	15 (6)	1 (0.4)	234 (93.6)
Cell line contamination check	0 (0)	0 (0)	250 (100)
Protocol identifiers	3 (1.2)	2 (0.8)	245 (98)
Code availability	4 (1.6)	1 (0.4)	245 (98)
Code identifiers	0 (0)	2 (0.8)	248 (99.2)
Data availability	24 (9.6)	0 (0)	226 (90.4)
Data identifiers	27 (10.8)	3 (1.2)	220 (88)
Key biological resources			
Antibody	2 (0.8)	5 (2)	243 (97.2)
Organism	3 (1.2)	7 (2.8)	240 (96)
Cell line	6 (2.4)	4 (1.6)	240 (96)
Software project and tools	8 (3.2)	38 (15.2)	204 (81.6)

XSL•FO
RenderX

Text Mining the Open Access Subset of PubMed Central

Overview

We downloaded and processed all PubMed Central (PMC; RRID: SCR_004166) articles whose full text was available in the PMC Open Archives Initiative (OAI) data set starting April 2021 (processing took approximately 2 months). The PMC-OAI data set was initially downloaded as multiple directories (1 per journal), containing articles available for text mining. These directories were consolidated into 4 shards (or parts), depending on the number of manuscripts available within each journal. Each shard was then processed using the proposed models. Consistent with our previous RTI study, abstract-only articles and articles without methods sections were excluded [14]. Similarly, articles only available as PDFs were not included within the open access (OA) subset, and, as such, were excluded from our analysis. We included data from journals, institutions, and countries that had published >10 papers per year. This information is available in the Multimedia Appendix 4. We limited our analyses to journals, institutions, and countries that had published >10 papers per category, such as year, if the data were only differentiated by year (eg, all by country vs all by country by year). We obtained data from 2,153,877 manuscripts representing 9398 journals, 37,648 research institutions, and 200 countries (based on research institution metadata in the Research Organization Registry [ROR]).

Deduplication and Disambiguation of Research Institutions

We sought to disambiguate the authors' affiliation strings using the standardized set of institutions listed in the ROR [27]. ROR provides unique identifiers and metadata for many institutions worldwide.

The ROR has developed an application programming interface (API) to search for and retrieve information from its registry. It is able to make a best guess at the institution identifier given an input affiliation string using a combination of substring searches, fuzzy word comparisons, and hard-coded heuristics. Although their API is offered as a web service, initial tests raised concerns of rate-limiting and slow response times for a large volume of requests. However, a developer version was obtained from the ROR [28], which allowed us to run an API instance on a local machine and avoid network concerns. We used the API end point organizations?affiliation= for disambiguating affiliation strings. For each query, a confidence score was provided along with a binary match or no match field. Almost all queries returned a best guess institution from the ROR, although the API did not declare confidence for all queries. We recorded all guesses in our database, whether the API was confident, the confidence using the chosen field of the API response, and the time that the API took on our local machine.

We also developed our own tool for disambiguating affiliations (available on GitHub [29]). We used a regular expression (Figure 1) to extract an institution's name from each affiliation string. Affiliation strings were split on all semicolons, regardless of length, to capture cases in which multiple affiliations were present in a single string. In these cases, a single paper could be included in the counts of multiple institutions (eg, *UCSD*; *UCLA*) where each institution's paper count would each increase by 1 or be counted multiple times by a single institution (eg, *Department of Computer Science, UCSD; Department of Biological Sciences, UCSD*) where the paper would be included twice in UCSD's count. ROR data were loaded onto a PostgreSQL instance, and institution names were stored in a *tsvector* column for fast lookup of the regular expression-extracted institution name. The workflow is illustrated in Figure 1.

To compare the performance of our tool with that of ROR's, 2 curators (JM and PE) matched 200 affiliation strings from a simple random sample of all affiliations from our PMC set (100 per curator) to institutions contained within the ROR database. For cases in which curators could not locate a matching ROR institution, the affiliation string was left blank. A total of 186 strings were matched to ROR institutions. The accuracy was calculated for each tool. Accuracy was defined as the percentage of institutions in which the result of the tool was equal to the result from the hand-curated set. Only when the tool and hand-curated set agreed exactly (ie, either both reported no matching ROR IDs or both reported the exact same ROR ID) was an accurate match declared. Calculations were performed for 2 cases: high confidence matches only and all matches (high and low confidence). The results of this comparison are shown in Table 3.

As shown in Table 3, both algorithms performed similarly in terms of accuracy. Our in-house tool's speed greatly differentiated itself from ROR's. As a result of this analysis, we elected to use our in-house tool over ROR's for institutional disambiguation.

XSL•FO
RenderX

Figure 1. Disambiguation of affiliation strings workflow. ROR: Research Organization Registry; regex[A]: exact regular expression.

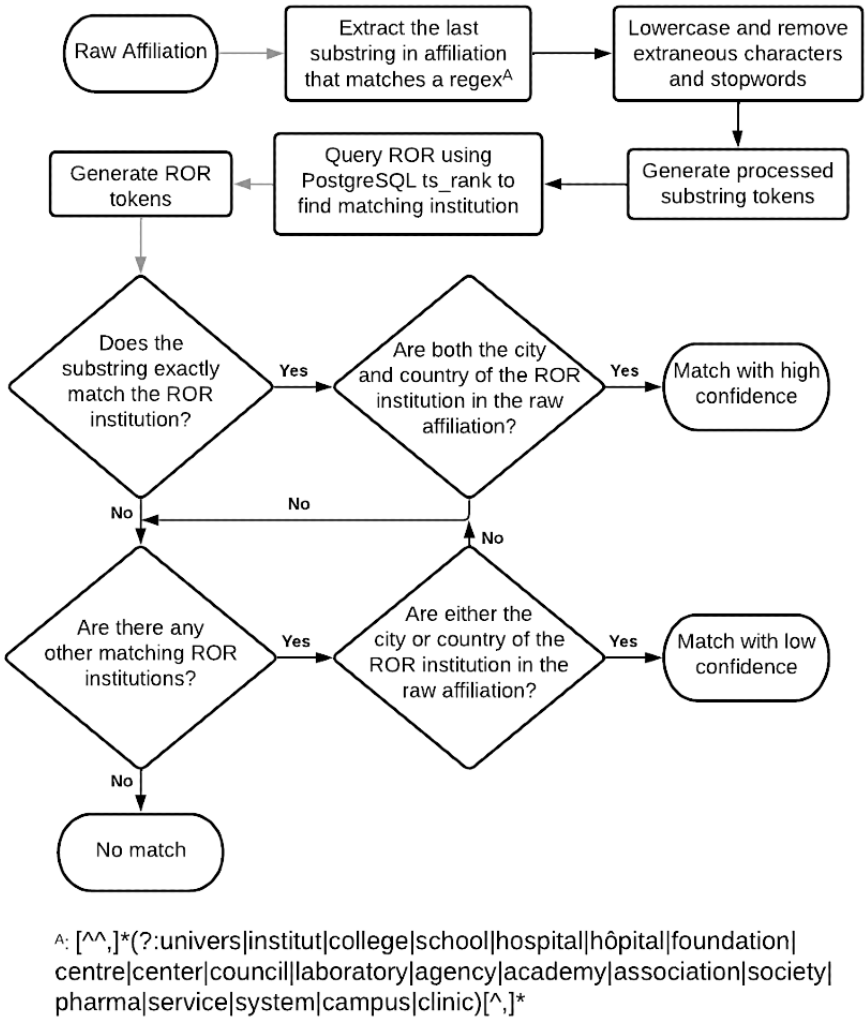


Table 3. Affiliation to institution matching: in-house tool compared with the Research Organization Registry (ROR) application programming interface and a human-curated set of 200 affiliations.

Confidence	Time per affiliation (ms)		Accuracy	
	In-house	ROR	In-house	ROR
High only	1.759	400.90	0.5323	0.6666
High and low	9.745	400.90	0.7043	0.7043

Department Identification and Grouping

To account for differing reporting standards and expectations across fields, we sought to measure how semantically similar papers are. Specifically, we used abstract similarity measures to group departments of major UK research institutions, so we could compare departments to their analogs at other institutions.

All affiliation strings that contained the strings *United Kingdom*, *Scotland*, *Wales*, or *England* were included. The following regex was used for extracting department names from the affiliation strings:

[^^,]*(?:department|centre|center|section|division|institute|institution|program|school|museum|group)[^,]*

Unwanted characters at the beginning of each affiliation string were removed according to the regex ^[^A-Z]*, and the surrounding whitespace was stripped. All affiliation data along with corresponding PubMed Identifiers were stored in a PostgreSQL table.

For judging semantic similarity across papers, we used the averaged word vectors (normalized by L2) of the abstracts. First, abstracts were extracted from the PMC XML data dumps (all data available before December 12, 2020), excluding articles with a publication type of *Comment*, *Published Erratum*, *Review*, or *Preprint*. Abstract text was stored in the PostgreSQL table along with the PubMed Identifier. Then, a random sample of 1% of all abstracts in the database was used to train fastText [30] word embeddings with default hyperparameters and dimensionality of 300. Then, for each abstract in the table, fastText's *getSentenceVector* function was used to determine the averaged L2 normalized word vector for each abstract, and the result was stored as a vector in the PostgreSQL table.

XSL•FO
RenderX

To cluster departments based on this similarity measure, we first found the average abstract vector for departments with >200 papers. This was a simple mean of all abstract vectors with an identical department name, previously described, and top-level institution as determined by our in-house disambiguator. Then, using t-SNE as implemented by scikit-learn ([31]; RRID: SCR_002577) with a perplexity of 7, we reduced 300 dimensions to 2 so that similarities between departments could be visualized. Finally, we used scikit-learn k-means clustering on the reduced data to identify 10 clusters of similar departments. To compare the RTIs across departments in each cluster, we found the RTIs across all papers in a given department and ranked departments based on the RTIs within each cluster.

Statistics

Journals, institutions, and countries were only included in our analyses if more than 10 papers were scored per year unless stated otherwise.

For SciScore named entity classifiers and disambiguation algorithms, we used the standard measures to quantify performance: recall (R), precision (P), and the harmonic mean of R and P (F_1). These values were determined using the following formulas:

(1)

(2)

(3)

False negatives are criteria that were missed by our models but were labeled by a human curator and false positives were incorrectly identified as an entity by our models.

The partial correlation coefficient was calculated using Spearman rank-order correlation coefficient using the following equation:

(4)

where Y_{ABC} is the correlation between A and B adjusted for C.

Ethics Approval

We did not obtain institutional review board approval to conduct this study, as we did not use any human or animal participants, thus making this study exempt.

Results

Overview

Using our institutional disambiguation model, we obtained data from 2,153,877 articles from 9398 unique journals representing 37,648 institutions across 200 countries. Of these articles, 1,971,824 (91.55%) contained rigor and transparency criteria (SciScore>0; RTI 3.99). The remaining 182,053 (8.45%) articles contained no mention of such criteria (SciScore=0; not applicable). As a result, we did not include these articles in our primary analyses; they did not contain a methods section or were out of scope [14]. We were able to confidently match 1,947,966 articles to 37,067 distinct institutions across 200 countries, where SciScore>0. The RTI data are available in Multimedia Appendix 5.

Criteria Trends Over Time

We determined the proportion of papers that addressed individual rigor criteria within the PMC-OAI subset. Data for RTI, version 1.0, represent PMC-OAI manuscripts published between 1997 and 2019. RTI, version 2.0, data are from the PMC-OAI manuscripts published between 1997 and 2020. Both the metrics steadily rise over time, although there is relatively little difference between RTI, version 1.0, and RTI, version 2.0, in terms of their RTIs. As shown in Figure 2, RTI has steadily increased over the last two decades, showing improved levels of transparency within machine-accessible PMC manuscripts. Out of the rigor criteria shown in Figure 3, author addressment of randomization increased the most between 1997 and 2020 (12% to 31%). Blinding (3% to 9%), power analysis (1% to 8%), and replication addressment (24% to 27%), all improved over this timeframe as well. Even at their maximum, blinding and power analysis were addressed in <10% of the studies. Replication addressment represents the percentage of papers that mention replication, number of replications, or type of replication. Figure 4 shows the data, code, and protocol presence across all the papers, regardless of score. Here, we considered a paper to address data presence if the paper had a data availability statement (eg, *all data used within this study is available in the supplementary methods* or *data is available upon request*) or a data identifier (ie, common accession number patterns in data repositories). Code accessibility was determined in a similar manner. We note that this is a conservative estimate of data and code accessibility, as we only checked the methods and materials sections, and some journals place these in a section completely separate from the materials and methods, whereas others use the references section. In addition, we were unable to check if identifiers actually exist owing to slow resolver resolution or if data or code is actually present in the supplementary files. Data addressment (5% to 17%), code addressment (0% to 3%), and the number of protocols cited (0 to 946 papers), all increased between 1997 and 2020.

In Figure 5, when looking at criteria commonly associated with cell line reporting standards (sex, cell line authentication, and contamination), we limited our analysis to papers containing at least one cell line and no IRB or IACUC, as detected by our models. As shown, the number of papers using cell lines continues to grow (470 to 21,854). Within this set, sex did not improve (14% to 13%), whereas the reporting of both cell line authentication (6% to 8%) and contamination (1% to 8%) increased but remained at relatively low levels. As shown in Figure 6, studies containing at least one organism were used to inform our analysis of the organism's demographic reporting rates. Reporting rates for sex (40% to 65%), age (31% to 54%), and weight (3% to 15%) improved steadily across the board.

XSL•FO
RenderX

Figure 2. Average score for Rigor and Transparency Index (RTI), version 1.0 (1997-2019) and version 2.0 (1997-2020). PubMed Central- Open Archives Initiative steadily increases over time. Differences between versions are negligible.

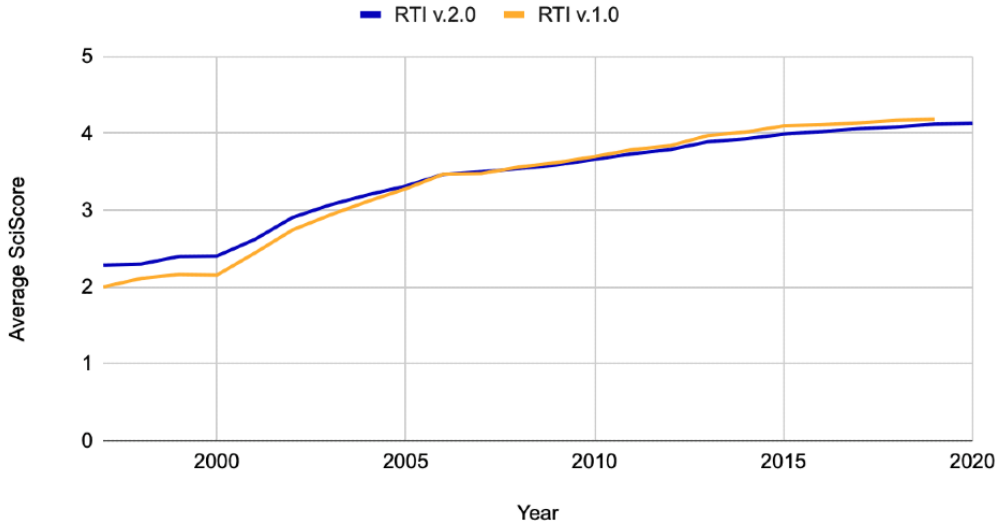


Figure 3. Proportion of papers addressing various bias limiting criteria (ie, blinding, randomization, power, and replication) across all scored papers (1997-2020).

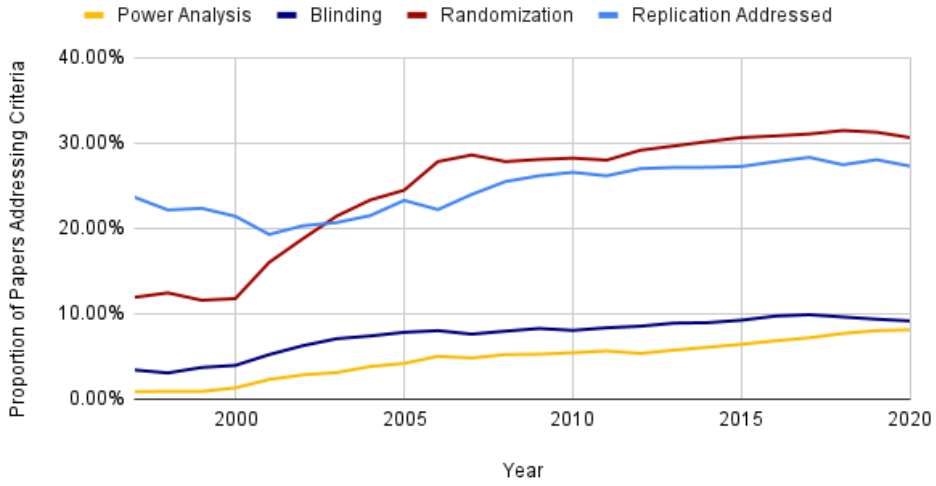
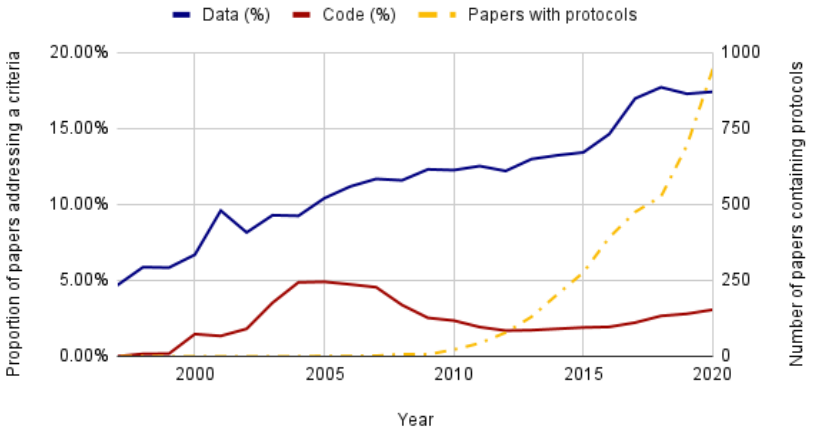


Figure 4. Data, code, and protocol addressment across all papers (1997-2020).



XSL•FO
RenderX

Figure 5. Data shown from 1997 to 2020. Left axis shows the percentage of papers containing cell lines that authenticate, check cells for contamination, and include sex. Right axis shows the number of papers using cell lines each year.

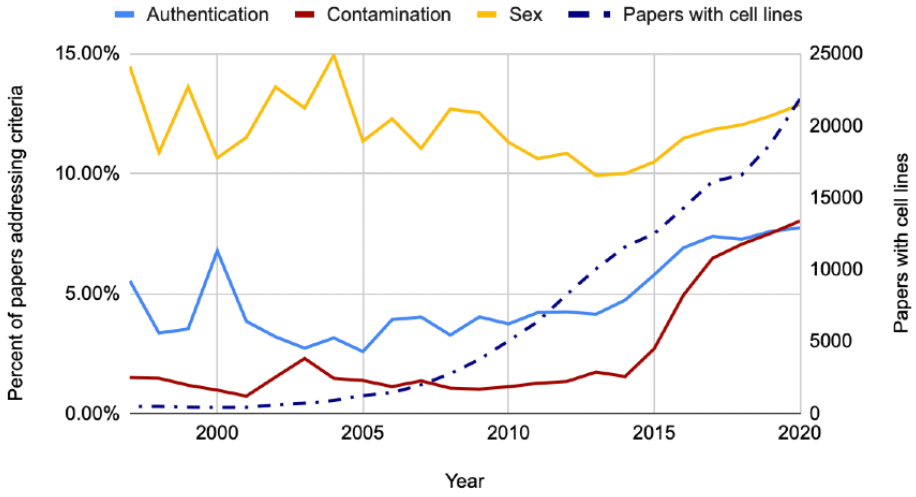
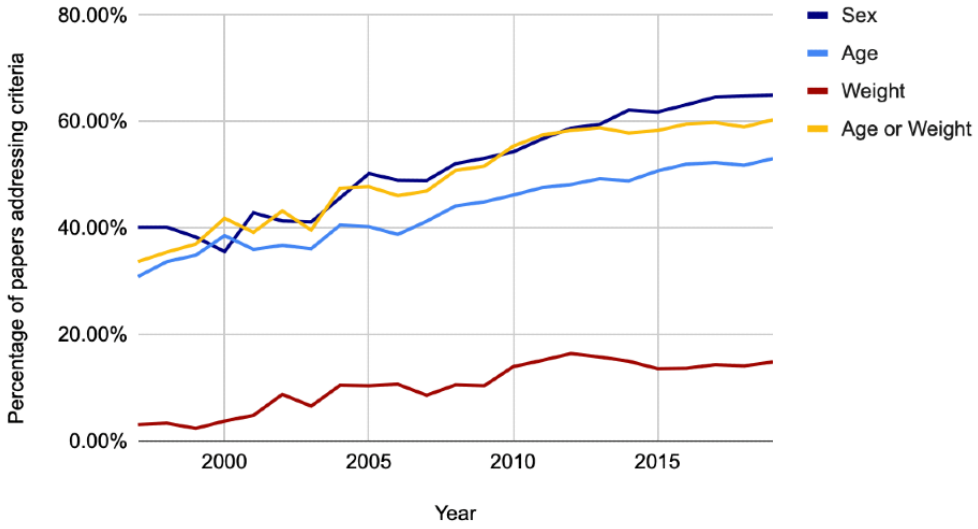


Figure 6. Data from 1997 to 2020. Percentage of papers describing demographic information (sex, age, or weight) that contain at least one transgenic organism.



Criteria Across Journals, Research Institutions, and Countries

Among the journals with >10 papers scored in 2020, the top performer in RTI was the *Journal of Neurochemistry* (RTI 6.24). Of the journals with >1000 papers scored in 2020, a total of 2 journals were tied for the lead in RTI, medicine, and nutrients (RTI 5.02). For reference, the RTI across all the papers scored in 2020 was 4.13. Further information on journal performance and journal performance by year is available in Multimedia Appendix 5.

The data in Figure 7 represent 186,045 OA papers published in 2020. The 2 countries with the greatest number of institutions, represented in Figure 7, were China (8/25, 32%) and the United States (5/25, 20%). Many other countries had either 1 or 2 institutions represented. Among individual institutions, Capital Medical University (n=10,125) had the highest RTI (4.75).

We were able to successfully match our institutional data (for institutions with ≥100 papers in 2013) to the names of 110 institutions listed in the data set used by Lepori et al [32] in 2019 to compare university revenues with their publication and citation counts. For the 110 matched institutions, Table 4 shows the correlation calculations between the 3 variables (all from 2013): total number of academic staffs, current total revenue, and RTI. As expected, there was a positive correlation (0.62) between the total number of academic staff and the current total revenue, which makes sense—as staff grows, so do costs. We also performed a partial correlation analysis between the total revenue and RTI, correcting for the total number of academic staffs. This shows that there is a weak negative relationship between an institution's total revenue and its RTI, although the correlation coefficient (−0.12) suggests that this is not significant. Correlation values were calculated using Spearman rank-order correlation coefficient.

XSL•FO
RenderX

Figure 7. Analysis of Rigor and Transparency Index (RTI) across research institutions in 2020. The left axis represents the RTI. The 50 institutions with the most papers published in 2020 were ranked according to their RTI. The 25 institutions with the highest RTI are shown.

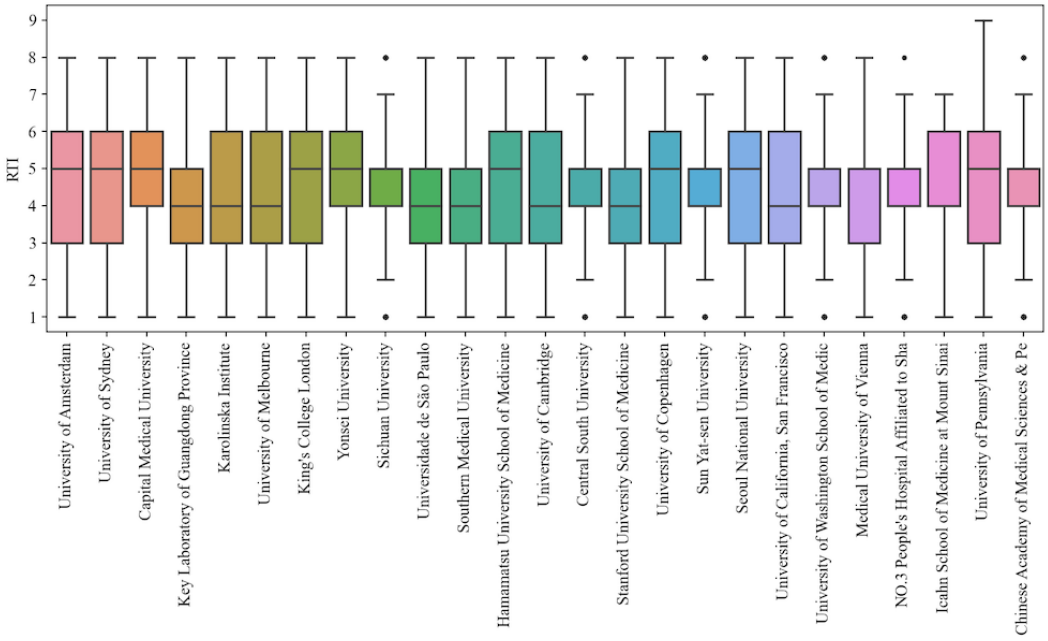


Table 4. Spearman rank-order correlation coefficient calculations between the number of academic staffs, the total revenue, and Rigor and Transparency Index (RTI).[a] The partial correlation coefficient between revenue and RTI was calculated to be -0.1154.

	Total academic staff	Current total revenue	RTI
Total academic staff	1	N/A[b]	N/A
Total current revenue	0.6208	1	N/A
RTI	−0.1209	−0.1648	1

[a]Data from 2013. A partial correlation was calculated between total revenue and RTI correcting for the number of academic staffs.

[b]N/A: not applicable.

Department Identification and Grouping

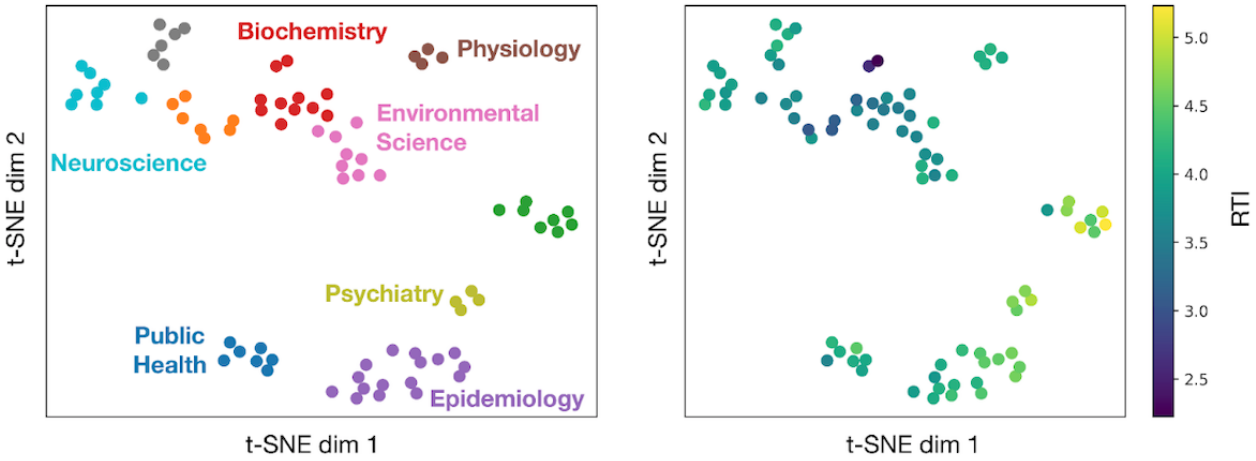
Institutional departments should be compared directly to meaningfully compare institutions at more granular levels, as reporting requirements and standards vary across fields. Therefore, we advise against interfield comparisons for this reason. We grouped the largest 80 UK departments by paper count, using the semantic similarities of their abstracts. Following the procedure described in Section 2.2, we computed a t-distributed stochastic neighbor embedding intraplate of abstract vectors across departments and then performed k-means clustering to generate discrete clusters. We visualized each department's RTI to allow intracluster comparisons (Figure 8). As shown in Figure 8, there are large differences between the RTIs of different fields; for example, the papers of chemistry departments tend to have lower RTIs than psychiatry departments. Therefore, such a clustering is necessary for a fair departmental comparison. We note that departments with alternative spelling are present in this data set, such as the *London School of Hygiene & Tropical Medicine* and *London School of Hygiene [& OR and] Tropical Medicine*. In this analysis, we did not remove these duplicates; however, it is perhaps a good validation that they tended to cluster together and their scores were reasonably similar.

We visualized the RTI for countries with 100 or more scored papers per year available in PMC-OAI between 2010 and 2020 (Multimedia Appendix 6). Each frame represents a different year, where blue represents relatively high scores, and yellow represents relatively low scores. Ethiopia was consistently one of the best performing countries, leading all countries in RTI in 9 out of the 11 years; Ethiopia achieved the highest country average in 2020 (4.98; for reference, RTI in 2020 was 4.13). Norway had the highest RTI papers published in 2010 and 2011. None of the countries consistently had the lowest RTI. The countries with the lowest average in multiple years were Russia (2011, 2013, and 2018), Romania (2012 and 2014), and Ukraine (2015-2017). In terms of volume, the United States and China consistently published the most papers, with the United Kingdom serving as a distant third.

A graphic with coloring scaled to a country's RTI has been shown over the last 10 years for countries with 100 or more papers. Blue indicates relatively high average values. Yellow indicates relatively low average values. This video is available as an .mp4 file in Multimedia Appendix 6.

XSL•FO
RenderX

Figure 8. Clustering and Rigor and Transparency (RTI) ranking of the top 80 UK departments by paper count are shown. The t-distributed stochastic neighbor embeddings of the semantic vector representation of each department's average paper abstract is shown, with k-means clusters indicated by coloring (left panel). Field names are shown for clusters with a single unifying theme among all departments. The labels were added by hand for presentation purposes. We also show the average RTIs of each department (right panel).

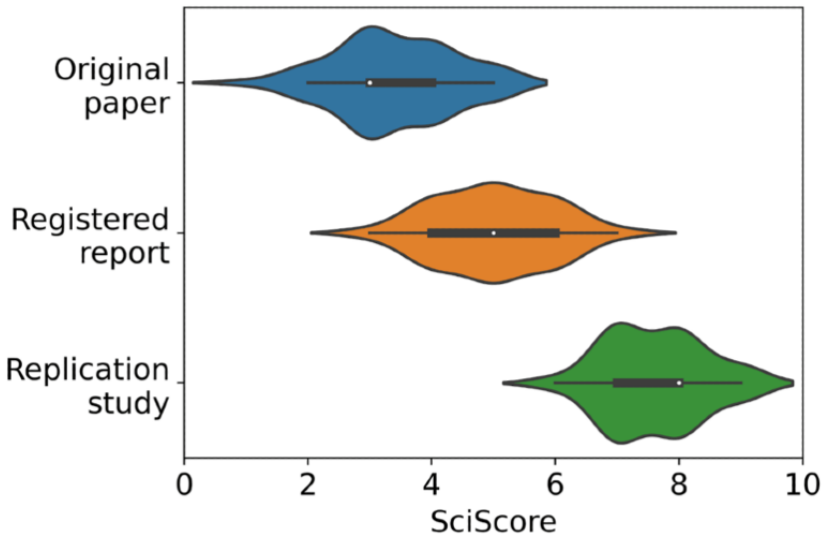


Criteria for Replicating a Study

The Cancer Reproducibility Project, headed by the Center for Open Science and Science Exchange, determined whether the top 50 cancer papers could be reproduced [6]. For each study, the project generated registered reports containing bulleted descriptions of the experimental protocols, data analyses, and replication study reports, which contained free-text descriptions of methods and results from each replicated experiment. The registered reports described their protocols step by step using bullet points, and resources were often only mentioned in reagent tables. Replication studies, in contrast, described both protocols and reagents in paragraphs throughout the methods sections. In addition, the registered reports seemed to focus more on protocol-specific best practices rather than on reporting best practices (eg, RRID use), which makes sense considering that they intend to report the results later. We expect that these differences largely contributed to the differences in scores between the registered reports and the replication studies.

To test our assumption that RTI may serve as a reasonable proxy for replicability, we compared the original studies, which often lacked sufficient detail for performing replication without contacting the original author, with the replicated studies. Figure 9 shows that the replicated reports (RTI 7.61, SD 0.78) were indeed significantly higher (P<.001) than their originating reports (RTI 3.39, SD 1.12). The scores of original papers that had responsive authors (RTI 3.45, SD 1.06) and those that did not have responsive authors (RTI 3.33, SD 1.06) were not significantly different on a paired, equal variance t test with 1 tail (P=.33). The underlying data are provided in Multimedia Appendix 4 [33].

Figure 9. Measured SciScores for Cancer Reproducibility Project papers. Original papers are in blue, registered reports are in orange, and replication studies are in green. A smoothed density plot of scores is shown in solid color. The white dot represents the median score, the thick black line the interquartile range (IQR), and the thin black line 1.5x IQR.



XSL•FO
RenderX

Discussion

Principal Findings

In this study, we introduced the latest version of the RTI, that is, RTI, version 2.0, a research reporting metric quantifying research quality and reporting transparency. The RTI lists journals, institutions, and countries with their composite scores and inclusion rates for rigor adherence and resource identifiability. We analyzed a significant number of manuscripts within the OAI subset of PMC, providing an opportunity to see general reporting trends within biomedicine and where we generally fall short within scientific reporting. In addition, we highlight the importance of high-quality reporting and demonstrate RTI's potential as a replication metric, using manuscripts from the Reproducibility Project: Cancer Biology. As with all generalized metrics, RTI is not perfect, and we do not expect all papers to score a perfect 10. This paper received a score of 7. As with any automated system, we cannot expect to handle all the edge cases. We expect RTI to be generally applicable to biomedical research. Other fields, for example, chemistry and physics, may not fit as well [14]. Although many of these less applicable papers are adequately handled as *not applicable* or through our more general rigor criteria, false positives do occur within automated systems. We are continuously working to improve RTI's generalizability through additional criteria (eg, data or code availability) and enhanced conditional scoring, where criteria are only factored in when relevant. Our overall aim is not to have every paper score a 10 but rather to help stakeholders improve papers that would otherwise score very poorly.

Technical Considerations

Unfortunately, the 2 primary limitations present in RTI, version 1.0, are still present in RTI, version 2.0. These issues can be summarized as follows. First, the OA subset represents only a fraction of the total biomedical literature and must therefore be considered a biased subsample. Second, papers with supplementary methods contained in PDFs are still unreadable to our algorithms, resulting in loss of data. We recognize that this is often due to constraints placed on the authors by the journal. As such, we again implore journals to lift restrictions that would limit the impact and reusability of a manuscript. These limitations have been described in our previous work [14].

Owing to the expanded abilities of SciScore, new considerations arose as well. Of these considerations, one of these stemmed from the addition of our data and code resolver, which attempts to resolve identifiers, URLs, and digital object identifiers by checking for their existence in external sources. To process millions of articles in a timely manner, we were forced to place a time restriction on the resolver. If the outside response time was too slow (≥5 seconds), we failed to resolve it, negatively affecting the reliability of our data. Therefore, we will not be able to comment on the validity of the identifiers detected, as we cannot differentiate between a slow outside resource and one that does not exist. In addition, because we only searched the materials and methods sections of the research manuscripts, as defined by Journal Article Tag Suites XML tags, we lost data only mentioned in other sections (eg, results). Anecdotally, this is especially true of criteria such as attrition, which is often mentioned in the results section and code or data availability statements, which can be listed within their own section at the end of manuscripts. We do report these but do not score these items for this reason. We expect to emend these issues in future versions of RTI.

SciScore's ability to process tables also improved in RTI, version 2.0, which had unintended side effects. Reagents were often counted twice in papers that used reagent tables (eg, STAR [structured, transparent, accessible reporting] methods) in addition to describing the reagents in their methods sections. In an extreme case, Hill et al [34] paper reported using 191 antibodies (listed in their STAR table), but SciScore identified 276 antibodies (identified from both the STAR table and the methods section text). The tool was not able to determine that the antibodies in the text and table were the same reagent for approximately half of the time in this study. This points to the need for continual improvement of artificial intelligence tools, as improvement in some aspects can lead to unintended consequences for others.

Analysis of Reporting Trends

Overview

After failing to replicate key findings in numerous scientific manuscripts, researchers introduced a variety of standards, guidelines, and checklists aimed at improving scientific reporting and with it, scientific reproducibility [10,13,35]. These guides appear to improve scientific reporting to some extent (Figure 2), although this effect seems be context specific [36]. Although researchers should try to ensure that their own manuscripts meet current best practices before submission, enforcing these standards should not fall entirely on journal staff. Researchers increasingly rely on multiple biological or software tools (antibodies, cell lines, plasmids, etc); these tools alone can have extremely complicated best practices, which may not be well understood by all researchers [37-39]. As such, authors, editors, and reviewers, especially in more general topic journals, may struggle to know which best practices to enforce and how to enforce them. In addition, 8% to 9% more papers are produced every year [40], and the current rate is roughly 2 papers added to PubMed every minute. This means that the task of spreading and checking best practices is difficult. Checklists can help guide best practices, and enforcing these checklists should lead to improved reporting standards [41], but given the scale of publishing, the use of automatic checklist tools such as SciScore and others, more focused tools such as Barzooka (continuous data in bar graphs), JetFighter (color-blind accessibility in visualizations), ODDPub (data and code availability), and RipetaScore (authorship, ethics, and data or code availability) [42-45], should help authors and reviewers improve manuscripts and address common checklist items and omissions consistently across many journals. In addition, automatic checklist completion can only help speed up the review process, which is a notoriously slow endeavor [46]. SciScore currently incorporates criteria from sources such as the ARRIVE (animal research: reporting of in vivo experiments) guidelines, the NIH standards, and the Materials Design,

XSL•FO
RenderX

Analysis, and Reporting checklist [9,13,35]. Other automated tools check for figure quality or the presence of limitations statements in the discussion section, which is an important part of several checklists. Additional checklist criteria (eg, PRISMA [Preferred Reporting Items for Systematic Reviews and Meta-Analyses]) should be added in future work, but automated tools such as this should be used to improve the reporting quality within the ever-growing literature. The RTI serves as a potential way to track how often these standards are met across a variety of stakeholders at various organizational levels.

Trends Across the General Literature

Experimental replication is a technique standardly used across many different fields. Replication metadata are important to report because readers need it to make accurate inferences about the trustworthiness of an experiment [47]. In 2020, Frommlet and Heinze [48] used meta-analysis to analyze 37 mouse experiments published in *Immunity* for experimental replication data. Although we did not replicate their exact study, our results are comparable. We limited our analysis to manuscripts containing a statement addressing IACUC approval and a *Mice* Medical Subject Headings term in 2020. We analyzed a few replication reporting criteria (the proportion of papers containing an explicit replication statement, number of replications, or type of replication). A major difference in our analyses is that Frommlet and Heinze [48] determined the presence of replication when the manuscript contained a figure indicating data representative of multiple experiments or when an explicit statement was made, whereas our classifier was trained exclusively on explicit statements (eg, "experiments were replicated in triplicate"). Of the manuscripts examined by them, 92% (34/37) contained some form of replication, whereas our data showed a far more conservative rate of 44% (1736/3917). In line with our data, Frommlet and Heinze observe that "the exact number [of replications] is frequently not even specified" and "in virtually all cases, [the replication information provided] is insufficient" [48]. Although not directly comparable, our data show that 42% of mouse research papers in 2020 mentioned a number associated with the amount of independent replications and only 6% explicitly mentioned the type of replication they were performing (ie, technical or biological). Although different in specifics, our results both indicate that replication metadata are generally underreported (at least in mice experiments), showing an easy source of potential improvement within research reporting.

Replication is not the only factor that negatively affects research reproducibility. Misidentified and contaminated cell lines continue to be a significant problem, with reported use rates varying between 10% and 50% [49-51]. Some reporting tools such as RRIDs appear to have lessened the incidence rates of problematic cell lines, as researchers are able to more easily look at a specific cell line's history [26], but there is still more work to be done. The most direct solution is to properly authenticate cell lines in the laboratory. Although different methods are continuing to be developed, short tandem repeat DNA profiling is currently most used [52-54]. However, this process is both time-consuming and expensive [55]. On the basis of our analysis of papers containing at least one cell line from 1997 to 2020, the rates of authentication have increased but are still low (6% to 8%). Similarly, the rate of contamination checks increased from 1% to 8% across the same time frame (Figure 5). In 2015, *Nature* reported that between 2013 and 2015, only 10% of authors submitting cell line–based papers (n=60) reported authenticating their cell lines [56]. The similarity in values indicates that cell line authentication is severely underreported (and most likely underperformed) in a large portion of biomedical literature. *Nature's* solution was to enhance its current submission policies to require authors to provide further details on cell line testing. This is easier said than done though. In 2010, the *International Journal of Cancer* became the first journal to require cell line authentication information [57]. Overall, this manual effort proved extremely effective, as the number of problematic cell lines published effectively went to 0 after implementation. This came at an administrative cost, as 240 additional hours were required to enforce these guidelines over the course of the 3-year study [58]. Fortunately, much of the work listed (eg, checking the manuscript and cell line–related data entry) can be automated. On the basis of this, we recommend that journals implement stringent cell line authentication requirements similar to those of the *International Journal of Cancer* and make use of automated tools to limit the administrative costs of best practice enforcement. Future studies could compare journal authentication and contamination rates against the specific guidelines implemented by each journal to determine which guidelines and enforcement strategies are most effective. Future models could also differentiate between authentication methods for more granular analysis.

Criteria Across Journals, Research Institutions, and Countries

By directly linking institutions with their research manuscripts, we created a way to track and rate an institution's published output. The latest version of the RTI, that is, RTI, version 2.0, lists an institution's adherence to various reproducibility-related criteria, as well as the identifiability of its research resources (antibodies, organisms, plasmids, etc). The RTI lists the composite scores for multiple entities (ie, journals, institutions, and countries). On the basis of our analysis of the data obtained from the study by Lepori et al [32], there is no strong correlation (r=−0.12) between an institution's RTI and its total revenue, after correcting for the size of the university through the number of academic staffs.

Although indicators such as global rank, funding, and even citations may be, to some extent, richness measures [32,59], the RTI is not. There is no significant correlation, which leads us to believe that research quality is not entirely driven by funding (or how rich a university or country is). Anecdotally, we believe this is largely owing to a researcher's knowledge of best practices and the community's ability to implement and enforce them. The first condition may appear to be met as an increasing number of journals implement best practice submission guidelines and checklists, but this is only the first step. These guidelines must be accessible and easily understood if they are to be effectively used [36]. Once the first condition is met, the second condition should follow more easily, especially if aided by automated tooling. We hope that by comparing research institutions based on the quality of their

XSL•FO
RenderX

research outputs, they consider rigor and transparency more in their decision-making with the ultimate goal being a shift from *publish or perish* to rigor and reproducibility.

To further encourage this, we aimed to apply RTI comparisons at the departmental level. Different fields can have drastically different reporting requirements and standards, making more granular comparisons far more tenuous. Nominal grouping alone may not be sufficient, as department names may not fully represent the breadth of a department or the nuances of the different subfields within. To mitigate this, we clustered the top 80 UK departments based on the semantic similarity of their abstracts. As shown, the generated clusters aligned remarkably well with department names, despite being fed only semantic abstract information (Figure 8). Not only do these clusters quantitate differences across departments but they also provide new information that cannot be obtained from name alone. For instance, based on other departments within the same cluster, it appears that the Department of Medicine at the University of Oxford focuses on epidemiological or public health research, whereas the Department of Medicine at the University of Cambridge tends to publish cellular biology research. Using our proposed clustering method, we can quantitate such nuanced differences between departments, allowing a like-to-like comparison of RTIs at the departmental level.

After adding both institution- and country-specific data (as well as expanding the entity types detected), we believe that the RTI's ability to serve as a proxy for good rigor and transparency practices has only been enhanced. Institutions and countries can now more easily identify areas where they fall short in rigor and reproducibility as well as monitor the impact of various scientific policies. We hope that the RTI will continue to highlight the importance of sound scientific practices.

Criteria for Replicating a Study

Although we scored >2 million papers across a range of fields, it is difficult to assess whether a particular score has any relevance to the ability of others to replicate a study. Using work done by the Center for Open Science's Reproducibility Project: Cancer Biology [6], we were able to look at the scores of all papers originally in their study (RTI 3.40, SD 0.95), which researchers used to attempt to replicate the experiments. According to Errington et al [6], none of the original manuscripts contained sufficient detail to attempt to replicate the study, and all required additional information from the authors. To begin replication attempts, Errington et al [6] had to email the original authors and were only able to replicate studies when the original authors responded with additional details. This process is unreliable and slow and results in the loss of a few experiments, as some authors did not respond. Following this, Errington et al [6] generated registered reports, documenting each protocol in a step-by-step manner. After the review, replication reports containing in-depth descriptions of their methods and results were published. These reports were intentionally as rigorous and transparent as possible, sharing all data and codes openly, following resource-specific best practices, and ensuring that all reagents were listed as transparently as possible. As a result, they scored significantly higher (RTI 7.61, SD 0.78) than their originating manuscripts (Figure 9). We assume that these replication papers, where authors paid as much attention to methodological detail as possible, are much more likely to be replicable without additional correspondence. Although we cannot simply describe all *2* papers as not replicable and all *8* papers as replicable, as numerous fields and their subsequent best practices exist, we can state that higher scores are associated with more methodological detail and as such are likely easier to replicate. We encourage biomedicine authors to aim for high scores by ensuring that their methods sections include as much detail as possible.

Acknowledgments

The authors would like to acknowledge the following grants for funding part of this work: R43 OD024432/OD/NIH HHS/United States, R44 MH119094/MH/NIMH NIH HHS/United States, U24 DA039832/DA/NIDA NIH HHS/United States, and U24 DK097771/DK/NIDDK NIH HHS/United States. The authors would also like to thank the curators who work with authors to help assign new RRIDs, especially Mr Mason Pairish, Ms Edyta Vieth, Dr Amos Bairoch, Dr Cynthia Smith, and Dr Shur-Jen Wang and curators who work with the literature to gather RRID data that were used to refine our algorithms, especially Nathan Anderson, Ethan Badger, Gabrielle Pine, and Zeljiana Babic. They would also like to thank the many journal editors, authors, and reviewers who added RRIDs to the manuscript.

Data and Code Availability

The code for retrieving and preprocessing OA subset XML data has been previously published and is open source [60]. SciScore's source code is not publicly available owing to its proprietary nature. However, the framework and core features of its underlying classifiers have been described in previous studies on Resource Disambiguator for the Web [24]. All Resource Disambiguator for the Web codes are available and may serve as reasonable equivalents to SciScore's named entity recognition classifier components. The regular expression patterns used for identifier extraction are available in Multimedia Appendix 3 [61]. The code used for institutional disambiguation is available in GitHub [29]. All SQL statements and spreadsheets used in the figures and analyses can be found in Multimedia Appendix 4 [33]. Data sets generated during OA subset *scoring* containing data from individual papers are not publicly available because of their potentially sensitive nature (ie, low scores assigned to published papers may negatively impact scientists producing OA works, significantly more so than researchers publishing closed access works, without giving them the ability to respond to criticism) but are available from the corresponding author upon reasonable

XSL•FO
RenderX

request. Summary data for each journal, institution, and country are provided in supplementary files and have been made available in Multimedia Appendix 5 and via SciScore's RTI webpage [62].

Conflicts of Interest

AB, JG, and IBO have an equity interest in SciCrunch Inc, the company behind the development of SciScore. PE and NA were employed by SciCrunch as the scientific curators. JM was employed as a scientific curator until 2020. The terms of this arrangement were reviewed and approved by the University of California, San Diego, California, in accordance with its conflicts of interest policies. MR and IBO serve as independent contractors for SciCrunch.

Multimedia Appendix 1
Criteria detected using SciScore with applicable guideline source, description, and example listed.
[DOCX File , 17 KB - jmir_v24i6e37324_app1.docx]

Multimedia Appendix 2
Individual Classifier Performance for Named-Entities. Training set size is shown as the number of entities, which represents the total number of entities tagged by our curators as either positive or negative and number of sentences, which represents the total number of sentences containing positive and negative examples as well as some sentences without any entities used in both training and testing.
[DOCX File , 25 KB - jmir_v24i6e37324_app2.docx]

Multimedia Appendix 3
JSON file containing regular expression patterns used for protocol, data, and code identifiers.
[TXT File , 8 KB - jmir_v24i6e37324_app3.txt]

Multimedia Appendix 4
Data underlying figures.
[XLSX File (Microsoft Excel File), 442 KB - jmir_v24i6e37324_app4.xlsx]

Multimedia Appendix 5
Rigor and Transparency Index.
[XLSX File (Microsoft Excel File), 25249 KB - jmir_v24i6e37324_app5.xlsx]

Multimedia Appendix 6
A graphic with coloring scaled to a country's Rigor and Transparency Index shown over the last 10 years (2010-2020) for countries with 100 or more papers. Blue shows relatively high averages. Yellow shows relatively low averages.
[MP4 File (MP4 Video), 209 KB - jmir_v24i6e37324_app6.mp4]

References

1. Ioannidis JP. Why most published research findings are false. PLoS Med 2005 Aug;2(8):e124 [FREE Full text] [doi: 10.1371/journal.pmed.0020124] [Medline: 16060722]
2. Begley CG, Ellis LM. Drug development: raise standards for preclinical cancer research. Nature 2012 Mar 28;483(7391):531-533. [doi: 10.1038/483531a] [Medline: 22460880]
3. Open Science Collaboration. PSYCHOLOGY. Estimating the reproducibility of psychological science. Science 2015 Aug 28;349(6251):aac4716. [doi: 10.1126/science.aac4716] [Medline: 26315443]
4. Vasilevsky NA, Brush MH, Paddock H, Ponting L, Tripathy SJ, Larocca GM, et al. On the reproducibility of science: unique identification of research resources in the biomedical literature. PeerJ 2013 Sep 5;1:e148 [FREE Full text] [doi: 10.7717/peerj.148] [Medline: 24032093]
5. Freedman LP, Cockburn IM, Simcoe TS. The economics of reproducibility in preclinical research. PLoS Biol 2015 Jun 9;13(6):e1002165 [FREE Full text] [doi: 10.1371/journal.pbio.1002165] [Medline: 26057340]
6. Errington TM, Denis A, Perfito N, Iorns E, Nosek BA. Challenges for assessing replicability in preclinical cancer biology. Elife 2021 Dec 07;10:e67995 [FREE Full text] [doi: 10.7554/eLife.67995] [Medline: 34874008]
7. Freedman LP, Venugopalan G, Wisman R. Reproducibility2020: progress and priorities. F1000Res 2017 May 2;6:604 [FREE Full text] [doi: 10.12688/f1000research.11334.1] [Medline: 28620458]
8. Viergever RF, Hendriks TC. The 10 largest public and philanthropic funders of health research in the world: what they fund and how they distribute their funds. Health Res Policy Syst 2016 Feb 18;14:12 [FREE Full text] [doi: 10.1186/s12961-015-0074-z] [Medline: 26892771]

XSL•FO
RenderX

9. Landis SC, Amara SG, Asadullah K, Austin CP, Blumenstein R, Bradley EW, et al. A call for transparent reporting to optimize the predictive value of preclinical research. Nature 2012 Oct 11;490(7419):187-191 [FREE Full text] [doi: 10.1038/nature11556] [Medline: 23060188]
10. Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. Nature 2014 Jan 30;505(7485):612-613 [FREE Full text] [doi: 10.1038/505612a] [Medline: 24482835]
11. Final NIH Policy for Data Management and Sharing. Publication NOT-OD-21-013. National Institutes of Health. 2020 Oct 29. URL: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html [accessed 2022-02-02]
12. Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, et al. SCIENTIFIC STANDARDS. Promoting an open research culture. Science 2015 Jun 26;348(6242):1422-1425 [FREE Full text] [doi: 10.1126/science.aab2374] [Medline: 26113702]
13. Macleod M, Collings AM, Graf C, Kiermer V, Mellor D, Swaminathan S, et al. The MDAR (Materials Design Analysis Reporting) framework for transparent reporting in the life sciences. Proc Natl Acad Sci U S A 2021 Apr 27;118(17):e2103238118 [FREE Full text] [doi: 10.1073/pnas.2103238118] [Medline: 33893240]
14. Menke J, Roelandse M, Ozyurt B, Martone M, Bandrowski A. The rigor and transparency index quality metric for assessing biological and medical science methods. iScience 2020 Oct 20;23(11):101698 [FREE Full text] [doi: 10.1016/j.isci.2020.101698] [Medline: 33196023]
15. Begley CG, Buchan AM, Dirnagl U. Robust research: institutions must do their part for reproducibility. Nature 2015 Sep 03;525(7567):25-27. [doi: 10.1038/525025a] [Medline: 26333454]
16. Quan W, Chen B, Shu F. Publish or impoverish: an investigation of the monetary reward system of science in China (1999-2016). Aslib J Inf Manag 2017 Sep 18;69(5):486-502. [doi: 10.1108/ajim-01-2017-0014]
17. Hammarfelt B. Recognition and reward in the academy: valuing publication oeuvres in biomedicine, economics and history. Aslib J Inf Manag 2017 Sep 18;69(5):607-623. [doi: 10.1108/AJIM-01-2017-0006]
18. Rouleau G. Open Science at an institutional level: an interview with Guy Rouleau. Genome Biol 2017 Jan 20;18(1):14 [FREE Full text] [doi: 10.1186/s13059-017-1152-z] [Medline: 28109193]
19. Macleod MR, O'Collins T, Horky LL, Howells DW, Donnan GA. Systematic review and metaanalysis of the efficacy of FK506 in experimental stroke. J Cereb Blood Flow Metab 2005 Jun;25(6):713-721. [doi: 10.1038/sj.jcbfm.9600064] [Medline: 15703698]
20. Schroyens N, Sigwald EL, Van Den Noortgate W, Beckers T, Luyten L. Reactivation-dependent amnesia for contextual fear memories: evidence for publication bias. eNeuro 2021 Jan 22;8(1):ENEURO.0108-20.2020 [FREE Full text] [doi: 10.1523/ENEURO.0108-20.2020] [Medline: 33355289]
21. Serghiou S, Contopoulos-Ioannidis DG, Boyack KW, Riedel N, Wallach JD, Ioannidis JP. Assessment of transparency indicators across the biomedical literature: how open is open? PLoS Biol 2021 Mar 1;19(3):e3001107 [FREE Full text] [doi: 10.1371/journal.pbio.3001107] [Medline: 33647013]
22. Howat AM, Mulhern A, Logan HF, Redvers-Mutton G, Routledge C, Clark J. Converting Access Microbiology to an open research platform: focus group and AI review tool research results. Access Microbiol 2021 Apr 19;3(4):000232 [FREE Full text] [doi: 10.1099/acmi.0.000232] [Medline: 34151179]
23. Lafferty JD, McCallum A, Pereira FC. Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: Proceedings of the 18th International Conference on Machine Learning. 2001 Presented at: ICML '01; June 28-July 1, 2001; Williamstown, MA, USA p. 282-289. [doi: 10.1145/1015330.1015422]
24. Ozyurt IB, Grethe JS, Martone ME, Bandrowski AE. Resource disambiguator for the Web: extracting biomedical resources and their citations from the scientific literature. PLoS One 2016 Jan 5;11(1):e0146300 [FREE Full text] [doi: 10.1371/journal.pone.0146300] [Medline: 26730820]
25. Identifiers.org Resolution Service. URL: https://identifiers.org/ [accessed 2022-06-07]
26. Babic Z, Capes-Davis A, Martone ME, Bairoch A, Ozyurt IB, Gillespie TH, et al. Incidences of problematic cell lines are lower in papers that use RRIDs to identify cell lines. Elife 2019 Jan 29;8:e41676 [FREE Full text] [doi: 10.7554/eLife.41676] [Medline: 30693867]
27. ROR Data Dump 7.1 (March 2021). figshare. URL: https://figshare.com/articles/dataset/ROR_Data_Dump_7_1_March_2021_/14273357 [accessed 2022-06-07]
28. Research Organization Registry (ROR) API. GitHub. URL: https://github.com/ror-community/ror-api [accessed 2022-02-15]
29. Fast ror disambiguator. GitHub. URL: https://github.com/PeterEckmann1/fast-ror-disambiguator [accessed 2022-02-15]
30. Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching word vectors with subword information. Trans Assoc Comput Linguist 2017 Dec 30;5:135-146 [FREE Full text] [doi: 10.1162/tacl_a_00051]
31. Pedregosa F, Varoquax G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12(85):2825-2830 [FREE Full text] [doi: 10.1145/2786984.2786995]
32. Lepori B, Geuna A, Mira A. Scientific output scales with resources. A comparison of US and European universities. PLoS One 2019 Oct 15;14(10):e0223415 [FREE Full text] [doi: 10.1371/journal.pone.0223415] [Medline: 31613903]
33. Data_underlying_figures_tables. Google Sheets. URL: https://docs.google.com/spreadsheets/d/1ySmU_7VleobYnAYgsY8StZ_9WBeml1e7hyZKt6myTTE/edit?usp=sharing [accessed 2022-02-15]

XSL•FO
RenderX

34. Hill SM, Nesser NK, Johnson-Camacho K, Jeffress M, Johnson A, Boniface C, et al. Context specificity in causal signaling networks revealed by phosphoprotein profiling. Cell Syst 2017 Jan 25;4(1):73-83.e10 [FREE Full text] [doi: 10.1016/j.cels.2016.11.013] [Medline: 28017544]
35. Percie du Sert N, Hurst V, Ahluwalia A, Alam S, Avey MT, Baker M, et al. The ARRIVE guidelines 2.0: updated guidelines for reporting animal research. PLoS Biol 2020 Jul 14;18(7):e3000410 [FREE Full text] [doi: 10.1371/journal.pbio.3000410] [Medline: 32663219]
36. Hepkema WM, Horbach SP, Hoek JM, Halffman W. Misidentified biomedical resources: journal guidelines are not a quick fix. Int J Cancer 2022 Apr 15;150(8):1233-1243. [doi: 10.1002/ijc.33882] [Medline: 34807460]
37. Baust JM, Buehring GC, Campbell L, Elmore E, Harbell JW, Nims RW, et al. Best practices in cell culture: an overview. In Vitro Cell Dev Biol Anim 2017 Sep;53(8):669-672. [doi: 10.1007/s11626-017-0177-7] [Medline: 28808859]
38. Brodaczewska KK, Szczylik C, Fiedorowicz M, Porta C, Czarnecka AM. Choosing the right cell line for renal cell cancer research. Mol Cancer 2016 Dec 19;15(1):83 [FREE Full text] [doi: 10.1186/s12943-016-0565-8] [Medline: 27993170]
39. Price PJ. Best practices for media selection for mammalian cells. In Vitro Cell Dev Biol Anim 2017 Sep;53(8):673-681. [doi: 10.1007/s11626-017-0186-6] [Medline: 28726187]
40. Landhuis E. Scientific literature: information overload. Nature 2016 Jul 21;535(7612):457-458. [doi: 10.1038/nj7612-457a] [Medline: 27453968]
41. Chaparro A, Keebler JR, Lazzara EH, Diamond A. Checklists: a review of their origins, benefits, and current uses as a cognitive aid in medicine. Ergon Des 2019 Jan 22;27(2):21-26. [doi: 10.1177/1064804618819181]
42. Weissgerber T, Riedel N, Kilicoglu H, Labbé C, Eckmann P, Ter Riet G, et al. Automated screening of COVID-19 preprints: can we help authors to improve transparency and reproducibility? Nat Med 2021 Jan;27(1):6-7 [FREE Full text] [doi: 10.1038/s41591-020-01203-7] [Medline: 33432174]
43. Saladi S. JetFighter: towards figure accuracy and accessibility. Elife. 2019 Apr 4. URL: https://elifesciences.org/labs/c2292989/jetfighter-towards-figure-accuracy-and-accessibility [accessed 2022-05-04]
44. Riedel N, Kip M, Bobrov E. ODDPub – a text-mining algorithm to detect data sharing in biomedical publications. Data Sci J 2020 Oct 29;19(1):42. [doi: 10.5334/dsj-2020-042]
45. Sumner JQ, Vitale CH, McIntosh LD. RipetaScore: measuring the quality, transparency, and trustworthiness of a scientific work. Front Res Metr Anal 2022 Jan 21;6:751734 [FREE Full text] [doi: 10.3389/frma.2021.751734] [Medline: 35128302]
46. Powell K. Does it take too long to publish research? Nature 2016 Mar 11;530(7589):148-151. [doi: 10.1038/530148a] [Medline: 26863966]
47. Blainey P, Krzywinski M, Altman N. Points of significance: replication. Nat Methods 2014 Sep;11(9):879-880. [doi: 10.1038/nmeth.3091] [Medline: 25317452]
48. Frommlet F, Heinze G. Experimental replications in animal trials. Lab Anim 2021 Mar;55(1):65-75 [FREE Full text] [doi: 10.1177/0023677220907617] [Medline: 32138592]
49. Schweppe RE, Klopper JP, Korch C, Pugazhenthi U, Benezra M, Knauf JA, et al. Deoxyribonucleic acid profiling analysis of 40 human thyroid cancer cell lines reveals cross-contamination resulting in cell line redundancy and misidentification. J Clin Endocrinol Metab 2008 Nov;93(11):4331-4341 [FREE Full text] [doi: 10.1210/jc.2008-1102] [Medline: 18713817]
50. Liang-Chu MM, Yu M, Haverty PM, Koeman J, Ziegle J, Lee M, et al. Human biosample authentication using the high-throughput, cost-effective SNPtrace(TM) system. PLoS One 2015 Feb 25;10(2):e0116218 [FREE Full text] [doi: 10.1371/journal.pone.0116218] [Medline: 25714623]
51. Huang Y, Liu Y, Zheng C, Shen C. Investigation of cross-contamination and misidentification of 278 widely used tumor cell lines. PLoS One 2017 Jan 20;12(1):e0170384 [FREE Full text] [doi: 10.1371/journal.pone.0170384] [Medline: 28107433]
52. Masters JR, Thomson JA, Daly-Burns B, Reid YA, Dirks WG, Packer P, et al. Short tandem repeat profiling provides an international reference standard for human cell lines. Proc Natl Acad Sci U S A 2001 Jul 03;98(14):8012-8017 [FREE Full text] [doi: 10.1073/pnas.121616198] [Medline: 11416159]
53. Mohammad TA, Tsai YS, Ameer S, Chen HI, Chiu YC, Chen Y. CeL-ID: cell line identification using RNA-seq data. BMC Genomics 2019 Mar 04;20(Suppl 1):81 [FREE Full text] [doi: 10.1186/s12864-018-5371-9] [Medline: 30712511]
54. Mzurikwao D, Khan MU, Samuel OW, Cinatl Jr J, Wass M, Michaelis M, et al. Towards image-based cancer cell lines authentication using deep neural networks. Sci Rep 2020 Nov 16;10(1):19857 [FREE Full text] [doi: 10.1038/s41598-020-76670-6] [Medline: 33199764]
55. Freedman LP, Gibson MC, Wisman R, Ethier SP, Soule HR, Reid YA, et al. The culture of cell culture practices and authentication--results from a 2015 survey. Biotechniques 2015 Oct 1;59(4):189-192 [FREE Full text] [doi: 10.2144/000114344] [Medline: 26458546]
56. Announcement: time to tackle cells' mistaken identity. Nature 2015 Apr 16;520(7547):264. [doi: 10.1038/520264a]
57. Lichter P, Allgayer H, Bartsch H, Fusenig N, Hemminki K, von Knebel Doeberitz M, et al. Obligation for cell line authentication: appeal for concerted action. Int J Cancer 2010 Jan 01;126(1):1 [FREE Full text] [doi: 10.1002/ijc.24985] [Medline: 19882693]
58. Fusenig NE, Capes-Davis A, Bianchini F, Sundell S, Lichter P. The need for a worldwide consensus for cell line authentication: experience implementing a mandatory requirement at the International Journal of Cancer. PLoS Biol 2017 Apr 17;15(4):e2001438 [FREE Full text] [doi: 10.1371/journal.pbio.2001438] [Medline: 28414712]

XSL•FO
RenderX

59. Gadd E. Mis-measuring our universities: why global university rankings don't add up. Front Res Metr Anal 2021 Sep 9;6:680023 [FREE Full text] [doi: 10.3389/frma.2021.680023] [Medline: 34568739]
60. Resource disambiguator. GitHub. 2021 Aug 27. URL: https://github.com/SciCrunch/resource_disambiguator [accessed 2022-02-15]
61. File_1. Google Drive. URL: https://drive.google.com/file/d/12L9QKlXeBouxM8fly9UNr18ZaEGVIWez/view [accessed 2022-02-15]
62. SciScore. Scicrunch Inc. URL: https://sciscore.com/RTI [accessed 2022-02-15]

Abbreviations

API: application programming interface
ARRIVE: animal research: reporting of in vivo experiments
CRF: conditional random field
NIH: National Institutes of Health
OA: open access
OAI: Open Archives Initiative
PMC: PubMed Central
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
ROR: Research Organization Registry
RRID: research resource identifier
RTI: Rigor and Transparency Index
STAR: structured, transparent, accessible reporting

Edited by T Leung; submitted 15.02.22; peer-reviewed by SS Amritphale, K Sward; comments to author 20.04.22; revised version received 10.05.22; accepted 23.05.22; published 27.06.22.

Please cite as:
Menke J, Eckmann P, Ozyurt IB, Roelandse M, Anderson N, Grethe J, Gamst A, Bandrowski A
Establishing Institutional Scores With the Rigor and Transparency Index: Large-scale Analysis of Scientific Reporting Quality
J Med Internet Res 2022;24(6):e37324
URL: https://www.jmir.org/2022/6/e37324
doi:10.2196/37324
PMID:35759334

©Joe Menke, Peter Eckmann, Ibrahim Burak Ozyurt, Martijn Roelandse, Nathan Anderson, Jeffrey Grethe, Anthony Gamst, Anita Bandrowski. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.06.2022. 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 Pharmacy-Based eHealth Intervention Promoting Correct Use of Medication in Patients With Asthma and COPD: Nonrandomized Pre-Post Study

Kyma Schnoor[1,2], MSc; Anke Versluis[1,2], PhD; Robbert Bakema[3], MSc; Sanne van Luenen[4], PhD; Marcel J Kooij[5], PharmD, PhD; J Maurik van den Heuvel[6], PhD; Martina Teichert[7], PharmD, PhD; Persijn J Honkoop[1], MD, PhD; Job F M van Boven[8], PhD, PharmD; Niels H Chavannes[1,2], MD, PhD; Jiska J Aardoom[1,2], PhD

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

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

[3]Nederlandse Service Apotheek Beheer, 's Hertogenbosch, Netherlands

[4]Department of Clinical Psychology, Leiden University, Leiden, Netherlands

[5]Service Pharmacy Koning, Amsterdam, Netherlands

[6]Department of Respiratory Medicine, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, Netherlands

[7]Department of Clinical Pharmacy and Toxicology, Leiden University Medical Center, Leiden, Netherlands

[8]Department of Clinical Pharmacy and Pharmacology, Groningen Research Institute for Asthma and COPD, University Medical Center Groningen, University of Groningen, Groningen, Netherlands

Corresponding Author:
Jiska J Aardoom, PhD
Department of Public Health and Primary Care
Leiden University Medical Center
Hippocratespad 21
Postzone V0-P, PO Box 9600
Leiden,
Netherlands
Phone: 31 7152668605
Email: j.j.aardoom@lumc.nl

Abstract

Background: Asthma and chronic obstructive pulmonary disease (COPD) affect millions of people worldwide. While medication can control and improve disease symptoms, incorrect use of medication is a common problem. The eHealth intervention SARA (Service Apothecary Respiratory Advice) aims to improve participants' correct use of inhalation medication by providing information and as-needed tailored follow-up support by a pharmacist.

Objective: The primary aim of this study was to investigate the effect of SARA on exacerbation rates in participants with asthma and COPD. Secondary aims were to investigate its effects in terms of adherence to maintenance medication and antimycotic treatment.

Methods: In this nonrandomized pre-post study, medication dispensing data from 382 Dutch community pharmacies were included. Exacerbation rates were assessed with dispensed short-course oral corticosteroids. Medication adherence between new and chronic users was assessed by calculating the proportion of days covered from dispensed inhalation maintenance medication. Antimycotic treatment was investigated from dispensed oral antimycotics in participants who were also dispensed inhaled corticosteroids (ICS). Outcomes were assessed 1 year before and 1 year after implementation of SARA and were compared between SARA participants and control participants. More specifically, for exacerbation rates and medication adherence, a difference score was calculated (ie, 1 year after SARA minus 1 year before SARA) and was subsequently compared between the study groups with independent-samples *t* tests. For antimycotics, the relative number of participants who were dispensed antimycotics was calculated and subsequently analyzed with a mixed-effects logistic regression.

Results: The study population comprised 9452 participants, of whom 2400 (25.39%) were SARA participants. The mean age of the population was 60.8 (15.0) years, and approximately two-thirds (n=5677, 60.06%) were female. The results showed an increase in mean exacerbation rates over time for both study groups (SARA: 0.05; control: 0.15). However, this increase in

XSL•FO
RenderX

exacerbation rates was significantly lower for SARA participants (t_{9450}=3.10, 95% CI 0.04-0.16; *P*=.002; Cohen *d*=0.06). Chronic users of inhalation medication in both study groups showed an increase in mean medication adherence over time (SARA: 6.73; control: 4.48); however, this increase was significantly higher for SARA participants (t_{5886}=–2.74, 95% CI –3.86 to –0.84; *P*=.01; Cohen *d*=–0.07). Among new users of inhalation medication, results showed no significant difference in medication adherence between SARA and control participants in the year after implementation of SARA (t_{1434}=–1.85, 95% CI –5.60 to 0.16; *P*=.06; Cohen *d*=–0.10). Among ICS users, no significant differences between the study groups were found over time in terms of the proportion of participants who were dispensed antimycotics (t_{5654}=0.29, 95% CI –0.40 to 0.54; *P*=.76; Cohen *d*=0).

Conclusions: This study provides preliminary evidence that the SARA eHealth intervention might have the potential to decrease exacerbation rates and improve medication adherence among patients with asthma and COPD.

(J Med Internet Res 2022;24(6):e32396) doi:10.2196/32396

KEYWORDS

asthma; COPD; medication adherence; exacerbations; pharmacy; eHealth

Introduction

Asthma and chronic obstructive pulmonary disease (COPD) are chronic respiratory diseases that affect millions of people worldwide [1,2]. Asthma and COPD place a significant health burden on patients and an economic burden on society [3-5]. Medication cannot cure these diseases but can reduce disease symptoms and improve control, which, in turn, can positively affect patients' quality of life [6-9]. Unfortunately, nonadherence to maintenance medication is common in patients with asthma and COPD. Indeed, adherence rates have been found to vary from 22% to 78% [7]. Nonadherence can have detrimental effects on clinical outcomes for individuals with asthma and COPD. Notably, it could negatively affect lung function, disease control, exacerbation rate, health-related quality of life, and work productivity [6,7,10]. In addition, nonadherence has been associated with higher health care use and costs [6,7].

Factors related to nonadherence to inhaled medication are multifaceted and can include intentional nonadherence (eg, concerns about side effects and complexity of medication regime) and unintentional nonadherence (eg, experiencing difficulties with how or when to use medication or lacking skills to use inhaler devices) [7,9,11-15]. Regarding incorrect use of the inhalers, Lavorini et al [12] systematically investigated the use of dry powder inhalers by patients with asthma or COPD. The results showed that between 4% and 94% of the patients did not use their inhalers correctly, with exact rates depending on the type of inhaler and the assessment method used. As such, these patients need additional care to support correct medication usage, and effective intervention strategies are required.

A variety of strategies have been investigated that aim to tackle the problem of nonadherence. Training and education on correct inhaler technique are considered crucial in combating nonadherence [9] and in effectively managing one's asthma or COPD [16]. A Cochrane systematic review demonstrated the efficacy of interventions intended to improve adherence to inhaled corticosteroids (ICS) among patients with asthma [17]. Adherence education, electronic trackers or reminders, and simplified regimens were found to improve adherence by 20%, 19%, and 4%, respectively [17]. Recently, a meta-analysis by Jeminiwa et al [18] also showed a positive effect of eHealth strategies on improving adherence to ICS among people with asthma. However, according to the Cochrane systematic review, clinical outcomes are often not improved with those interventions [17].

In the Netherlands, the eHealth intervention SARA (Service Apothecary Respiratory Advice; in Dutch, Service Apotheek Raad en Advies) was developed to promote correct use of inhalation medication for patients with asthma and COPD. The goal of this self-management intervention is to reduce the burden of lung disease and reduce exacerbations by stimulating correct use and adherence of inhaler medication in patients with asthma and COPD. SARA combines several interventions' components, including education, self-management strategies, and as-needed follow-up care by a pharmacist.

This study aimed to investigate the effectiveness of SARA in participants with asthma and COPD by comparing pharmacy dispensing data over time, that is, before and after the implementation of SARA, as well as between SARA participants and a control group. The primary aim of this study was to investigate the effect of SARA on exacerbation rates. The secondary aims were to investigate the effect of SARA on medication adherence and antimycotic treatment.

Methods

The SARA eHealth Intervention

The SARA eHealth intervention was developed by the Service Pharmacy organization. The Service Pharmacy organization supports independent but affiliated community pharmacies (ie, Service Pharmacies) in their day-to-day business operations to provide high-quality pharmaceutical care and provide offline and online communication. The Service Pharmacy organization developed SARA to support and prepare pharmacies for the second dispensing of inhalation medication. Pilot studies were then conducted with SARA and its corresponding questionnaire. Relevant input on how to improve the intervention was gathered by conducting two focus group interviews with pharmacists as well as patients with asthma and COPD, gaining insight into their needs and preferences. Their input was used to improve the intervention where possible.

SARA aims to improve the correct use of inhalation medication by providing information and supporting knowledge about this type of medication. SARA is an online platform that contains

XSL•FO
RenderX

the following: (1) comprehensive information about inhalation medication, its usage, and side effects; (2) inhalation instruction videos; (3) informational videos about asthma and COPD; (4) a pollen forecast; and (5) a questionnaire that is emailed to individuals on the 15th day after starting SARA. A 7-item questionnaire was developed by the Service Pharmacy organization, assessing patients' inhalation medication usage and related experiences, concerns and doubts, difficulties, and side effects (Multimedia Appendix 1). The questionnaire was based on the national Dutch guideline for pharmaceutical patient consultation, specifically regarding the second dispensing of inhalation medication, which was in development at the time [19]. This consultation guideline aims to support the community pharmacist in providing patient-centered care during pharmaceutical consultations provided by the pharmacist to the patient. The seven drafted questions were discussed in a focus group with pharmacists, and the feedback was subsequently used to improve the questionnaire to maximize its reliability. The outcomes of the questionnaires are automatically forwarded to the corresponding pharmacy. Next, the pharmacist can provide as-needed follow-up care in case any important issues are encountered, such as experiencing one or more severe side effects. The type and intensity of follow-up care can be tailored to the identified patient needs and preferences and the pharmacist's resources. Pharmacists are trained to identify patients' individual needs before delivering additional support, especially because SARA identifies those with extra needs. The follow-up care can entail additional detailed inhalation instructions or training (eg, when a patient experiences difficulties inhaling), providing additional information on how to properly use the medication (eg, when a patient reports not knowing when to take the medication or whether one can use the medication in combination with other medication), or providing additional information on the importance of taking the medication and its effects (eg, when a patient reports not having taken the medication because of doubts about whether it will work). The follow-up care can be offered through extra pharmacy visits, extra house visits, telephone calls, or digital communication tools, such as chats.

Design

This study entailed a nonrandomized pre-post study design. Pharmacy dispensing data were used to compare patient-level medication dispensing data over time (ie, the year before versus after implementation of SARA, hereafter often referred to as "over time") and between groups (ie, SARA versus control participants).

Ethical Considerations

No ethics approval was applied for because this study was declared to not fall within the scope of the Dutch Medical Research Involving Human Subjects Act by the Medical Ethics Committee (MEC) of the Leiden University Medical Center (MEC No. G20.030).

Participant Flow

From the beginning of 2017 onward, SARA has been implemented in approximately 400 Service Pharmacies in the Netherlands. Not all Service Pharmacies participated in SARA. Some pharmacies could not participate in SARA because of conflicting software programs, among other reasons. Other pharmacies declined to participate in SARA due to personnel problems, thereby resulting in not having the resources to implement a different and new way of working.

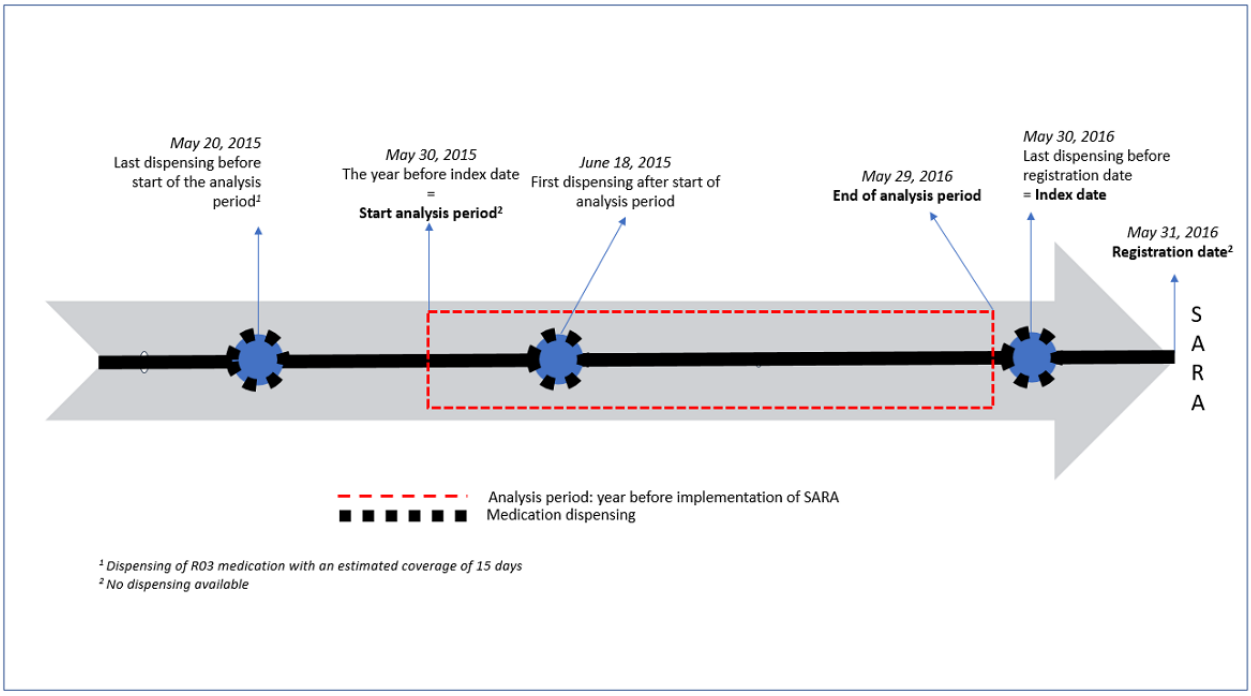
In the participating pharmacies, individuals were offered SARA during a pharmacy visit when collecting inhalation medication for their asthma, COPD, bronchitis, or another indication. More specifically, individuals were offered SARA when they were dispensed medication for obstructive airway disease within the R03 class of drug, according to the use of the Anatomical Therapeutic Chemical (ATC) classification as developed by the World Health Organization (WHO) [20]. The trigger for pharmacists to invite a patient to participate in SARA was dispensing of an R03 class of drug. However, pharmacists could choose not to offer SARA to patients if they considered them ineligible for participation in SARA, for example, those living in a nursing home or those with very limited digital literacy levels. When interested in SARA, participants were subsequently enrolled in the intervention. Otherwise, they were asked to indicate whether they were not interested in SARA at that specific point in time or would never be interested. Patients' choices were registered by the pharmacists in the pharmacy dispensing database, as well as the date their choices were registered, from here on referred to as the "registration date." If patients wanted to participate, they were enrolled by their pharmacist in the SARA program, after which they were sent a registration confirmation link and were able to start the program accordingly. The process of registering patients' choices in the database was sometimes delayed in daily practice, with pharmacists conducting the formal registration in the pharmacy dispensing database a while after the actual dispensing. Patients who were interested and subsequently agreed to participate in SARA were categorized as SARA participants. Those who were not interested were categorized as control participants. Additionally, patients who collected their inhalation medication and who were never offered SARA were categorized as control participants as well.

The index date was calculated using one of the following two options: (1) if there was an R03-medication dispensing available on the registration date, the registration date was defined as the study index date, or (2) if there was no R03-medication dispensing available on the registration date, the last dispensing date before the registration date was defined as the study index date. Subsequently, the index date was used to calculate the specific period of analysis (ie, the year before as well as the year after implementation of SARA) for each participant. More specifically, the index date was coded as the starting date of the year of analysis after the implementation of SARA. The exact year of analysis before implementation of SARA was coded as the year before the index date, not including the index date itself. Figure 1 presents an example of the index date calculation using option 2, in which case the registration date of the participant was May 31, 2016. As no medication dispensing was available for this date, the last dispensing date before the registration date (ie, May 30, 2016) was taken as the index date. Subsequently, May 30, 2016, was set as the starting date of the year after implementation, whereas the year before implementation of

XSL•FO
RenderX

SARA would cover the period up to and including May 29, 2016.

Figure 1. Operationalized analysis period for the year before the implementation of SARA. Step 1: the index date (ie, May 30, 2016) was used to calculate the specific period of analysis (ie, the day before the index date = the end of the analysis period before the implementation of SARA). Step 2: medication adherence scores were calculated based on the proportion of days covered with the "at least one" method. SARA: Service Apothecary Respiratory Advice; in Dutch, Service Apotheek Raad en Advies.



Study Population

Medication dispensing data from January 2015 to September 2020, from 382 Service Pharmacies located in different regions of the Netherlands, were obtained by information and communications technology service provider NControl. Patients' data in the NControl database are pseudonymized, meaning that their data cannot be directly connected to the natural person (ie, data subject) to whom they belong without the use of additional information, which is kept separately, according to Article 4(5) of the General Data Protection Regulation [21]. NControl provided a selection of this pseudonymized data to the main researchers of the Leiden University Medical Center, including data on patient demographics (ie, year of birth and gender), disease indication (ie, asthma, COPD, bronchitis, or other), the name of the Service Pharmacy, and medication dispensing records with detailed information on the type of the dispensed medication, ATC codes, corresponding dispensing date, amount dispensed, estimated covering days, and prescribed daily dosage. These data were not attributable to specific data subjects; these subjects were represented by personal identifier numbers that could not be used to directly identify a natural person (ie, data subject).

The study population consisted of individuals collecting R03 medication at one of the included 382 Service Pharmacies. Eligibility criteria to be included in the analyses were as follows: (1) patients aged 18 years or older at the time of their first available dispensing date record, (2) patients registered as SARA or control participants (ie, no missing data on SARA participation status), and (3) the time between the index date and the most recent R03-medication dispensing was a maximum of 30 days. This third inclusion criterion was chosen because SARA was always offered during a participant's pharmacy visit for collecting one's R03 medication, and if the time between this dispensing date and the registration date was more than 30 days, we considered it as a potential source of bias. We then presumed that it indicated a significant delay in the pharmacists' registration of SARA participation, which would result in uncertainty about what period to operationalize as "before implementation of SARA" and what period to operationalize as "after implementation of SARA." The fourth eligibility criterion was that patients had to have a disease indication from the pharmacy for asthma or COPD, excluding patients with indications other than asthma or COPD. The fifth and final eligibility criterion was that patients had to have at least one medication dispensing record before starting the 2-year analysis period and at least one record after, in order to ensure complete and up-to-date dispensing data during the analysis period. Besides the five eligibility criteria mentioned above, additional outcome-specific eligibility criteria were in place for the secondary outcomes of medication adherence and antimycotic treatment (see the respective subsections in the Outcome Measures section).

Outcome Measures

Exacerbation Rates

The primary outcome measure was the difference in exacerbation rates over time (ie, before versus after implementation of SARA) between SARA and control participants. The medication dispensing data of short-course

XSL•FO
RenderX

prednisone and prednisolone, hereafter referred to as prednisone, were used to estimate exacerbation rates, as prednisone is prescribed to inhibit the inflammation of exacerbations. Prescriptions with ATC codes H02AB06 (prednisolone) and H02AB07 (prednisone) were used to estimate exacerbation rates. The medication dispensing records were categorized as exacerbations based on the Dutch College of General Practitioners' guidelines for asthma and COPD [22,23], that is, in the case of a dispensing record reflecting a daily dosage of 30 or 40 mg of prednisone for a minimum of 5 days and a maximum of 14 days. The mean number of exacerbations in the year before and after implementation of SARA was summed into a mean total score of exacerbations for each of these analysis periods.

Medication Adherence

One of the secondary outcomes was the difference in medication adherence over time between SARA and control participants. In addition to the general eligibility criteria as mentioned in the Study Population section, another inclusion criterion was formulated for this outcome measure. Participants needed to have at least three dispensing records of R03 medication during the 2-year analysis period in order to exclude fully nonadherent participants and validate the method of calculating medication adherence. In this way, participants with early cessation were excluded from the calculation, and only patients who were pharmacologically treated were included in the analyses.

The WHO definition of adherence was used to operationalize medication adherence, that is, the extent to which a person's behavior corresponds with the agreed-upon recommendations from a health care provider [15]. Studying medication adherence using medication dispensing records of pharmacies is a common method for assessing adherence [24]. Relevant groups of inhalation medication according to the WHO ATC classification included R03 medication, that is, medication for obstructive airway diseases [25]. All medication dispensings of the maintenance R03 medications represented by the following codes were included in the database: R03BA01, R03BA02, R03BA05, R03BA08, R03AK06, R03AK07, R03AK08, R03AK10, R03AK11, R03AL03, R03AL04, R03AL05, R03AL08, R03AL09, R03AC18, R03AC13, R03AC12, R03BB04, R03BB05, R03BB06, and R03BB07. These included ICS, long-acting beta agonists, long-acting muscarinic antagonists, and fixed-dose combinations. Nebulizers were excluded from the analyses.

Medication adherence was operationalized as the proportion of days covered (PDC). The PDC is the preferred method for calculating adherence at a population level and has been operationalized by the Pharmacy Quality Alliance [26]. In this study, the PDC was defined as the ratio of the number of days that a patient had medication available for at least one type of R03 medication during exactly 1 analysis year (ie, before and after the implementation of SARA, respectively) to the total number of days that the patient was dispensed the medication during that same period (ie, estimated covering days of the medication). Hence, the PDC reflected the proportion of days that the individual had at least one type of R03 medication available during the corresponding year of analysis.

More specifically, the "at least one" method was applied, which is a standardized method for measuring concurrent adherence to multiple related medications, in this case, the broad class of R03 medications. When the estimated coverage period of dispensed R03 medication did not precisely cover all 365 days of the 1-year analysis period, the data from the first available R03-medication dispensing record before or after the analysis period, respectively (ie, depending on whether it concerned the analysis period before or after implementation of SARA), was used to determine the coverage of days belonging to the analysis period. Two assumptions were made in this process: (1) participants would only come to collect R03 inhalation medication once they finished their previously collected medication; in this way, the stock was not taken into account, and (2) participants would fully adhere to the prescribed dosage from the dispensing date onward until the end of the prescribed covering days. The above-mentioned methods and flow of this calculation of the PDC is presented in Figure 1.

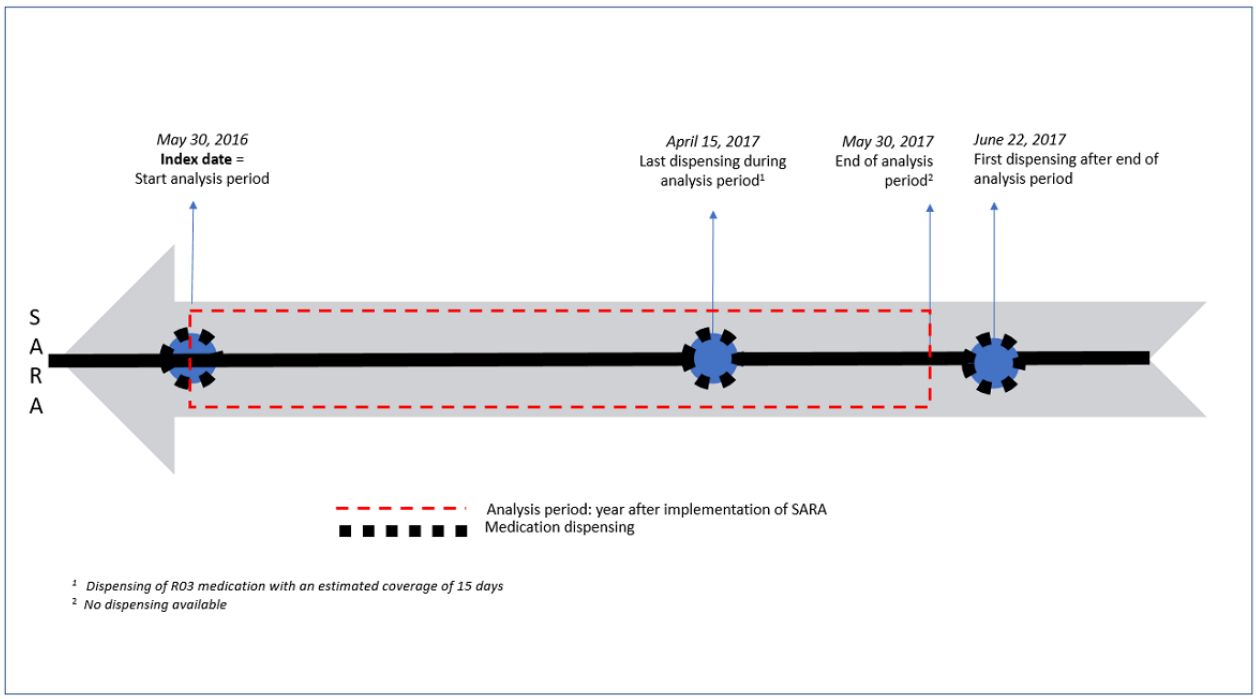
Looking at Figure 1, a patient's analysis period before implementation of SARA started on May 30, 2015, but no medication dispensing was available for this date. The last dispensing before the start of this analysis period was on May 20, 2015, with an estimated coverage of 15 days, that is, the period of May 20 to June 3, 2015. The period from June 4, 2015, onward to the day before the next medication dispensing on June 18, 2015 (ie, the period from June 4 up to and including June 17, 2015), would be coded as "not covered." Similarly, looking at Figure 2, for example, a patient's analysis period after implementation of SARA ended on May 30, 2017, and the last available dispensing record concerned a dispensing of R03 medication on April 15, 2017, with an estimated coverage of 15 days. This last dispensing thus covered the period from April 15 to 29, 2017. No records of dispensing data were available for the period from April 30 to May 30, 2017; hence, this period was coded as "not covered." Medication adherence scores could range from 0 to 100, where 100 would reflect all 365 days of the analysis year being covered.

As it is commonly a cutoff point for good adherence, the PDC of 0.8 was used [26,27]. If it could not be determined whether or not a patient was covered by medication for a specific day of the year, a PDC could not be calculated; this would be considered a missing value.

The analyses were performed separately for *new users* and *chronic users* of R03 medication because different behaviors were expected for these two groups [28]. New users refer to participants starting with inhalation medication, operationalized as zero R03 dispensing records in the year before the index date. Chronic users refer to those already using R03 medication, operationalized as having at least one R03 dispensing record in the year before the index date.

XSL•FO
RenderX

Figure 2. Operationalized analysis period for the year after the implementation of SARA. Step 1: the index date (ie, May 30, 2016) was used to calculate the specific period of analysis (ie, index date = the start of the analysis period after the implementation of SARA). Step 2: medication adherence scores were calculated based on the proportion of days covered with the "at least one" method. SARA: Service Apothecary Respiratory Advice; in Dutch, Service Apotheek Raad en Advies.



Antimycotic Treatment

Antimycotic treatment was operationalized as the difference over time in dispensed antimycotics between the SARA and control participants. The prevalence of oral candidiasis, potentially associated with ICS use, was estimated based on dispensing data of antimycotics in the subpopulation of participants who were dispensed ICS during the analysis period. Therefore, an additional inclusion criterion was formulated: participants needed to have at least one medication dispensing record of ICS (ie, ATC code R03BA01, R03BA02, R03BA05, or R03BA08) during the analysis period. If a participant was dispensed antimycotics (ie, ATC code J02AC01 [fluconazole], J02AC02 [itraconazole], A07AA02 [nystatin], A07AA07 [amphotericin B], or A07AC01 [miconazole]) during the analysis period, the outcome was coded as 1 (“yes”); if not, the outcome was coded as 0 (“no”). Next, the percentage of participants with an antimycotic dispensing was calculated per study condition and subsequently compared before and after the implementation of SARA.

Statistical Analyses

The study population characteristics, per outcome measure, were summarized by descriptive statistics: means and SDs for continuous variables, and counts and percentages for dichotomous and categorical variables. Potential differences between SARA and control participants were analyzed using *t* tests for normally distributed continuous variables and chi-square tests for categorical variables.

Differences in the outcome measures of exacerbation rates and medication adherence were analyzed using independent *t* tests to examine potential differences between the two study groups over time. More specifically, difference scores were calculated per patient by subtracting the outcome scores (ie, exacerbation rates and PDC sores for the subpopulation of chronic users of inhalation medication) of the year before implementation of SARA and the scores in the year after. Additionally, for the subpopulation of new users of inhalation medication, an independent-samples *t* test was conducted to investigate differences in medication adherence in the year after implementation of SARA between SARA and control participants. The potential effects of covariates (ie, age and gender) were tested by means of analysis of covariance. The results of these analyses were only presented in the case of significant effects of covariates.

A mixed-effects logistic regression was conducted to analyze the change over time between the two study groups regarding the relative number of patients who were dispensed antimycotics. In this analysis, an interaction term of time (ie, before and after the index date) and the study condition (ie, SARA vs control) was included to analyze the change over time across groups. The potential effects of covariates (ie, age and gender) were tested by adding those as interaction terms to the model. The results of these analyses were only presented in the case of significant effects of covariates.

All analyses were conducted in the total population consisting of both patients with asthma and those with COPD. For exploratory purposes, separate analyses for the subpopulations of patients with asthma and those with COPD were conducted. For all the analyses, a significance level of $P \leq .05$ was used, and a Cohen *d* was calculated to measure effect sizes. All analyses were conducted in SPSS Statistics for Windows (version 25.0; IBM Corp).

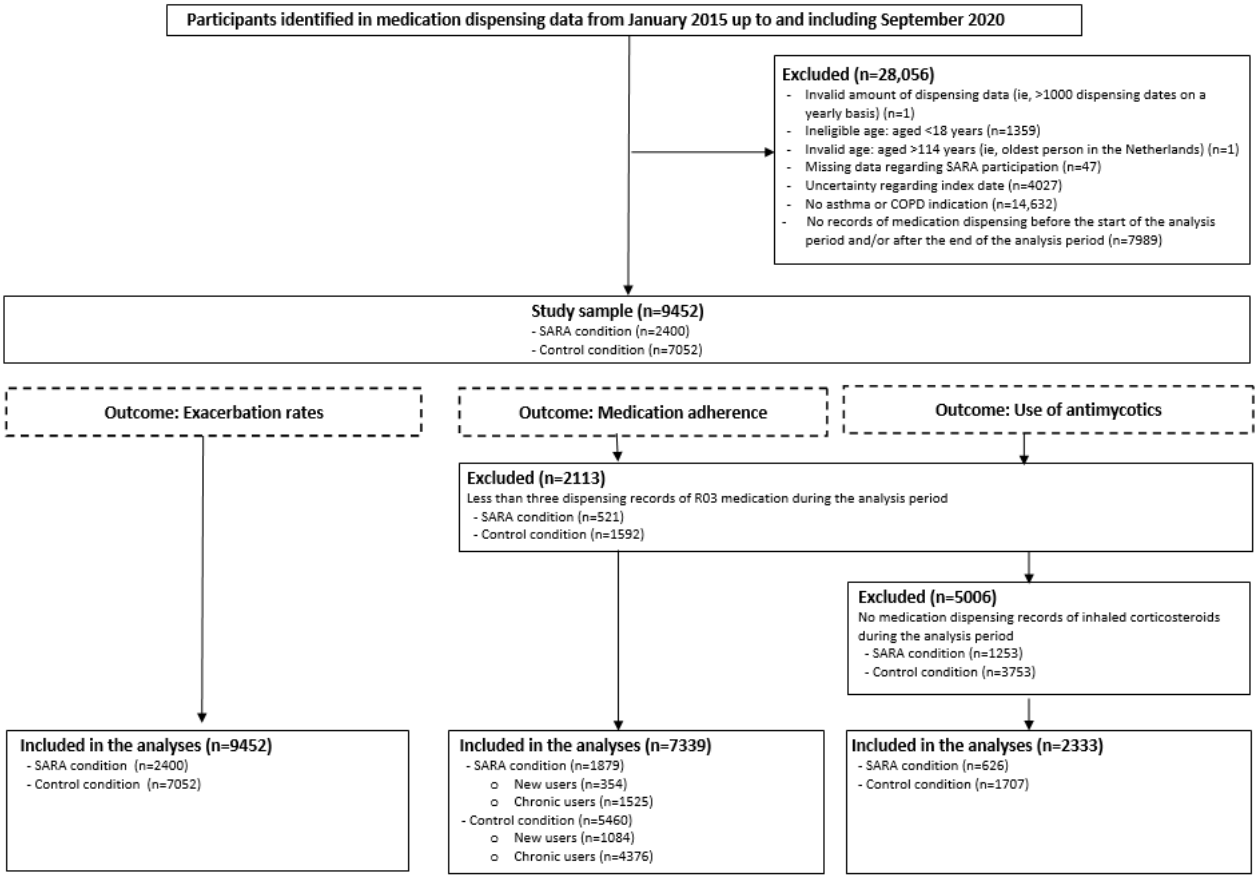
XSL•FO
RenderX

Results

Study Population

The flow of included patients is presented in Figure 3. The total study population comprised of 9452 individuals with either asthma or COPD. Of those, 25.39% (n=2400) were enrolled in SARA, 25.73% (n=2432) indicated that they were not interested in using SARA, and 48.88% (n=4620) were not invited to participate or indicated that they did not want to start using SARA at that particular moment in time. As the inclusion criteria differed per outcome measure, the demographic characteristics are presented separately for each outcome measure (Table 1). Overall, the mean age of the study population was 60.8 (SD 15.0) years, and almost two-thirds of the study population were female. In all the different subpopulations, the mean age of patients using SARA was significantly lower than that of patients in the control group. In general, there was a significantly larger proportion of men in the control group as compared to the SARA group. Table S1 in Multimedia Appendix 2 shows the characteristics of the study samples separately per disease indication for asthma and COPD.

Figure 3. Flow of participants for the different outcome measures and corresponding analyses. SARA: Service Apothecary Respiratory Advice; in Dutch, Service Apotheek Raad en Advies.



XSL•FO
RenderX

Table 1. Demographic characteristics of the study populations analyzed for the different outcome measures.

Outcome measure and characteristics	SARA[a] participants (n=2400)	Control participants (n=7052)	Total population (N=9452)	*P* value[b]
Exacerbation rate				
Gender, n (%)				
Male	882 (36.75)	2851 (40.43)	3733 (39.49)	.002
Female	1504 (62.67)	4173 (59.17)	5677 (60.06)	
Unknown	14 (0.58)	28 (0.40)	42 (0.44)	
Age (years), mean (SD)	57.7 (13.8)	61.9 (15.3)	60.8 (15.0)	<.001
Medication adherence				
Total group (SARA: n=1879; control: n=5460; total: n=7339)				
Gender, n (%)				
Male	693 (36.88)	2200 (40.29)	2893 (39.42)	.01
Female	1175 (62.53)	3239 (59.32)	4414 (60.14)	
Unknown	11 (0.58)	21 (0.38)	32 (0.44)	
Age (years), mean (SD)	60.9 (13.4)	65.1 (14.5)	64.0 (14.4)	<.001
Subgroup: new users[c] (SARA: n=354; control: n=1084; total: n=1438)				
Gender, n (%)				
Male	128 (36.16)	420 (38.74)	548 (38.11)	.38
Female	225 (63.56)	658 (60.70)	883 (61.40)	
Unknown	1 (0.28)	6 (0.55)	7 (0.49)	
Age (years), mean (SD)	59.4 (14.2)	62.7 (16.5)	61.9 (16.0)	.002
Subgroup: chronic users[d] (SARA: n=1525; control: n=4376; total: n=590)				
Gender, n (%)				
Male	565 (37.05)	1780 (40.68)	2345 (39.74)	.02
Female	950 (62.29)	2581 (58.98)	3531 (59.84)	
Unknown	10 (0.66)	15 (0.34)	25 (0.42)	
Age (years), mean (SD)	61.3 (13.1)	65.7 (14.0)	64.6 (13.9)	.04
Antimycotic treatment (SARA: n=626; control: n=1707; total: n=2333)				
Gender, n (%)				
Male	196 (31.31)	612 (35.85)	808 (34.63)	.04
Female	428 (68.37)	1090 (63.85)	1518 (65.07)	
Unknown	2 (0.32)	5 (0.29)	7 (0.30)	
Age (years), mean (SD)	55.1 (14.2)	59.0 (16.2)	58.0 (15.8)	<.001

[a]SARA: Service Apothecary Respiratory Advice; in Dutch, Service Apotheek Raad en Advies.

[b]*P* values represent comparisons between the SARA group and the control group; for characteristics with multiple subcategories (ie, gender), values for the group are reported in the top row of the group.

[c]New users are participants with zero R03 dispensing records in the year before the index date.

[d]Chronic users are participants with at least one R03 dispensing record in the year before the index date.

Exacerbation Rates

In the year before the implementation of SARA, 63.00% (5955/9452) of the total study population had 0 exacerbations (range 0-12). In the year after the implementation of SARA, 56.00% (5293/9452) of the study population had 0 exacerbations (range 0-14). In both study groups, the mean rate of exacerbations was higher in the year after the implementation of SARA (SARA: mean 0.73; control: mean 0.82) than in the year before (SARA: mean 0.68; control: mean 0.67). Yet, as shown in Table 2, there was a significant difference between the SARA and control participants regarding the exacerbation rate over time, showing that the increase in exacerbations was significantly less in the SARA group (*P*=.002). The results of

XSL•FO
RenderX

the exploratory analyses are presented in Table S2 in Multimedia Appendix 2. In both participants with asthma and those with COPD, the mean exacerbation rate increased over time in both the SARA group (asthma: mean increase 0.07; COPD: mean increase 0.03) and the control group (asthma: mean increase 0.17; COPD: mean increase 0.12). As presented in Table S2 in Multimedia Appendix 2, among the asthma participants, the difference in exacerbation rates differed significantly between study groups (P=.003), indicating that SARA participants had a significantly lower increase in exacerbation rates over time in comparison to the control participants. No significant difference between the SARA and control participants was found in the COPD population regarding the change in exacerbation rate over time (Table S2 in Multimedia Appendix 2).

Table 2. Outcome results in terms of exacerbation rates.

Study group and periods[a]	Exacerbation rate, mean (SD)	Difference score[b]	Participants (N=9452), n (%)	*t* test (*df*)[c]	*P* value[c]	95% CI[c]	Cohen *d*[c]
Control							
1 year before	0.67 (1.2)		7052 (74.61)	3.10 (9450)	.002	0.037-0.163	0.06
1 year after	0.82 (1.3)	0.15	7052 (74.61)				
SARA							
1 year before	0.68 (1.2)		2400 (25.39)				
1 year after	0.73 (1.2)	0.05	2400 (25.39)				

[a]The study periods were 1 year before and 1 year after the implementation of SARA (Service Apothecary Respiratory Advice; in Dutch, Service Apotheek Raad en Advies).

[b]The difference score was calculated as the exacerbation rate the year after SARA minus the rate the year before SARA; values are only reported in the "1 year after" rows.

[c]Statistics comparing study groups are reported only in the top row of values.

Medication Adherence

In both study groups, the mean PDC in the subpopulation of chronic users was higher in the year after compared to the year before implementation of SARA for both SARA participants (after: mean 77.26; before: mean 70.53) and control participants (after: mean 77.77; before: mean 73.29). However, there was a significant difference in change over time between the SARA and the control groups, showing that the increase in medication adherence was significantly higher in the SARA group (Table 3).

The exploratory results, repeating the analyses for the chronic user subgroup of participants with asthma and participants with COPD, are presented in Table S3 in Multimedia Appendix 2. For patients with asthma who were chronic users, there was an increase in medication adherence with no significant difference between the SARA and control participants. Gender was found to be a significant covariate for the patients with COPD who were chronic users. Splitting the analyses for men and women within this subpopulation showed that the increase in medication adherence for men was significantly higher for SARA participants than for control participants. For women, there was no significant difference between SARA and control participants over time in terms of medication adherence.

When comparing medication adherence in the year after implementation of SARA between the study groups for new users with COPD, this population showed significantly higher medication adherence in the SARA group as compared to the control group (Table S4 in Multimedia Appendix 2). No significant difference between the study groups was found in the subpopulation of new users with asthma.

XSL•FO
RenderX

Table 3. Outcome results in terms of medication adherence among the chronic user subpopulation.

Study group and periods[a]	PDC[b], mean (SD)	Days covered, mean (SD)	Difference score[c]	Participants (n=5888),n (%)	*t* test (*df*)[d]	*P* value[d]	95% CI[d]	Cohen *d*[d]
Control								
1 year before	73.29 (28.3)	267.50 (103.4)		4368 (74.18)	–2.74 (5886)	.01	–3.856 to –0.839	–0.07
1 year after	77.77 (25.2)	283.86 (91.8)	4.48	4368 (74.18)				
SARA								
1 year before	70.53 (29.8)	257.45 (108.6)		1520 (25.82)				
1 year after	77.26 (25.0)	282.01 (91.1)	6.73	1520 (25.82)				

[a]The study periods were 1 year before and 1 year after the implementation of SARA (Service Apothecary Respiratory Advice; in Dutch, Service Apotheek Raad en Advies).

[b]PDC: proportion of days covered.

[c]The difference score was calculated as the PDC 1 year after SARA minus 1 year before SARA; values are only reported in the "1 year after" rows.

[d]Statistics comparing study groups are reported only in the top row of values.

Antimycotic Treatment

As shown in Table 4, the relative mean number of participants who had been dispensed antimycotics was higher after the implementation of SARA as compared to the year before for both SARA participants (6.4% vs 5.4%) and control participants (6.1% vs 4.7%). Results showed no significant differences in the relative number of participants who had been dispensed both ICS and antimycotics between the SARA and control groups (*P*=.82). Additionally, in the exploratory results, no significant differences were found with respect to antimycotic treatment over time between SARA and the control participants in the subgroups of participants with asthma and COPD (Table S5 in Multimedia Appendix 2).

Table 4. Results of the mixed-effects logistic regression regarding dispensed antimycotics among participants who were dispensed ICS.

Study group and periods[a]	Participants dispensed antimycotics, n (%)	Participants dispensed ICS[b], n (%)	*t* test (*df*)[c]	*P* value[c]	95% CI[c]	Cohen *d*[c]
Control (n=1707)						
1 year before	80 (4.69)	1707 (73.17)	0.23 (4662)	.82	–0.461 to 0.584	0
1 year after	104 (6.09)	1707 (73.17)				
SARA (n=626)						
1 year before	34 (5.43)	626 (26.83)				
1 year after	40 (6.39)	626 (26.83)				

[a]The study periods were 1 year before and 1 year after the implementation of SARA (Service Apothecary Respiratory Advice; in Dutch, Service Apotheek Raad en Advies).

[b]ICS: inhaled corticosteroids; percentages are based on total participants in both groups (n=2333).

[c]Statistics comparing study groups are reported only in the top row of values.

Discussion

Principal Findings

This study investigated the effectiveness of the pharmacy-based eHealth intervention SARA by comparing pharmacy dispensing data between SARA and control participants over time before and after the implementation of SARA. The results showed a smaller increase in exacerbation rates over time for SARA participants as compared to control participants. Furthermore, in the SARA group, chronic users of inhalation medication had a significantly larger increase in medication adherence over time as compared to control participants. Finally, no significant differences between the study groups were found with respect to antimycotic treatment over time.

Although the observational data do not entirely allow for causal conclusions, the significantly smaller increase in exacerbation rates over time among SARA participants may suggest a beneficial effect of SARA. Earlier clinical intervention studies comprising a behavioral intervention and integrated disease management program have also found positive effects on exacerbation rates among asthma participants [29,30]. Yet, SARA has the potential to help control exacerbations in a less invasive and less time-consuming way; this is potentially apparent in reduced material and immaterial costs, such as less time spent conducting follow-ups by pharmacists.

The results regarding medication adherence showed that chronic users of inhalation medication in the SARA group had a significantly higher increase in medication adherence as

XSL•FO
RenderX

compared to control participants. This finding aligns with a previous meta-analysis examining eHealth strategies to improve medication adherence in ICS users [18]. However, it is essential to note that the mean medication adherence was lower for SARA participants than control participants, both before and after the implementation of SARA. A potential explanation is selection bias. Patients with more severe symptoms may have been more likely to be invited to participate in the SARA intervention by the pharmacists because they may visit the pharmacy more often, and patients with more severe symptoms typically show lower medication adherence [10]. On the other hand, patients with more severe symptoms may simply have been more interested in participating in the SARA intervention considering their higher disease burden, which may have, in turn, biased the results. The finding that new users of inhalation medication generally had lower medication adherence scores than chronic users emphasizes the importance of analyzing those two patient groups separately, as they appear to have different adherence patterns.

An interesting difference between men and women was found in the analysis of patients with COPD who were chronic users of inhalation medication. The results suggested that men within this subpopulation benefitted more from SARA (ie, increased medication adherence in comparison to controls) than women (ie, no differences between SARA and control participants). Little research is available on gender-associated differences in response to self-management interventions. A narrative review did discuss some evidence that women have more trouble with using inhalation medication correctly [31]. Furthermore, a systematic review discussed mixed results regarding gender-associated differences in response to pulmonary rehabilitation [32]. Thus, there appears to be some evidence of gender-associated differences that could explain our finding; however, more research is needed to investigate individual differences of patients regarding adherence based on their characteristics, beliefs, and attitudes to adherence.

With respect to antimycotic treatment for oral candidiasis in a subpopulation of ICS users, no difference was found between the study groups over time. These results should be interpreted carefully because the included sample was small, possibly limiting the power to detect statistical significance. To our knowledge, this was the first study that analyzed the effect of an eHealth intervention for patients with asthma and COPD on antimycotic treatment. The exploratory analyses showed a more favorable course of exacerbation rates over time for SARA versus control participants in the subpopulation of patients with asthma. This effect was not found in the subpopulation of patients with COPD. Our results are in line with previous research investigating a clinic-based intervention aiming to improve inhaler techniques, which only showed a positive effect in patients with asthma but not in patients with COPD [33]. It might be that patients with asthma benefit more from the educational intervention elements than patients with COPD. Alternatively, it might be due to more difficulties in managing COPD symptoms as the disease progresses, or the fact that COPD often results from smoking and that smoking cessation is quite challenging.

Furthermore, exploratory results showed that new users of inhalation medication had higher medication adherence in the year after SARA implementation among SARA participants as compared to control participants, but only in the subpopulation of patients with COPD and not in patients with asthma. In addition, patients with COPD generally had higher medication adherence than patients with asthma. This is in line with literature showing that patients with COPD generally have better adherence rates than patients with asthma, and there are multiple explanations for this [34]. First, it can be related to the different disease courses; in patients with asthma, the use of medication can, for example, be more dependent on the season than in patients with COPD [34]. Second, patients with COPD generally experience more consistent and severe disease symptoms [34]. Third, older age is associated with being more adherent, and patients with COPD are generally older than patients with asthma [35].

The findings of this study should be interpreted in light of several strengths and limitations. A major strength of this study pertains to the large amount of pharmacy dispensing data stemming from thousands of patients from hundreds of pharmacies geographically located throughout different areas in the Netherlands. This is likely to benefit the generalizability of the study results. In addition, these kinds of trials can contribute to external validity more than a randomized controlled trial [36]. Furthermore, the data set allowed for longitudinal research comparing data before and after the implementation of SARA with continuous enrollment of patients instead of during a specific period of time. For that reason, the impact of seasonal effects or national guidelines are expected to have been limited. Regarding the study limitations, the study results were based on retrospective pharmacy dispensing data. This design has several limitations, such as data that were not originally designed to answer specific research questions. Indeed, pharmacy dispensing data were limited in terms of not providing information about actual usage of the medication, more specifically if, when, and how often dispensed medication was used. Still, dispensing data are commonly used as a proxy for medication adherence [37,38]. Future studies could consider including other measures of medication adherence, for example, self-reports of medication use, smart inhaler devices, or measurements of metabolite levels [37,39-41]. Another study limitation is related to the commonly used “at least one” method to calculate the PDC as an indicator of medication adherence. This methodology does not take into account potential overuse of medication. Besides, the PDC can slightly differ when using the highest stock records of medication [42,43]. In addition, our assumption when interpreting the results was that better medication adherence was a consequence of better self-management skills. However, it could be the case that lower medication adherence is a sign of good self-management, as the patients may only take their medication when actually needed. This is an interesting topic for future research. In addition, future research could combine multiple methods to calculate medication adherence to provide a more comprehensive picture of this outcome measure. A recent publication by Menditto et al [43] proposes measuring persistence as a pragmatic and informative measure of medication adherence behaviors, which would allow for

XSL•FO
RenderX

benchmarking of adherence strategies. Such strategies would thus facilitate cross-study comparisons and might help to identify a gold standard for calculating medication adherence [37,38,44]. This pragmatic trial only allowed for adherence measures based on pharmacy dispensing data. More specifically, the PDC is a preferred method of assessing medication adherence in case of treatment with multiple types of medications. An alternative metric such as the medication possession ratio (MPR) would be unable to cover multiple medication treatments since its numeration is the sum of days supplied in the period. In case of multiple medications, the MPR has to be averaged for each individual medication, leading to skewed results with possibilities of invalid ratios over 100%. So there are biases, such as not taking into account overuse and stockpiling, but using the PDC was a well-considered choice.

Another study limitation was that it was unknown what kind or intensity of support was offered by pharmacists. Hence, different pharmacists may have provided different types of support to patients. Even though this is inherent to tailored interventions, it would be worthwhile to investigate *what* type of support has the most beneficial effect. This also includes identifying *when*, *how*, and *how much* support should be offered. Addressing these questions can help to develop and strengthen evidence-based interventions [45]. A final study limitation that needs to be mentioned was the difference in demographic characteristics between the SARA and control participants. More specifically, SARA participants were generally younger and more often female. Even though such differences are not unusual in nonrandomized studies, they may have created selection bias [46]. However, SARA was, in principle, offered to all kinds of participants with varying degrees of symptoms. Therefore, the possibly biased selection of participants in the SARA group is likely to be representative of the group of potential future users of eHealth interventions for these groups. An important aspect to also take into account is that the questionnaire for the SARA intervention might increase patients' awareness for medication adherence, but it is unlikely that this strongly affected adherence behavior directly. In future research, this could be something to take into account. More research is needed to draw firm conclusions on the effectiveness of SARA. A randomized controlled trial is needed to allow causal conclusions, which can then be used for a cost-effectiveness analysis as well, where, next to pharmacy dispensing data, other data can be collected, such as the following: (1) other sources that measure medication adherence, (2) objective data regarding exacerbation rates, (3) the actual and correct use of inhalation medication, and (4) health system characteristics that may impact adherence (eg, patient-provider interaction quality and procedural elements) [46]. In addition, qualitative research would allow for more insight into user experiences and could subsequently be used to optimize the intervention. In parallel, it would be interesting to investigate patients' acceptability and effectiveness of the different components of the SARA intervention (eg, education materials and online support by a pharmacist). Also, it would be worthwhile to get a better understanding of the pharmacist perspective, for instance, what is their attitude toward eHealth in general and SARA specifically, what is the usability of SARA, and how is SARA used in the pharmacy (ie, does it add to the efficiency of care processes?)? Another recommendation for future research is to analyze the long-term effectiveness of SARA.

This research shows that SARA has the potential to help patients in decreasing exacerbation rates and improving medication adherence. Before large-scale implementation, it would be valuable to investigate both the patient and pharmacist perspective more thoroughly, both quantitatively and qualitatively. In this way, the full potential of the intervention can be maximized, making sure the intervention fits the needs and preferences of both of these stakeholders. Implementation barriers and facilitators can be investigated and taken into account when considering implementation strategies, such as integration of SARA into the workflow of pharmacists as well as the capacities of pharmacists to offer tailored follow-up care [47,48].

Conclusions

This was the first study that assessed the effectiveness of a multi-component eHealth intervention stimulating correct use of medication. The results suggest that such an intervention has the potential to decrease exacerbation rates and improve medication adherence. This could subsequently have important clinical implications and lead to better patient outcomes and potentially reduced health care costs.

Acknowledgments

Funding for this study was provided by Service Apotheek Beheer BV; this funding body had no role in the analysis or interpretation of the findings in this paper.

Conflicts of Interest

RB is an employee of Service Apotheek Beheer BV, who also funded the study but had had no role in the analysis or interpretation of the findings in this paper. During the last 3 years, JFMvB has received consultancy fees, research funding, or both from Aardex; AstraZeneca; Chiesi; COST Action CA19132 "ENABLE," funded by COST (European Cooperation in Science and Technology); GSK; Novartis; Nutricia; Pfizer; Pill Connect; Teva; and Trudell Medical to consult, give lectures, provide advice, and conduct independent research, all unrelated to this study and all paid to his institution, the University Medical Center Groningen.

Multimedia Appendix 1

XSL•FO
RenderX

The 7-item questionnaire included in the SARA eHealth intervention. SARA: Service Apothecary Respiratory Advice; in Dutch, Service Apotheek Raad en Advies.
[DOCX File , 15 KB - jmir_v24i6e32396_app1.docx]

Multimedia Appendix 2
Results of exploratory analyses.
[DOCX File , 41 KB - jmir_v24i6e32396_app2.docx]

References

1. The Global Asthma Report 2018. Auckland, New Zealand: Global Asthma Network; 2018. URL: http://globalasthmareport.org/resources/Global_Asthma_Report_2018.pdf [accessed 2022-05-30]
2. Chronic obstructive pulmonary disease (COPD). World Health Organization. 2022 May 19. URL: https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd) [accessed 2022-05-30]
3. Ehteshami-Afshar S, FitzGerald JM, Doyle-Waters MM, Sadatsafavi M. The global economic burden of asthma and chronic obstructive pulmonary disease. Int J Tuberc Lung Dis 2016 Jan;20(1):11-23. [doi: 10.5588/ijtld.15.0472] [Medline: 26688525]
4. Quaderi SA, Hurst JR. The unmet global burden of COPD. Glob Health Epidemiol Genom 2018;3:e4 [FREE Full text] [doi: 10.1017/gheg.2018.1] [Medline: 29868229]
5. Prince MJ, Wu F, Guo Y, Gutierrez Robledo LM, O'Donnell M, Sullivan R, et al. The burden of disease in older people and implications for health policy and practice. Lancet 2015 Mar 07;385(9967):549-562. [doi: 10.1016/S0140-6736(14)61347-7] [Medline: 25468153]
6. van Boven JFM, Chavannes N, van der Molen T, Rutten-van Mölken MPMH, Postma M, Vegter S. Clinical and economic impact of non-adherence in COPD: A systematic review. Respir Med 2014 Jan;108(1):103-113 [FREE Full text] [doi: 10.1016/j.rmed.2013.08.044] [Medline: 24070566]
7. Mäkelä MJ, Backer V, Hedegaard M, Larsson K. Adherence to inhaled therapies, health outcomes and costs in patients with asthma and COPD. Respir Med 2013 Oct;107(10):1481-1490 [FREE Full text] [doi: 10.1016/j.rmed.2013.04.005] [Medline: 23643487]
8. Dekhuijzen R, Lavorini F, Usmani O, van Boven JFM. Addressing the impact and unmet needs of nonadherence in asthma and chronic obstructive pulmonary disease: Where do we go from here? J Allergy Clin Immunol Pract 2018;6(3):785-793. [doi: 10.1016/j.jaip.2017.11.027] [Medline: 29339126]
9. George M, Bender B. New insights to improve treatment adherence in asthma and COPD. Patient Prefer Adherence 2019;13:1325-1334 [FREE Full text] [doi: 10.2147/PPA.S209532] [Medline: 31534319]
10. Murphy AC, Proeschal A, Brightling CE, Wardlaw AJ, Pavord I, Bradding P, et al. The relationship between clinical outcomes and medication adherence in difficult-to-control asthma. Thorax 2012 Aug;67(8):751-753. [doi: 10.1136/thoraxjnl-2011-201096] [Medline: 22436168]
11. Rubin BK. What does it mean when a patient says, "my asthma medication is not working?". Chest 2004 Sep;126(3):972-981. [doi: 10.1378/chest.126.3.972] [Medline: 15364781]
12. Lavorini F, Magnan A, Dubus JC, Voshaar T, Corbetta L, Broeders M, et al. Effect of incorrect use of dry powder inhalers on management of patients with asthma and COPD. Respir Med 2008 Apr;102(4):593-604 [FREE Full text] [doi: 10.1016/j.rmed.2007.11.003] [Medline: 18083019]
13. Chorão P, Pereira A, Fonseca J. Inhaler devices in asthma and COPD--An assessment of inhaler technique and patient preferences. Respir Med 2014 Jul;108(7):968-975 [FREE Full text] [doi: 10.1016/j.rmed.2014.04.019] [Medline: 24873873]
14. van Boven JF, Trappenburg JC, van der Molen T, Chavannes NH. Towards tailored and targeted adherence assessment to optimise asthma management. NPJ Prim Care Respir Med 2015 Jul 16;25:15046 [FREE Full text] [doi: 10.1038/npjpcrm.2015.46] [Medline: 26181850]
15. Sabaté EE. Adherence to Long-Term Therapies: Evidence for Action. Geneva, Switzerland: World Health Organization; 2003. URL: http://apps.who.int/iris/bitstream/handle/10665/42682/9241545992.pdf?sequence=1 [accessed 2022-05-30]
16. Melani AS, Bonavia M, Cilenti V, Cinti C, Lodi M, Martucci P, Gruppo Educazionale Associazione Italiana Pneumologi Ospedalieri. Inhaler mishandling remains common in real life and is associated with reduced disease control. Respir Med 2011 Jun;105(6):930-938 [FREE Full text] [doi: 10.1016/j.rmed.2011.01.005] [Medline: 21367593]
17. Normansell R, Kew KM, Stovold E. Interventions to improve adherence to inhaled steroids for asthma. Cochrane Database Syst Rev 2017 Apr 18;4:CD012226 [FREE Full text] [doi: 10.1002/14651858.CD012226.pub2] [Medline: 28417456]
18. Jeminiwa R, Hohmann L, Qian J, Garza K, Hansen R, Fox BI. Impact of eHealth on medication adherence among patients with asthma: A systematic review and meta-analysis. Respir Med 2019 Mar;149:59-68 [FREE Full text] [doi: 10.1016/j.rmed.2019.02.011] [Medline: 30803887]
19. Consultvoering. KNMP. URL: https://www.knmp.nl/dossiers/consultvoering [accessed 2022-05-30]
20. WHO Collaborating Centre for Drug Statistics Methodology. Guidelines for ATC Classification and DDD Assignment 2022. Oslo, Norway: WHO Collaborating Centre for Drug Statistics Methodology; 2021. URL: https://www.whocc.no/filearchive/publications/2022_guidelines_web.pdf [accessed 2022-05-30]

XSL•FO
RenderX

21. Garritsen F, van den Heuvel JM, Bruijnzeel-Koomen CAFM, Maitland-van der Zee AH, van den Broek MPH, de Bruin-Weller MS. Use of oral immunosuppressive drugs in the treatment of atopic dermatitis in the Netherlands. J Eur Acad Dermatol Venereol 2018 Aug;32(8):1336-1342. [doi: 10.1111/jdv.14896] [Medline: 29485224]
22. Smeele I, Barnhoorn MJM, Broekhuizen BDL, Chavannes NH, In 't Veen JCCM, Van der Molen T, et al. NHG-Standaard Astma bij volwassenen (derde herziening). Huisarts Wet 2015;58(3):142-154.
23. Snoeck-Stroband JB, Schermer TRJ, Van Schayck CP, Muris JW, Van der Molen T, In 't Veen JCCM, et al. NHG-Standaard COPD (derde herziening). Huisarts Wet 2015;58(4):198-211.
24. Bourbeau J, Bartlett SJ. Patient adherence in COPD. Thorax 2008 Sep;63(9):831-838 [FREE Full text] [doi: 10.1136/thx.2007.086041] [Medline: 18728206]
25. ATC/DDD Index 2022. WHO Collaborating Centre for Drug Statistics Methodology. URL: https://www.whocc.no/atc_ddd_index/ [accessed 2022-06-01]
26. Nau DP. Proportion of days covered (PDC) as a preferred method of measuring medication adherence. Yahoo. 2012. URL: http://ep.yimg.com/ty/cdn/epill/pdcmpr.pdf [accessed 2022-05-30]
27. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: Impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006 May 16;144(10):742-752 [FREE Full text] [doi: 10.7326/0003-4819-144-10-200605160-00125] [Medline: 16702590]
28. van Boven JFM, Koponen M, Lalic S, George J, Bell J, Hew M, et al. Trajectory analyses of adherence patterns in a real-life moderate to severe asthma population. J Allergy Clin Immunol Pract 2020 Jun;8(6):1961-1969.e6. [doi: 10.1016/j.jaip.2019.12.002] [Medline: 31857262]
29. Baptist AP, Hao W, Song PX, Carpenter L, Steinberg J, Cardozo LJ. A behavioral intervention can decrease asthma exacerbations in older adults. Ann Allergy Asthma Immunol 2020 Mar;124(3):248-253.e3 [FREE Full text] [doi: 10.1016/j.anai.2019.12.015] [Medline: 31877361]
30. Jain V, Allison R, Beck S, Jain R, Mills P, McCurley J, et al. Impact of an integrated disease management program in reducing exacerbations in patients with severe asthma and COPD. Respir Med 2014 Dec;108(12):1794-1800 [FREE Full text] [doi: 10.1016/j.rmed.2014.09.010] [Medline: 25294691]
31. Matera MG, Ora J, Calzetta L, Rogliani P, Cazzola M. Sex differences in COPD management. Expert Rev Clin Pharmacol 2021 Mar;14(3):323-332. [doi: 10.1080/17512433.2021.1888713] [Medline: 33560876]
32. Robles PG, Brooks D, Goldstein R, Salbach N, Mathur S. Gender-associated differences in pulmonary rehabilitation outcomes in people with chronic obstructive pulmonary disease: A systematic review. J Cardiopulm Rehabil Prev 2014;34(2):87-97. [doi: 10.1097/HCR.0000000000000018] [Medline: 24280903]
33. Maricoto T, Madanelo S, Rodrigues L, Teixeira G, Valente C, Andrade L, et al. Educational interventions to improve inhaler techniques and their impact on asthma and COPD control: A pilot effectiveness-implementation trial. J Bras Pneumol 2016 Dec;42(6):440-443. [doi: 10.1590/s1806-37562016000000098]
34. Covvey JR, Mullen AB, Ryan M, Steinke DT, Johnston BF, Wood FT, et al. A comparison of medication adherence/persistence for asthma and chronic obstructive pulmonary disease in the United Kingdom. Int J Clin Pract 2014 Oct;68(10):1200-1208. [doi: 10.1111/ijcp.12451] [Medline: 24797899]
35. Rand C, Nides M, Cowles M, Wise R, Connett J. Long-term metered-dose inhaler adherence in a clinical trial. The Lung Health Study Research Group. Am J Respir Crit Care Med 1995 Aug;152(2):580-588. [doi: 10.1164/ajrccm.152.2.7633711] [Medline: 7633711]
36. Price D, Bateman ED, Chisholm A, Papadopoulos NG, Bosnic-Anticevich S, Pizzichini E, et al. Complementing the randomized controlled trial evidence base. Evolution not revolution. Ann Am Thorac Soc 2014 Feb;11(Supplement 2):S92-S98. [doi: 10.1513/annalsats.201308-276rm]
37. Anghel LA, Farcas AM, Oprean RN. An overview of the common methods used to measure treatment adherence. Med Pharm Rep 2019 Apr;92(2):117-122 [FREE Full text] [doi: 10.15386/mpr-1201] [Medline: 31086837]
38. Torres-Robles A, Wiecek E, Cutler R, Drake B, Benrimoj SI, Fernandez-Llimos F, et al. Using dispensing data to evaluate adherence implementation rates in community pharmacy. Front Pharmacol 2019;10:130 [FREE Full text] [doi: 10.3389/fphar.2019.00130] [Medline: 30863308]
39. Burgess S, Wilson S, Cooper D, Sly P, Devadason S. In vitro evaluation of an asthma dosing device: The smart-inhaler. Respir Med 2006 May;100(5):841-845 [FREE Full text] [doi: 10.1016/j.rmed.2005.09.004] [Medline: 16216485]
40. van Boven JFM, Cushen B, Sulaiman I, Greene G, MacHale E, Mokoka M, et al. Personalising adherence-enhancing interventions using a smart inhaler in patients with COPD: An exploratory cost-effectiveness analysis. NPJ Prim Care Respir Med 2018 Jun 27;28(1):24 [FREE Full text] [doi: 10.1038/s41533-018-0092-8] [Medline: 29950601]
41. Stirratt MJ, Dunbar-Jacob J, Crane HM, Simoni JM, Czajkowski S, Hilliard ME, et al. Self-report measures of medication adherence behavior: Recommendations on optimal use. Transl Behav Med 2015 Dec;5(4):470-482 [FREE Full text] [doi: 10.1007/s13142-015-0315-2] [Medline: 26622919]
42. Arnet I, Kooij MJ, Messerli M, Hersberger KE, Heerdink ER, Bouvy M. Proposal of standardization to assess adherence with medication records: Methodology matters. Ann Pharmacother 2016;50(5):360-368. [doi: 10.1177/1060028016634106]

XSL•FO
RenderX

43. Menditto E, Cahir C, Malo S, Aguilar-Palacio I, Almada M, Costa E, et al. Persistence as a robust indicator of medication adherence-related quality and performance. Int J Environ Res Public Health 2021 May 03;18(9):4872 [FREE Full text] [doi: 10.3390/ijerph18094872] [Medline: 34063641]
44. Malo S, Aguilar-Palacio I, Feja C, Lallana MJ, Rabanaque MJ, Armesto J, et al. Different approaches to the assessment of adherence and persistence with cardiovascular-disease preventive medications. Curr Med Res Opin 2017 Jul;33(7):1329-1336. [doi: 10.1080/03007995.2017.1321534] [Medline: 28422521]
45. Johansson R, Andersson G. Internet-based psychological treatments for depression. Expert Rev Neurother 2012 Jul;12(7):861-869; quiz 870. [doi: 10.1586/ern.12.63] [Medline: 22853793]
46. Khan R, Socha-Dietrich H. Investing in Medication Adherence Improves Health Outcomes and Health System Efficiency: Adherence to Medicines for Diabetes, Hypertension, and Hyperlipidaemia. OECD Health Working Papers No. 105. Paris, France: OECD Publishing; 2018. URL: https://www.oecd-ilibrary.org/deliver/8178962c-en.pdf?itemId=%2Fcontent%2Fpaper%2F8178962c-en&mimeType=pdf [accessed 2022-05-30]
47. Versluis A, van Luenen S, Meijer E, Honkoop PJ, Pinnock H, Mohr DC, et al. SERIES: eHealth in primary care. Part 4: Addressing the challenges of implementation. Eur J Gen Pract 2020 Dec;26(1):140-145 [FREE Full text] [doi: 10.1080/13814788.2020.1826431] [Medline: 33025820]
48. 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 Mar 12;10:21 [FREE Full text] [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]

Abbreviations

ATC: Anatomical Therapeutic Chemical
COPD: chronic obstructive pulmonary disease
COST: European Cooperation in Science and Technology
ICS: inhaled corticosteroids
MEC: Medical Ethics Committee
MPR: medication possession ratio
PDC: proportion of days covered
SARA: Service Apothecary Respiratory Advice; in Dutch, Service Apotheek Raad en Advies
WHO: World Health Organization

Edited by T Leung; submitted 28.07.21; peer-reviewed by YM Huang, R Halkes; comments to author 01.11.21; revised version received 10.02.22; accepted 12.05.22; published 08.06.22.

Please cite as:
Schnoor K, Versluis A, Bakema R, van Luenen S, Kooij MJ, van den Heuvel JM, Teichert M, Honkoop PJ, van Boven JFM, Chavannes NH, Aardoom JJ
A Pharmacy-Based eHealth Intervention Promoting Correct Use of Medication in Patients With Asthma and COPD: Nonrandomized Pre-Post Study
J Med Internet Res 2022;24(6):e32396
URL: https://www.jmir.org/2022/6/e32396
doi:10.2196/32396
PMID:35675120

©Kyma Schnoor, Anke Versluis, Robbert Bakema, Sanne van Luenen, Marcel J Kooij, J Maurik van den Heuvel, Martina Teichert, Persijn J Honkoop, Job F M van Boven, Niels H Chavannes, Jiska J Aardoom. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 08.06.2022. 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

Possible Impact of a 12-Month Web- and Smartphone-Based Program to Improve Long-term Physical Activity in Patients Attending Spa Therapy: Randomized Controlled Trial

Florie Fillol[1], PhD; Ludivine Paris[2], PhD; Sébastien Pascal[1], MSc; Aurélien Mulliez[3], MSc; Christian-François Roques[4], MD; Sylvie Rousset[5], PhD; Martine Duclos[5,6,7], MD, PhD

[1]Biomouv SAS Inc, Paris, France

[2]University of Auvergne, Clermont-Ferrand, France

[3]Biostatistics Unit (Clinical Research and Innovation Direction), University-Hospital Clermont-Ferrand, Clermont-Ferrand, France

[4]Physical and Rehabilitation Medicine, Paul Sabatier University, Toulouse University, Toulouse, France

[5]Human Nutrition Unity, Centre de Recherche en Nutrition Humaine Auvergne, French National Institute for Agriculture, Food and Environment (INRAE), Clermont-Ferrand, France

[6]Department of Sport Medicine and Functional Explorations, University-Hospital Clermont-Ferrand, G. Montpied Hospital, Clermont-Ferrand, France

[7]Unité fonctionnelle de Recherche Médecine, Clermont University, University of Auvergne, Clermont-Ferrand, France

Corresponding Author:
Florie Fillol, PhD
Biomouv SAS Inc
259 Rue Saint Honoré
Paris, 75001
France
Phone: 33 0664421389
Email: florie.fillol@gmail.com

Abstract

Background: Lack of physical activity (PA) and sedentary behaviors are leading risk factors for noncommunicable diseases (NCDs). Web- and smartphone-based interventions are effective in increasing PA in older adults and in patients with NCD. In many countries, spa therapy, commonly prescribed to patients with NCD, represents an ideal context to initiating lifestyle changes.

Objective: This study aimed to evaluate, in patients attending spa therapy, the effectiveness of an intervention combining a face-to-face coaching and, when returning home, a web- and smartphone-based PA program on the achievement of PA guidelines (PAG) 12 months after the end of spa therapy.

Methods: This was a 12-month, prospective, parallel-group randomized controlled trial. Patients were enrolled during spa therapy and randomized 1:1 to intervention or control group who received PA usual advice. From the end of spa therapy, PA, weight, waist circumference, and quality of life of the participants were assessed by phone every 2 months. Primary outcome was meeting PAG (PA ≥600 metabolic equivalent of task) at 12 months. Secondary outcomes were meeting current PAG at 6 months; sedentary time, weight, waist circumference, PA, and quality of life at 6 and 12 months. Objective use data of the web- and smartphone-based PA program were collected. Analytic methods included intention to treat and constrained longitudinal data analyses.

Results: The study sample included 228 participants (n=176, 77.2% females) with a mean age of 62.4 (SD 6.7) years and a mean BMI of 28.2 (SD 4.2) kg/m^2. Approximately 53.9% (123/228) of the participants were retired. No group differences were found for any baseline variable. At 12 months, the proportion of patients achieving PAG was significantly higher in intervention group than in the control group (81% vs 67% respectively, odds ratio 2.34, 95% CI 1.02-5.38; *P*=.045). No difference between intervention and control group was found neither in achieving PAG at 6 months nor for sedentary time, weight, and waist circumference at 6 and 12 months. Regarding quality of life, the physical component subscale score was significantly higher at 12 months in the intervention group than in the control group (mean difference: 4.1, 95% CI 1.9-6.3; *P*<.001). The mean duration use of the program was 7.1 (SD 4.5) months. Attrition rate during the first 2 months was 20.4% (23/113) whereas 39.8% (45/113) of the participants used the program for at least 10 months.

XSL•FO
RenderX

Conclusions: PA increased in both the intervention group and the control group. However, at 12 months, more participants met PAG in the intervention group compared with the controls. This indicates that the web- and smartphone-based program could have maintained PA in the intervention group. In addition, a spa therapy seems to be an ideal time and framework to implement PA education.

Trial Registration: ClinicalTrials.gov NCT02694796; https://clinicaltrials.gov/ct2/show/NCT02694796

(J Med Internet Res 2022;24(6):e29640) doi:10.2196/29640

KEYWORDS

physical activity; spa; mobile phone; older adults; internet; exercise; aged; sedentary behavior; quality of life; follow-up studies

Introduction

Background

Lack of physical activity (PA) and excess sedentary behaviors are now recognized as leading risk factors for noncommunicable diseases (NCDs), such as cardiovascular diseases, chronic obstructive pulmonary disease, cancers, and type 2 diabetes, which, taken together, are the primary causes of death worldwide [1]. Insufficient PA, or physical inactivity, is defined as a level of PA below the recommended 150 minutes of moderate PA per week, and sedentary behaviors are defined as "any waking behaviors characterized by an energy expenditure <1.5 metabolic equivalents of task (METs), while in a sitting, reclining or lying posture" [2]. In 2017, of the 41 million NCD-related deaths, 1.26 million were because of insufficient PA [3]. In 2016, >1 in 4 (27.5%) adults worldwide were physically inactive [4]. In 2015, a national survey in France showed that among adults aged 55 to 74 years, 42.2% of women and 28.4% of men did not achieve the recommendation of 150 minutes of moderate PA per week [5]. Engaging people in healthy behaviors such as stopping tobacco use, reducing alcohol consumption, adopting healthy diets, increasing PA, and limiting sedentary time is crucial to tackling the rise of NCDs [1]. Although the health benefits of PA are widely recognized [6], engaging older adults and those with NCDs in long-term lifestyle modifications is very challenging. Although many studies have shown the benefits of PA interventions on the health of patients with NCDs [7], a decrease in PA adherence is frequently observed in the long term, leading to a loss of the acquired health benefits [8].

To maintain adherence to PA, information and communication technologies appear to be promising tools that provide personalized follow-up, real-time feedback, and recommendations. Recent reviews and meta-analyses have found that web- and smartphone-based interventions are effective in increasing PA in the general population [9,10], in older adults [11,12], and in patients with an NCD [13,14]. However, another systematic review [15] suggested that multicomponent interventions, where the use of an app was one of several intervention components such as physical education, provision of PA equipment, parental education, face-to-face counselling, might be more effective than stand-alone app interventions.

In many countries (continental Europe, Japan, China, South America, and North Africa), a course of spa therapy is accepted as treatment by the health insurance system and is commonly prescribed to patients with chronic diseases such as rheumatic conditions, respiratory diseases, and skin diseases and patients convalescing from cancer, as well as to those who are overweight or obese. In France, the 3-week courses of therapy delivered in spa centers are reimbursed by the national social security. The context and environment of a stay in a spa therapy center have been shown to be conducive to educating patients about their disease and initiating lifestyle changes, including increasing PA, through patient therapeutic education in PA programs [16-20].

Objectives

We hypothesized that an intervention combining individual face-to-face coaching during spa therapy with a subsequent 12-month web- and smartphone-based PA program would improve PA in patients undergoing spa therapy. The main objective of this study was to evaluate the effectiveness of the intervention compared with the usual advice (ie, standard advice on PA provided during spa therapy) on the achievement of PA recommendations 12 months after the end of spa therapy. Secondary outcomes were to evaluate, throughout the 12-month follow-up, at 6 and 12 months, the effectiveness of the intervention on PA, sedentary time, weight, waist circumference, quality of life of the patients, and engagement with the program (the number of performed PA sessions and frequency of use of the program).

Methods

Study Design

This was a 12-month, prospective, parallel-group, open, multicenter, single-blinded randomized controlled trial (RCT) that enrolled patients attending a 3-week spa therapy treatment. It evaluated the effectiveness of individual face-to-face PA coaching during the stay at the spa therapy facility followed by a 12-month web- and smartphone-based PA program, including a connected wrist pedometer and a connected weighing scale. Participants were randomized 1:1 to either the intervention group or the control group. The participants were enrolled in 1 of 8 French spa therapy facilities: Amélie-les-Bains, Bourbon-Lancy, Brides-les-Bains, Le-Boulou, Chaudes-Aigues, Eugénie-les-Bains, Vals-les-Bains, and Vichy.

Participants and Recruitment

Enrollment and follow-ups were conducted between September 2015 and December 2017. Patients were recruited through posters and flyers displayed in spa therapy facilities and spa physicians' surgeries. A PA instructor was allocated to each spa center to prescreen all potential patients and evaluate their eligibility. Spa physicians participating in the study could also

XSL•FO
RenderX

refer their patients to the PA instructor for prescreening. The inclusion criteria were as follows: age of 50 to 79 years, diagnosis of a stabilized chronic disease (cardiovascular disease, obesity, type 2 diabetes, chronic obstructive pulmonary disease, rheumatic conditions, and breast cancer), BMI between >19 kg/m^2 and <35 kg/m^2, undertaking PA for <150 minutes per week, and having smartphone access to the internet. Exclusion criteria included having a cardiac pacemaker, nonstabilized chronic disease, locomotor disability, evolving metastatic cancer, or a contraindication to PA. Eligible participants underwent a medical examination with the spa physician, who after checking that they could safely follow the study protocol, included them in the trial after the participants provided informed consent. Randomization of the participants to the intervention or control group was stratified by gender and center (thermal spa resort) and performed by the spa PA instructor using a centralized secured management system, REDCap (Research Electronic Data Capture; Vanderbilt University).

Intervention and Control

The intervention comprised a 1-hour individual coaching session with a PA instructor during the 3-week spa therapy stay in 1 of the 8 spa care facilities and then access to the web- and mobile-based PA program and associated connected devices for the 12 months following the end of the spa therapy. All PA instructors received the same training and used the same material. The first part of the consultation aimed to introduce or remind the participants of the benefits of PA for health and disease management. The PA instructor provided advice on how to reach the recommended level of PA and examples of PA adapted to the patient's particular condition. Subsequently, the PA instructor presented the automated web and mobile-based PA program (Thermactive, BIOMOUV SAS Inc) together with the use of connected devices (weighing scales and wrist pedometer; Multimedia Appendices 1 and 2). The PA instructor downloaded the mobile app onto the patient's smartphone and showed him or her how to log into the mobile app and connect and use the weighing scale. The PA instructor also explained access to the website and showed participants the main functionalities of the program. The patients were registered in the program by the PA instructor who completed a web-based questionnaire to determine the patient's PA profile: age, weight, height, physical fitness (endurance, strength, flexibility, and balance measured by the PA instructor), PA, joint disabilities, and pathology. The patient also declared his availability for PA sessions, his PA preferences, and his sports material (such as dumbbells, yoga mats, bands, and wrist weights). Participants in the intervention group followed the web- and mobile-based PA program for 12 months from the end of their 3-week stay in the spa therapy center.

The automated program aimed to help patients achieve recommended levels of PA in 2 ways: by proposing personalized and structured PA sessions and by increasing daily PA (number of steps). The PA sessions were automatically generated based on the patient's profile. To generate personalized PA sessions, an algorithm was developed to select and associate exercises from a database of >1500 different exercises. Each exercise was classified according to its nature (aerobic, strengthening, and balance), part of the body concerned (leg, arm, and trunk), exercise intensity, and duration. The algorithm selected exercises appropriate to a patient's physical capacity and availability and constructed a PA session adapted to the patient. Each PA session comprised 3 phases: a 5-minute warm-up period; either 10 to 35 minutes of exercise to develop muscle strength and flexibility or 10 to 50 minutes of endurance during walking or cycling (mixing continuous and intermittent effort); and finally, a 5-minute recovery phase comprising stretching and relaxation or a return to calm after walking sessions. The PA sessions were either automatically compiled videos or PDF files. The program of PA sessions followed international guidelines regarding the number of sessions per week, resting periods, type of exercise (resistance and endurance), duration, and intensity of each exercise [21]. For each participant, their PA sessions evolved during the course of the intervention taking into account the number of PA sessions completed (recorded by the patient) and any difficulty perceived at the end of the PA sessions (collected using a Borg scale [22]). To increase daily PA, the program generated a daily goal of the number of steps to be achieved based on data from the pedometer over 7 consecutive days. The achievement of these goals determined the subsequent goals, and every day, participants received a notification on their mobile app about the achievement of their personal goals. They also received emails about new PA sessions available on the website and emails reminding them whether a PA session had not been performed and inviting them to do it when possible. Participants had the possibility to record or add activities on the mobile app, which were not planned in the program, such as walking, cycling, swimming, or fitness sessions. The website and the mobile app also allowed participants to record their daily PA and amount of sedentary time to visualize their evolution over time.

Patients allocated to the control group received the usual advice on PA and a booklet providing advice and examples of PA suited to their pathology. At the end of the 12-month follow-up period, the patients included in the control group received free connected devices and access to the Thermactive program for 12 months.

Measurements and Follow-up

Data collected during the study and follow-up were recorded using an electronic case report form in a centralized secured management system, REDCap.

Demographic variables of the participants, including sex, age, weight, waist circumference, highest level of formal education (high school or less and higher education), occupation (nonworking [retired or unemployed], manager [artisan or intellectual profession], and employee [employee, intermediate occupation, and worker]), condition treated by spa therapy, medical-surgical and family history, medical treatments, physical fitness, PA, and quality of life were collected at baseline (month 0 [M0]) by the PA instructor. PA was assessed using the validated International Physical Activity Questionnaire (IPAQ)-short version [23]. The IPAQ measures the frequency (days per week) and duration (minutes) of PA during the past 7 days in the following domains: work, transportation, work at home, and leisure activities [23]. Different levels of PA

XSL•FO
RenderX

(walking, moderate, vigorous, and total) were calculated and expressed in METs minutes per week (a product of PA intensity and duration). PA was classified as low (<600 MET minutes per week), moderate (600-3000 METs), or high (>3000 METs) [23]. Meeting current PA guidelines (PAG) was defined as a total PA of ≥600 METs [23].

At inclusion (M0), physical fitness was evaluated in both groups using validated physical fitness field tests from *Eurofit for Adults* [24] and the *Senior Fitness Tests* [25], with the 6-minute walk test to assess cardiorespiratory fitness (endurance), the arm curl test, the 30-second chair stand test for muscle strength, the lateral side–bending test for flexibility and patients' balance by the one-leg standing test, and the timed up and go test for balance. Quality of life was assessed using the Short Form Health Survey-12 (SF-12; version 2) [26]. The SF-12 assesses limitations in role functioning with 12 items. It consists of 2 subscales measuring physical health (physical component subscale [PCS]) and mental health (mental component subscale). The presence and severity of different impairments over the past 4 weeks are rated. Subscale scores can vary between 0 and 100, with higher scores indicating less impairment or greater health well-being.

From the end of the 3-week spa therapy, PA, body weight, waist circumference, and quality of life of the participants in both groups were assessed at month 2 (M2), month 4, month 6 (M6), month 8, month 10, and month 12 (M12) by interviewers (masked to the participant's randomization group) by phone. Data were collected every 2 months to avoid a loss to follow-up and to allow more precise measurement of change in outcome over time.

To limit missing data, participants were contacted 3 times for each follow-up phone interview. First, the participants were contacted by email to plan the phone interview; in case of no answer, an SMS text message was sent to his or her cell phone within 7 days, and after failing to respond within 3 days of the SMS text message, he or she was contacted directly by phone. The interviewer tried to contact nonresponders for 1 month after the theoretical follow-up date.

Outcomes

The primary outcome was meeting the current PAG at 12 months after the end of spa therapy, defined as reporting total PA ≥600 METs [23] measured by the IPAQ short form.

Secondary outcomes were meeting the current PAG at 6 months after the end of spa therapy; sedentary time, weight, waist circumference, PA, and quality of life at 6 and 12 months; and changes in these indicators evaluated every 2 months during the 12-month follow-up.

The use of the program was evaluated by the number of connections to the Thermactive website, number of PA sessions conducted (structured PA sessions+recorded PA sessions), and number of months for which use of the program was maintained.

Sample Size

With a risk of 0.05, a power (1-b) of 0.90, and assuming a detectable difference in patients meeting the PAG of 15% between the 2 groups [27], the sample size required was 462, with 231 participants in each study arm.

Statistics

Continuous variables were described as mean (SD or 95% CI) or median (IQR). The normal distribution of continuous variables was checked using the Shapiro-Wilk test. To compare between-group differences, a Student *t* test (2-tailed) was used for variables with normal distribution; otherwise, the Mann-Whitney test was used. Categorical variables are presented as frequencies and percentages and were compared between groups using a chi-square test. To test effectiveness, the data were analyzed using intention-to-treat principles [28]. As all randomized patients were included in the analyses and considering that assessment every 2 months should limit the loss to follow-up, attrition was not considered to increase the sample size [29]. To compare between-group differences from baseline for repeated outcomes, a constrained longitudinal data analysis (CLDA) was used. This mixed model is a constrained full-likelihood approach, whereby both the baseline and postbaseline values are modeled as dependent variables (the constrained model assumes that both the baseline and postbaseline measurements are jointly multivariate and normally distributed as the baseline value is treated as part of the response vector), and the true baseline values are constrained to be the same for the 2 treatment groups. Such methods based on maximum likelihood are consistent under the *missing at random* assumption. This model allows the inclusion of patients for whom either the baseline or postbaseline measurements are missing, thereby increasing efficiency [30]. Hence, this analysis provides an adjustment for the observed baseline difference in estimating the intervention effects. Time was treated as a categorical variable so that no restriction was imposed on the trajectory of the means over time. In addition to adjusting for baseline covariates, the analysis model was also adjusted for the intervention, time, sex, and interaction of time and intervention. Random effects at the patient and center levels were also included. The results are expressed as odds ratios (ORs) with 95% CI and *P* values for categorical variables and as differences in mean change from baseline to 1 year with 95% CI for continuous variables. All statistical tests were 2-sided, and $P<.05$ was considered statistically significant. Data were analyzed using Stata 12.

Safety

All serious adverse events (AEs) were recorded and notified to the French clinical trials pharmacovigilance system.

Ethics Approval

The trial, funded by Association Française pour la Recherche Thermale (grant number 2015-02), a nonprofit independent organization, was approved by the National Agency for the Safety of Medicine and Health Products and the regional ethics committee (Comité de Protection des Personnes Sud-Est N 6; registration number: CPP AU1196; registration number IDRCB:2015-A00855-44) and registered at ClinicalTrials.gov (NCT02694796) before enrollment of the participants began.

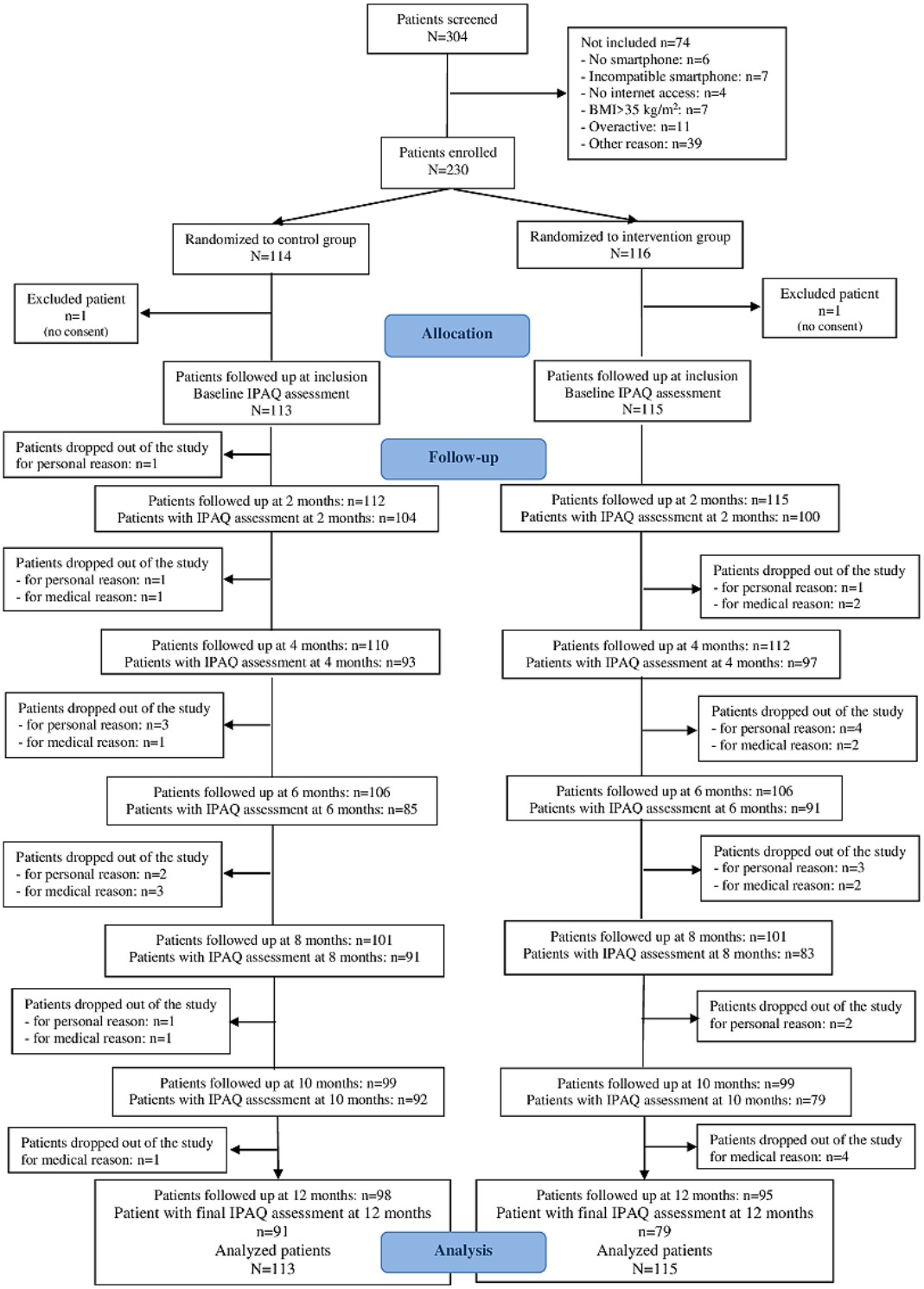
XSL•FO
RenderX

Results

Patients

Recruitment was conducted from September 2015 to December 2016. Of the 304 patients screened, 230 (75.6%) were enrolled and randomly assigned to either the control group (n=114, 49.6%) or intervention group (n=116, 50.4%; Figure 1). After randomization, 0.9% (2/230) of patients (1 in each group) withdrew their participation; thus, a total of 228 patients were included in the analyses. Patient characteristics are presented in Table 1. More than 3 participants out of 4 were women (176/228, 77.2%). The mean age of the sample was 62.4 (SD 6.7) years, and 53.9% (123/228) of the participants were retired. The 2 groups did not differ in any variable recorded at baseline.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flowchart. IPAQ: International Physical Activity Questionnaire.



XSL•FO
RenderX

Table 1. Baseline characteristics of the participants in the control and intervention group (N=228).

Characteristics	Control group (n=113)	Intervention group (n=115)	Total
Female, n (%)	86 (76.1)	90 (78.3)	176 (77.2)
Age (years), mean (SD)	62.3 (6.9)	62.6 (6.6)	62.5 (6.7)
Weight (kg), mean (SD)	76.3 (15.1)	77.0 (14.3)	76.7 (14.6)
BMI, mean (SD)	28.3 (4.4)	28.2 (4.0)	28.3 (4.2)
Waist circumference (cm), mean (SD)	96.7 (13.6)	95.0 (12.4)	95.8 (13.0)
Educational level, n (%)			
High school or less	62 (54.9)	55 (47.8)	117 (51.3)
Higher education	51 (45.1)	60 (52.2)	111 (48.7)
Occupation, n (%)			
Nonworking (retired, unemployed, housewife or househusband, disability, or long-term leave)	77 (68.1)	74 (64.3)	151 (66.2)
Manager (artisan, trader, senior executive, or intellectual profession)	18 (15.9)	16 (13.9)	34 (14.9)
Employee (intermediate occupation or worker)	17 (15)	24 (20.9)	41 (18)
Indication for spa treatment, n (%)			
Arthrosis	82 (72.6)	83 (72.2)	165 (72.4)
Cardiovascular diseases	14 (12.4)	11 (9.6)	25 (11)
Obesity	17 (15)	19 (16.5)	36 (15.8)
Diabetes (type 1 and type 2)	8 (7.1)	12 (10.4)	20 (8.8)
COPD[a]	2 (1.8)	1 (0.9)	3 (1.3)
Cancer	3 (2.7)	4 (3.5)	7 (3.1)
Other	26 (23)	22 (19.1)	48 (21.1)
Physical fitness, mean (SD)			
Resting heart rate (beats per minute)	70.1 (9.8)	71.1 (10.7)	70.6 (10.3)
6-minute walk test (minutes)	463 (97.6)	464.6 (94.6)	463.8 (95.9)
Arm curl test (number of flexions)	22 (6.9)	22.1 (7.2)	22.0 (7.0)
30-second chair stand test (number of up-and-down)	14.3 (4.1)	13.8 (4.4)	14.0 (4.2)
Lateral side–bending test (right side; cm)	15.9 (4.2)	15.2 (3.5)	15.6 (3.9)
Lateral side–bending test (left side; cm)	16 (4.3)	15.4 (3.7)	15.7 (4.0)
One-leg standing test (seconds)	6.2 (9.0)	6.0 (6.3)	6.1 (7.8)
Timed up and go test (seconds)	6.2 (1.5)	6.2 (1.7)	6.2 (1.6)
PA[b] (IPAQ[c]; MET[d] minutes per week), median (IQR)			
Continuous score for vigorous intensity	0 (0-960)	0 (0-480)	0 (0-960)
Continuous score for moderate intensity	120 (0-240)	240 (0-360)	130 (0-360)
Continuous score for walking	198 (66-396)	198 (66-346.5)	198 (66-396)
Continuous score for overall activity	396 (198-664)	419 (238-720)	396 (198-686)
Sedentary time (IPAQ; minutes), median (IQR)			
Time spent sitting on a week day	300 (240-420)	360 (270-480)	360 (240-480)
Time spent sitting on a weekend day	300 (240-360)	300 (240-360)	300 (240-360)
Time spent watching television on a week day	120 (120-180)	120 (120-180)	120 (120-180)
Time spent watching television on a weekend day	120 (120-240)	120 (120-180)	120 (120-180)
Time spent in front of computer or tablet on a week day	120 (60-180)	120 (60-240)	120 (60-210)
Time spent in front of computer or tablet on a weekend day	60 (30-120)	60 (45-150)	60 (30-120)

XSL•FO
RenderX

Characteristics	Control group (n=113)	Intervention group (n=115)	Total
Quality of life (SF-12[e]; 0-100), mean (SD)			
Physical health (PCS[f])	43.2 (8.5)	43.3 (8.5)	43.2 (8.5)
Mental health (MCS[g])	47.0 (9.5)	48.1 (8.9)	47.6 (9.2)

[a]COPD: chronic obstructive pulmonary disease.

[b]PA: physical activity.

[c]IPAQ: International Physical Activity Questionnaire.

[d]MET: metabolic equivalent of task.

[e]SF-12: Short Form Health Survey-12.

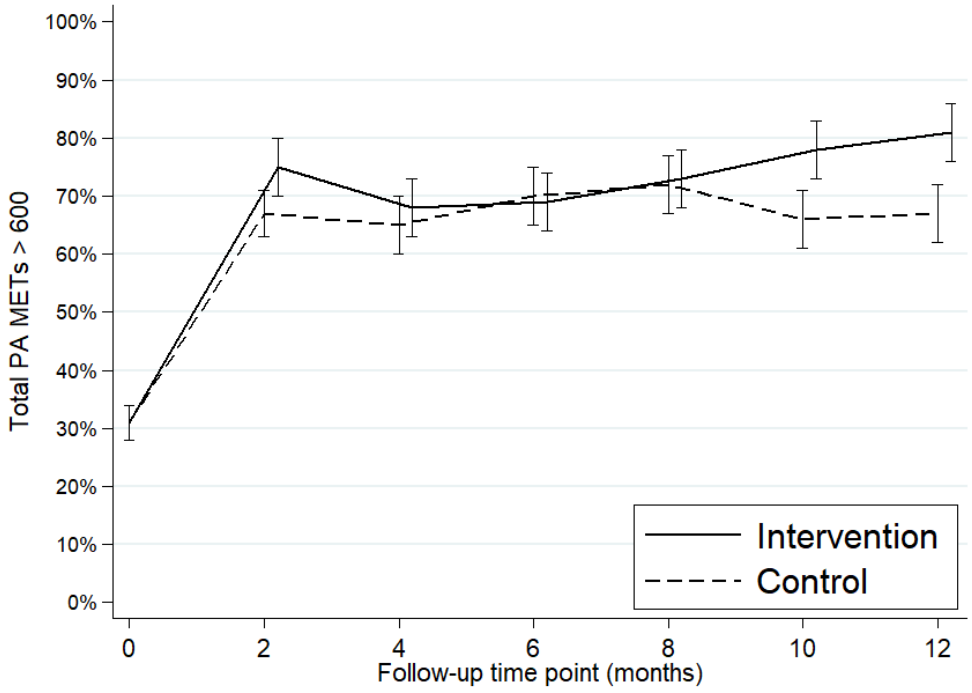
[f]PCS: physical component subscale.

[g]MCS: mental component subscale.

Primary Outcome

The change in the percentage achieving PAG marginal values according to CLDA modeling for each group is presented in Figure 2, and the statistical comparison between the groups for PAG achievement is shown in Table 2. The achievement of PAG significantly increased in both groups from M0 to M12 (Table 2), with the greatest increase occurring between M0 and M2 (Figure 2). At 12 months, the proportion of patients achieving PAG was significantly higher in the intervention group than in the control group (64/79, 81% vs 61/91, 67%, respectively; Figure 2; OR 2.34, 95% CI 1.02-5.38; *P*=.045; Table 2). The CLDA analysis also showed that significantly fewer women achieved PAG than men (*P*=.005; Table 2).

Figure 2. Change in the percentage of PA guidelines achievement (total PA MET≥600) marginal values according to constrained longitudinal data analysis model for each group over time. MET: metabolic equivalent of task; PA: physical activity.



XSL•FO
RenderX

Table 2. Constrained longitudinal data analysis model of the achievement of physical activity guidelines (total physical activity metabolic equivalents of task ≥600) over time.

Characteristics	Odds ratio (95% CI)	*P* value
Female	0.52 (0.33-0.83)	.005
Inclusion visit (month 0)	N/A[a]	N/A
2-month visit (month 2)	6.3 (3.55-11.17)	<.001
4-month visit (month 4)	5.49 (3.05-9.87)	<.001
6-month visit (month 6)	7.41 (3.97-13.85)	<.001
8-month visit (month 8)	8.37 (4.5-15.55)	<.001
10-month visit (month 10)	5.79 (3.2-10.48)	<.001
12-month visit (month 12)	6.29 (3.45-11.46)	<.001
Intervention group×month 2	1.58 (0.77-3.23)	.21
Intervention group×month 4	1.18 (0.58-2.41)	.65
Intervention group×month 6	0.95 (0.45-2.01)	.89
Intervention group×month 8	1.04 (0.47-2.26)	.93
Intervention group×month 10	2.12 (0.95-4.74)	.07
Intervention group×month 12	2.34 (1.02-5.38)	.045

[a]N/A: not applicable.

Secondary Outcomes

PA and Sedentary Times

At 6 months follow-up, the achievement of PAG did not differ between the intervention and control groups (63/91, 69.2% and 59/84, 70.2% of patients reached the PAG, respectively; Figure 2; OR 0.95, 95% CI 0.45-2.01; *P*=.89; Table 2). Regarding the PA level (Figure 3), the IPAQ score of total PA at M12 was significantly higher in the intervention group than in the control group (intervention group total PA 1618 METs, 95% CI 1491-1744 METs vs control group total PA 1275 METs, 95% CI 1140-1385 METs; *P*=.04), whereas no significant difference was observed at M6 (intervention group total PA 1427 METs, 95% CI 1303-1564 METs vs control group total PA 1274 METs, 95% CI 1146-1392 METs; *P*=.30).

There were no statistically significant differences between the 2 groups at M6 or M12 regarding the IPAQ scores for walking, moderate, and vigorous PA (Figure 3) or for sitting time or time spent in front of a screen (television or computer) during weekdays or weekends (Figure 4).

Nevertheless, the time spent in front of a screen (computer or television) decreased significantly over the follow-up in both the groups during both weekdays and weekends (Table 3).

XSL•FO
RenderX

Figure 3. International Physical Activity Questionnaire scores for total, moderate, intense, and walking physical activity margin values according to constrained longitudinal data analysis model for each group over time. MET: metabolic equivalent of task; PA: physical activity.

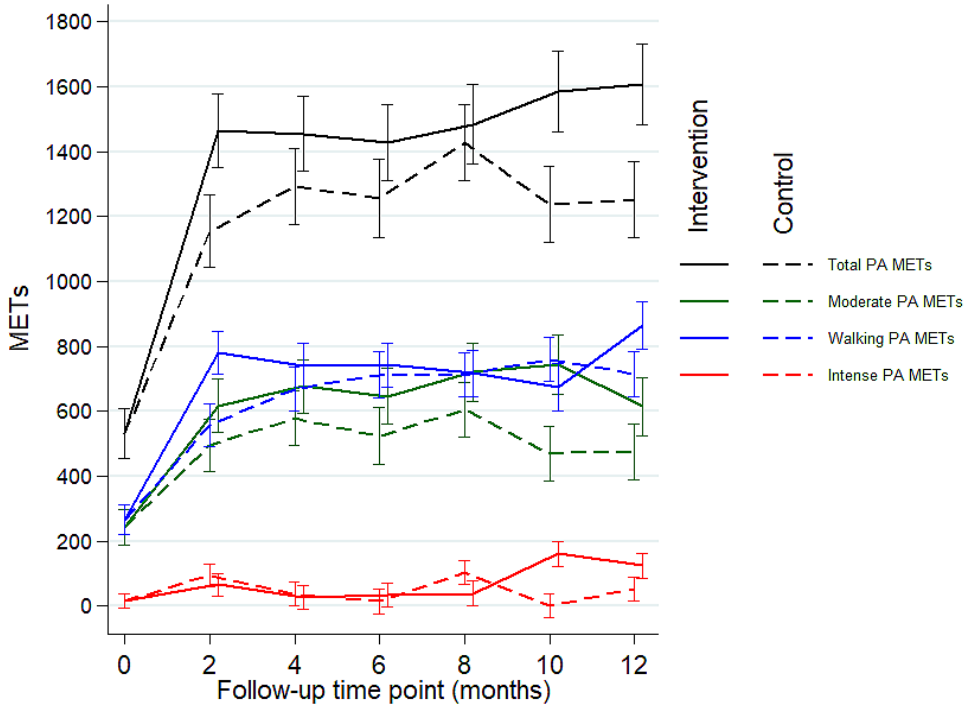
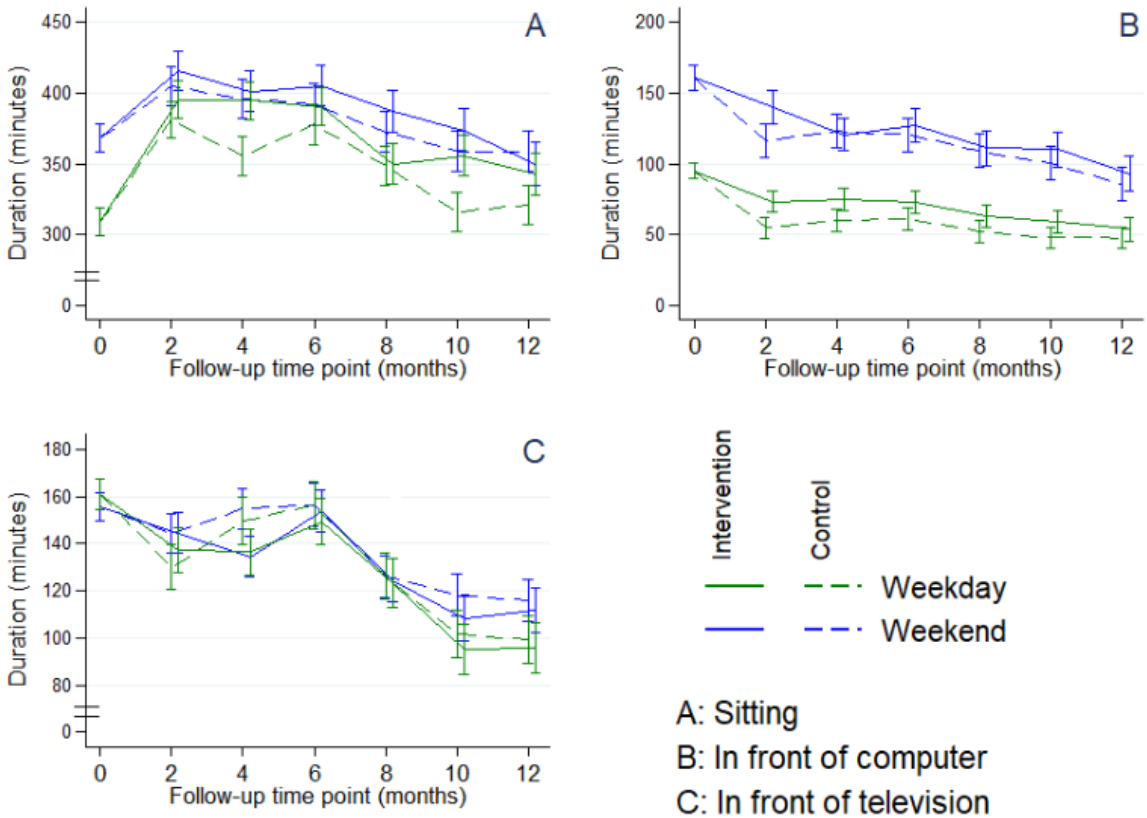


Figure 4. Sedentary times marginal values according to constrained longitudinal data analysis model for each group over time.



XSL•FO
RenderX

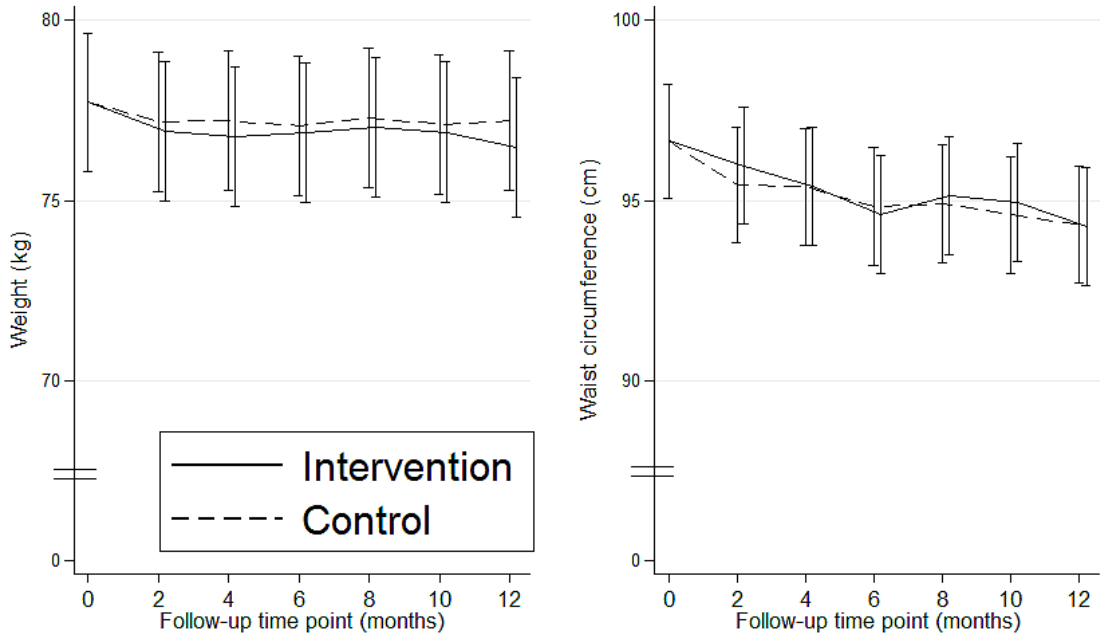
Table 3. Change in time spent in front of screens (computer or television) for control and intervention groups pooled.

Duration of time	Change between month 12 and month 0, mean (SE; 95% CI)	*P* value
Time spent in front of a computer during the week (minutes)	−75.2 (10.2; −95.3 to −55.2)	<.001
Time spent in front of a computer during the weekend (minutes)	−47.2 (7.8; −62.6 to −31.8)	<.001
Time spent in front of a television during the week (minutes)	−39.7 (9.3; −58.0 to −21.4)	<.001
Time spent in front of a television during the weekend (minutes)	−61.9 (10.8; −83.0 to −40.7)	<.001

Body Weight and Waist Circumference

There was no statistically significant difference between the 2 groups for body weight or waist circumference at M6 and M12 (Figure 5). However, the mean waist circumference for the 2 groups had significantly decreased at 6 months by 1.9 cm (95% CI −3.0 to −0.8 cm; *P*=.001) and at 12 months by 2.4 cm (95% CI −3.5 to −1.3 cm; *P*<.001).

Figure 5. Weight and waist circumference marginal values according to constrained longitudinal data analysis model for each group over time.

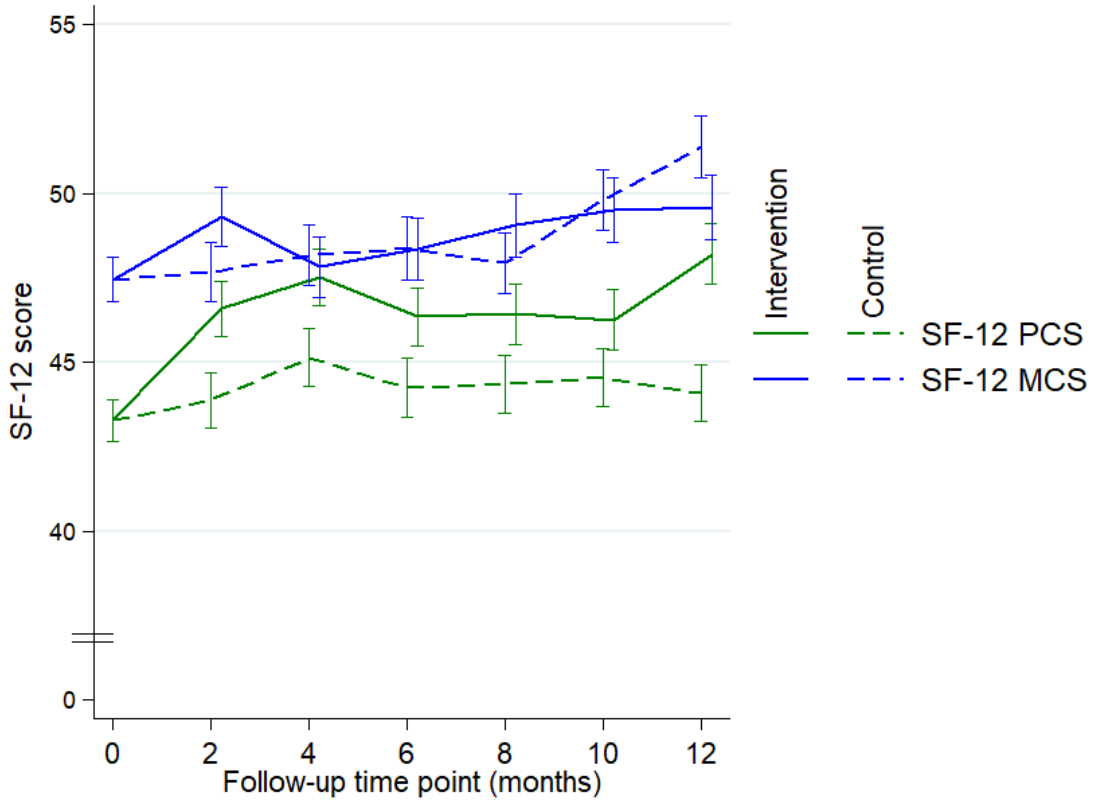


Quality of Life

The quality of life assessment showed that the PCS score was significantly higher at M12 in the intervention group than in the control group (Figure 6; mean difference at M12 4.1, 95% CI 1.9-6.3; *P*<.001). At M6, the PCS score tended to be higher in the intervention group than in the control group (PCS score 2.1, 95% CI 0.0-4.3; *P*=.06). There were no statistically significant differences between the 2 groups in the mental component subscale score at M6 or M12 (Figure 6).

XSL•FO
RenderX

Figure 6. SF-12 scores (PCS and MCS) marginal values to constrained longitudinal data analysis model for each group over time. MCS: mental component subscale; PCS: physical component subscale; SF-12: Short Form Health Survey-12.



Use of the Program

Monitoring of the program use results is presented in Table 4. The patients used the program for an average of 7.1 (SD 4.5) months. Approximately 20.4% (23/113) dropped out of the program before 2 months of use; however, 39.8% (45/113) of the participants used the program for ≥10 months (Table 4). Among the participants, 62.8% (71/113) had at least one structured PA session.

XSL•FO
RenderX

Table 4. Use of the program (N=115).

Characteristics	Intervention group
Logging into the program	
Patients who logged into the program at least once, n (%)	113 (98.3)
Total number of log-ins into the program, N	16,325
Number of log-ins by patients, mean (SD)	143.2 (179.4)
Number of log-ins by patients, median (IQR)	76 (24.3-208.8)
Duration of program use	
Duration of use (months), mean (SD)	7.1 (4.5)
Patients who used the program for <2 months, n (%)	23 (20.4)
Patients who used the program for 2 to 4 months, n (%)	14 (12.4)
Patients who used the program between 4 and 6 months, n (%)	13 (11.5)
Patients who used the program between 6 and 8 months, n (%)	8 (7.1)
Patients who used the program between 8 and 10 months, n (%)	10 (8.8)
Patients who used the program for >10 months, n (%)	45 (39.8)
Total PA[a] sessions (recorded+structured)	
Patients who conducted at least one PA session, n (%)	81 (71.7)
Total number of PA sessions conducted, N	2588
Number of PA sessions conducted, median (IQR)	16 (3-47)
Structured PA sessions	
Patients who conducted at least one structured PA session, n (%)	71 (62.8)
Total number of structured PA sessions conducted, N	1836
Number of structured PA sessions conducted, median (IQR)	8 (2-34)
Patients who conducted <1 structured PA session by month of use, n (%)	25 (35.2)
Patients who conducted 1 to 4 structured PA sessions by month of use, n (%)	26 (36.6)
Patients who conducted 4 to 8 structured PA sessions by month of use, n (%)	16 (22.5)
Patients who conducted >8 structured PA sessions by month of use, n (%)	4 (5.6)

[a]PA: physical activity.

Safety

AEs recorded during the study are presented in Table 5. None of the severe AEs were attributed to the intervention. One patient reported an aggravation of lymphedema in the left arm because of wearing a wrist pedometer. This adverse effect was resolved by physiotherapy.

Table 5. Adverse events recorded during the follow-up (N=228).

Adverse events	Control group (n=113), n (%)	Intervention group (n=115), n (%)
Adverse events	102 (49.6)	70 (60.9)
Severe adverse events	13 (11.5)	11 (9.6)
Increased arthrosis	2 (1.8)	0 (0)
Hospitalizations or care for a disorder unrelated to the spa indication	11 (9.7)	11 (9.6)

Discussion

Principal Findings and Comparison With Prior Work

This RCT aimed to assess the effectiveness of an intervention, including an initial face-to-face coaching and a web- and mobile-based PA program, to meet PAG among patients attending a 3-week spa therapy treatment. The results showed that significantly more participants met the PAG at the 12-month follow-up in the intervention group than in the controls; however, no difference was observed between the 2 groups for reaching PAG at 6 months. The intervention significantly improved the physical component of the quality of life at 12 months. Sedentary times and waist circumference were

XSL•FO
RenderX

significantly reduced in both groups at 6 and 12 months of follow-up without significant differences between the groups.

The level of PA increased in both groups but was significantly higher at 12 months in the intervention group. The increase in PA in the control group might be explained by the usual advice on PA and lifestyle changes provided during the 3-week spa therapy by health care professionals. Indeed, a number of studies have shown that the context and environment of spa treatments represent an opportunity to educate patients on their chronic diseases and initiate behavioral changes [16-20], such as PA.

Our analyses showed that the effect of usual advice on PA in the control group was the highest during the first 2 months after the spa therapy; subsequently, this tended to stabilize and finally slightly decreased after 8 months. Although the PA in the intervention group followed the same dynamic for the first 8 months, it increased after 8 months and became significantly higher at 12 months.

The maintenance of the level of PA to reach the PAG at 12 months in the intervention group could be explained by the web- and mobile-based PA program. This result is in line with the results observed in other RCTs aimed at improving PA among older adults using web-based PA interventions [31,32]. A systematic review and meta-analysis evaluated the effects of eHealth interventions on promoting PA in older adults [12]. The results of this meta-analysis showed that the effects of the eHealth intervention (vs controls) on PA time measured by questionnaires and objective wearable devices on energy expenditure and step counts were all significant with minimal heterogeneity.

Our findings also highlight that the intervention significantly improved the physical component of quality of life at 12 months, which is consistent with the increase in physical abilities because of the improvement in PA level. Limited studies have reported on the effect of web- or mobile-based PA interventions on quality of life among older adults. A randomized control trial that included 235 participants indicated that after 3 months, an internet-based intervention aimed at increasing PA significantly improved the quality of life of inactive older adults [33]. Another study conducted by Irvine et al [34] also showed a significant improvement in the SF-12 PCS among sedentary older adults aged >55 years who engaged in a web-based PA program.

Our results indicate that waist circumference was significantly reduced in both groups at 6 and 12 months of follow-up without a significant difference between the groups.

A meta-analysis [35], including 31 RCTs, emphasized that internet-based interventions showed a significant reduction in waist circumference (mean change −2.99 cm, 95% CI −3.68 to −2.30 cm; I^2=93.3%) compared with minimal interventions such as information-only groups. Our findings indicate a similar mean change in waist circumference in the 2 groups (−2.4 cm; 95% CI −3.5 to −1.3 cm). Therefore, this reduction did not seem to be explained by the intervention. The inclusion in a research study and the focus on their medical conditions should motivate them to adopt better health behaviors. The time spent sitting was higher at M2, month 4, and M6 in both groups than that at baseline. This could be because of fatigue related to the increase in PA [36], which induced compensatory time spent being sedentary, probably at the expense of light PA (unassessed by the IPAQ questionnaire, but which can represent most PA in older adults). This hypothesis should be confirmed in future studies.

Our results also indicate that men were more likely to successfully reach the PAG than women. The present findings are consistent with those of previous studies. Blanchard et al [37] evaluated PA levels in patients with heart disease over 12 months (with or without cardiac rehabilitation) and showed a more pronounced decline in PA over time in women than in men. Jenkins and Gortner [38] specifically examined gender disparity in PA in people living with heart disease who did not receive cardiac rehabilitation. The results showed that men walked significantly more than women at 1, 2, 6, and 12 months after hospitalization. However, analyzing the determinants of parameters that establish which factors predict which participants are successful in reaching PAG was not a part of our research question. Such a determinant analysis will be performed in forthcoming studies and will address different research questions with the ultimate aim of better targeting different populations.

Limitations and Strengths

The effect of the intervention on maintaining long-term PA and reaching PAG needs to be viewed cautiously as, despite an extension of the enrollment period, the a priori sample size was not met. Two main reasons explain the difficulties in including participants in the trial. First, it appeared that many patients with a web connection and smartphone were already meeting the PAG. Second, we encountered difficulties in the recruitment of qualified PA instructors who played an essential role in the prescreening of participants and face-to-face coaching of the intervention group.

Another limitation of our trial is the self-reported assessment measures, making them potentially subject to social desirability bias [39]. Furthermore, the Hawthorne effect [40] (referring to a tendency in some individuals to alter their behavior in response to their awareness of being observed) along with contamination bias could also affect the magnitude of the differences observed in the results. However, the contamination bias cannot call into question our main result as it reduced the size of the difference between the 2 groups. Therefore, we can hypothesize that without contamination bias, the difference between the 2 groups would have been greater.

The Hawthorne effect and the repeated assessment of outcomes every 2 months could motivate participants to become more active, leading them to overestimate the report of PA and consequently bias our findings. Although this bias could have occurred in both the control and intervention groups and, therefore, would not bias the comparison between the 2 groups, the proportion of participants achieving PAG might be overestimated. Moreover, we cannot exclude that participants in the intervention group may be influenced by the expectation that they will perform better as they received the promising PA program, especially at the end of the program, resulting in an overestimation of their PA level.

XSL•FO
RenderX

A greater number of patients was assessed at M12 in the control group (91/113, 80.5%) than in the intervention group (79/115, 68.7%). One of the reasons for this higher compliance of the control group may be the promise to have free access to the program at the end of the follow-up.

The use of the program can be considered satisfactory as patients used the program for an average of 7.1 (SD 4.5) months; 78.3% (90/115) of the patients used the program for at least 2 months and 39.1% (45/115) for at least 10 months. Approximately 61.7% (71/115) of patients reported engaging in structured PA sessions (median 8 sessions), emphasizing the clear interest of participants in the value of the program, as well as its acceptability and usability. Indeed, the attrition rate for web and smartphone interventions in PA is often quite high [41] (ranging from 30% [42] to 80% [43]), and declining rates of engagement over time are often reported by researcher-led web-based health interventions.

In a secondary analysis of a randomized trial [42], attrition at 3 months of a 100-day PA intervention delivered via an app ranged from 32% to 39%. Another RCT found that 80% of participants ceased using a web-based PA intervention by week 80 (20 months), and the attrition rate was approximately 70% to 75% at 12 months [43].

The percentage of patients who stopped using the web application and mobile app before 4 months was 32.8% (37/113) in this study.

Thus, the attrition rate observed in our study was consistent with that reported in the literature.

A recent study [44] examined the effect of individualized follow-up with an app for 1 year on peak oxygen uptake in patients undergoing cardiac rehabilitation. The results of this study showed high levels of use of the app in the intervention group: 84% (46/55) of the patients used it to set and achieve personal goals and tasks. The intervention group improved in the peak oxygen uptake to a larger extent than the control group (without the app). Adherence to app use was more than twice the web and app adherence estimated in this study (45/113, 39.8%). This could be mainly explained by the fact that in the study of Lunde et al [44], monitoring and feedback were provided by a real person to the patients, whereas in our study, the PA program was fully automated. The authors explained that the high level of individualization (having a real person behind the app, as well as quite simple technology) may have been crucial to maintaining adherence to app use.

Therefore, adherence in the long term (>10 months) to the web- and mobile-based PA program studied here would be enhanced by introducing engagement with a real PA instructor in the follow-up of the patients.

In our analyses (not shown in the manuscript), we compared the *respondents* and those with *missing* data at 12 months by baseline characteristics.

Those with *missing* data differed from the *respondents* by the baseline declaration of *high PA* and *sitting time during the weekend*. The proportion of those with missing data who declared practicing high PA at baseline was higher than the proportion of the *respondents* (4/58, 6.9% vs 3/170, 1.8%, respectively; *P*=.05). For sitting time, those with *missing* data declared, at baseline, to spend less time sitting during the weekend than the *respondents* (280 vs 320 minutes, respectively; *P*=.046). Nevertheless, those with *missing* data were more frequent in the intervention group than in the control group (36/115, 31.3% vs 22/113, 19.5%, respectively). Therefore, if we hypothesized that those with *missing* data were more active than the *respondents*, the level of PA of the intervention group would have been higher if we had been able to collect data from those with *missing* data.

Finally, in the present analyses, we did not investigate the determinants of which participants were adherent to the program. Such analyses, along with the presentation of the results on the step counts, will be the topic of ongoing analyses.

This clinical trial provided results on the PA of participants attending spa treatment. The generalizability of the results to the general population of older adults with NCDs without spa treatment or rehabilitation programs might be limited. Thus, attending a spa treatment or rehabilitation program is proof of interest in one's health.

To our knowledge, this study is the first to combine education during a spa treatment and the use of a web- and mobile-based PA program over a 12-month follow-up. The 3-week stay at a spa resort favors the building of strong relations and exchanges with health care professionals and other patients and has an educational dimension [16,20]. This could help explain why various studies have shown that coaching and information on PA administered during spa therapy produces a lasting benefit on PA [17-20] in the intervention groups and also produces an improvement in the controls [20]. Thus, these findings could partly explain why no large differences in PA were observed among patients receiving information in different forms. Moreover, the periodic follow-up by the interviewers in the 2 groups could also be a potential reason for the increasing motivation to practice PA.

Conclusions

The limitations, especially the impossibility of reaching the required sample size, indicate that it is necessary to interpret the results with caution. Nonetheless, this study demonstrates the potential of a web- and mobile-based PA program associated with an initial face-to-face coaching during a spa treatment to maintain PA in older adults over a 12-month period to achieve PAG and improve quality of life. A spa treatment appears to offer the ideal time and setting to implement education in PA and initiate patients to the use of web- and mobile-based PA programs.

Increasing PA and reducing the excessive sedentariness of inactive patients reduce the risk of NCD aggravation and pain in some nonmalignant chronic conditions, favoring a lasting improvement in personal physical capacity and quality of life.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank the physicians in the spa resorts who enrolled patients in the study: Dr Noël Aprile, Dr Jean-Pierre Armand, Dr Anne-Marie Baque-Gensac, Dr Marie-André Bernard, Dr Georges Breteau, Dr Gisèle Briancon-Bobillion, Dr Marc Chedal, Dr Pierre Coutant, Dr Jean-Claude Desprez, Dr Agnes Dolci, Dr James El Farkh, Dr Valérie Escudero, Dr Isabelle Fay, Dr Romain Forestier, Dr Martine Fournier-Treme, Dr Alain Françon, Dr Bruno Figgeri, Dr Didier Gleizes, Dr Jacqueline Guerin, Dr Philippe Guerin, Dr Riad Hamoui, Dr Catherine Hernandez, Dr Denis Hours, Dr Eric Jouret, Dr Marie-Reine Mahieu, Dr Claudine Marcerou, Dr Chantal Melis, Dr Martine Merlen, Dr Christian Perrard, Dr Philippe Rougerie, Dr Pascale Segonne, Dr Nicolas Thiebaud, Dr Luis Veilla, and Dr Mercedes Vico. The authors would also like to thank Dr Agnès Dechartres for preparing the statistical analysis plan and Professor Emma Boyland, University of Liverpool for proofreading and revising the manuscript.

Authors' Contributions

FF designed the study, supervised the trial progress and implementation, wrote the manuscript (*Introduction*, *Methods*, *Results and Discussion* sections), and revised the manuscript. LP participated in the management of logistics and monitoring of data. SP conceived and designed the digital program and implemented the web platform. AM implemented the statistical analysis and revised the manuscript. SR revised the manuscript. CFR revised the manuscript. MD designed the study, supervised the progression of the trial, and revised the manuscript.

Conflicts of Interest

FF was an employee of Biomouv SAS Inc and SP was the chief executive officer of Biomouv SAS Inc, who provided the web and smartphone-based physical activity program. CFR was the president of the Association Française pour la Recherche Thermale Scientific Committee.

Multimedia Appendix 1
Use guide of the website.
[PPT File (Microsoft PowerPoint Presentation), 3915 KB - jmir_v24i6e29640_app1.ppt]

Multimedia Appendix 2
Use guide of the mobile app.
[PPT File (Microsoft PowerPoint Presentation), 2890 KB - jmir_v24i6e29640_app2.ppt]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1219 KB - jmir_v24i6e29640_app3.pdf]

References

1. Global status report on noncommunicable diseases 2014. World Health Organization. 2014. URL: https://apps.who.int/iris/handle/10665/148114 [accessed 2021-10-05]
2. Thivel D, Tremblay A, Genin PM, Panahi S, Rivière D, Duclos M. Physical activity, inactivity, and sedentary behaviors: definitions and implications in occupational health. Front Public Health 2018 Oct 5;6:288 [FREE Full text] [doi: 10.3389/fpubh.2018.00288] [Medline: 30345266]
3. GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1736-1788 [FREE Full text] [doi: 10.1016/S0140-6736(18)32203-7] [Medline: 30496103]
4. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. Lancet Glob Health 2018 Oct;6(10):e1077-e1086 [FREE Full text] [doi: 10.1016/S2214-109X(18)30357-7] [Medline: 30193830]
5. Étude de santé sur l'environnement, la biosurveillance, l'activité physique et la nutrition (Esteban), 2014-2016. Volet Nutrition. Chapitre Activité physique et sédentarité. 2e édition. Santé publique France. 2020 Feb 13. URL: https://www.santepubliquefrance.fr/import/etude-de-sante-sur-l-environnement-la-biosurveillance-l-activite-physique-et-la-nutrition-esteban-2014-2016.-volet-nutrition.-chapitre-activit [accessed 2021-10-05]
6. Leavitt MO. 2008 Physical Activity Guidelines for Americans. U.S. Department of Health and Human Services. 2008. URL: https://health.gov/sites/default/files/2019-09/paguide.pdf [accessed 2021-10-05]
7. Macera CA, Cavanaugh A, Bellettiere J. State of the art review: physical activity and older adults. Am J Lifestyle Med 2016 Jun 23;11(1):42-57 [FREE Full text] [doi: 10.1177/1559827615571897] [Medline: 30202313]
8. Dorn J, Naughton J, Imamura D, Trevisan M. Correlates of compliance in a randomized exercise trial in myocardial infarction patients. Med Sci Sports Exerc 2001 Jul;33(7):1081-1089. [doi: 10.1097/00005768-200107000-00003] [Medline: 11445753]

XSL•FO
RenderX

9. Gal R, May AM, van Overmeeren EJ, Simons M, Monninkhof EM. The effect of physical activity interventions comprising wearables and smartphone applications on physical activity: a systematic review and meta-analysis. Sports Med Open 2018 Sep 03;4(1):42 [FREE Full text] [doi: 10.1186/s40798-018-0157-9] [Medline: 30178072]
10. Jahangiry L, Farhangi MA, Shab-Bidar S, Rezaei F, Pashaei T. Web-based physical activity interventions: a systematic review and meta-analysis of randomized controlled trials. Public Health 2017 Nov;152:36-46. [doi: 10.1016/j.puhe.2017.06.005] [Medline: 28734170]
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. Kwan RY, Salihu D, Lee PH, Tse M, Cheung DS, Roopsawang I, et al. The effect of e-health interventions promoting physical activity in older people: a systematic review and meta-analysis. Eur Rev Aging Phys Act 2020 Apr 21;17:7 [FREE Full text] [doi: 10.1186/s11556-020-00239-5] [Medline: 32336996]
13. Boekhout JM, Berendsen BA, Peels DA, Bolman CA, Lechner L. Evaluation of a computer-tailored healthy ageing intervention to promote physical activity among single older adults with a chronic disease. Int J Environ Res Public Health 2018 Feb 15;15(2):346 [FREE Full text] [doi: 10.3390/ijerph15020346] [Medline: 29462862]
14. Lunde P, Nilsson BB, Bergland A, Kværner KJ, Bye A. The effectiveness of smartphone apps for lifestyle improvement in noncommunicable diseases: systematic review and meta-analyses. J Med Internet Res 2018 May 04;20(5):e162 [FREE Full text] [doi: 10.2196/jmir.9751] [Medline: 29728346]
15. 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]
16. Gremeaux V, Benaïm C, Poiraudeau S, Hérisson C, Dupeyron A, Coudeyre E. Evaluation of the benefits of low back pain patients' education workshops during spa therapy. Joint Bone Spine 2013 Jan;80(1):82-87. [doi: 10.1016/j.jbspin.2011.12.014] [Medline: 22342470]
17. Kwiatkowski F, Mouret-Reynier MA, Duclos M, Leger-Enreille A, Bridon F, Hahn T, et al. Long term improved quality of life by a 2-week group physical and educational intervention shortly after breast cancer chemotherapy completion. Results of the 'Programme of Accompanying women after breast Cancer treatment completion in Thermal resorts' (PACThe) randomised clinical trial of 251 patients. Eur J Cancer 2013 May;49(7):1530-1538. [doi: 10.1016/j.ejca.2012.12.021] [Medline: 23352440]
18. Gin H, Demeaux JL, Grelaud A, Grolleau A, Droz-Perroteau C, Robinson P, et al. Observation of the long-term effects of lifestyle intervention during balneotherapy in metabolic syndrome. Therapie 2013;68(3):163-167. [doi: 10.2515/therapie/2013025] [Medline: 23886461]
19. Maitre J, Guinhouya B, Darrieutort N, Paillard T. Physical education in a thermal spa resort to maintain an active lifestyle at home: a one-year self-controlled follow-up pilot study. Evid Based Complement Alternat Med 2017;2017:1058419 [FREE Full text] [doi: 10.1155/2017/1058419] [Medline: 28539960]
20. Schnebelen-Berthier C, Negro N, Jaruga A, Roques CF, Lecerf JM. Long term effect of spa therapy combined with patient education program on subjects with overweight and obesity - a controlled study. Obes Res Clin Pract 2019;13(5):492-498. [doi: 10.1016/j.orcp.2019.06.005] [Medline: 31383564]
21. Jones CJ, Rose DJ. International guidelines for training physical activity instructors for older adults. J Aging Phys Act 2004 Jan;12(1):1-2. [doi: 10.1123/japa.12.1.1] [Medline: 15211016]
22. Borg G. Psychophysical scaling with applications in physical work and the perception of exertion. Scand J Work Environ Health 1990;16 Suppl 1:55-58 [FREE Full text] [doi: 10.5271/sjweh.1815] [Medline: 2345867]
23. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
24. Ojad P, Tuxworth B. Eurofit for Adults: Assessment of Health-related Fitness. New York, NY, USA: Council of Europe, Committee for the Development of Sport; 1995.
25. Fournier J, Vuillemin A, Le Cren F. Mesure de la condition physique chez les personnes âgées. Évaluation de la condition physique des seniors : adaptation française de la batterie américaine « Senior Fitness Test ». Sci Sports 2012 Sep;27(4):254-259. [doi: 10.1016/j.scispo.2012.07.005]
26. Ware JE, Kosinski M, Keller SD. SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales. Boston, MA, USA: The Health Institute, New England Medical Center; 1995.
27. Kolt GS, Schofield GM, Kerse N, Garrett N, Ashton T, Patel A. Healthy Steps trial: pedometer-based advice and physical activity for low-active older adults. Ann Fam Med 2012;10(3):206-212 [FREE Full text] [doi: 10.1370/afm.1345] [Medline: 22585884]
28. Gupta SK. Intention-to-treat concept: a review. Perspect Clin Res 2011 Jul;2(3):109-112 [FREE Full text] [doi: 10.4103/2229-3485.83221] [Medline: 21897887]
29. Wittes J. Sample size calculations for randomized controlled trials. Epidemiol Rev 2002;24(1):39-53. [doi: 10.1093/epirev/24.1.39] [Medline: 12119854]

XSL•FO
RenderX

30. Lu K. On efficiency of constrained longitudinal data analysis versus longitudinal analysis of covariance. Biometrics 2010 Sep;66(3):891-896. [doi: 10.1111/j.1541-0420.2009.01332.x] [Medline: 19764951]
31. Alley SJ, Kolt GS, Duncan MJ, Caperchione CM, Savage TN, Maeder AJ, et al. The effectiveness of a Web 2.0 physical activity intervention in older adults - a randomised controlled trial. Int J Behav Nutr Phys Act 2018 Jan 12;15(1):4 [FREE Full text] [doi: 10.1186/s12966-017-0641-5] [Medline: 29329587]
32. 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]
33. 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]
34. Irvine AB, Gelatt VA, Seeley JR, Macfarlane P, Gau JM. Web-based intervention to promote physical activity by sedentary older adults: randomized controlled trial. J Med Internet Res 2013 Feb 05;15(2):e19 [FREE Full text] [doi: 10.2196/jmir.2158] [Medline: 23470322]
35. Seo DC, Niu J. Evaluation of Internet-based interventions on waist circumference reduction: a meta-analysis. J Med Internet Res 2015 Jul 21;17(7):e181 [FREE Full text] [doi: 10.2196/jmir.3921] [Medline: 26199208]
36. Melanson EL, Keadle SK, Donnelly JE, Braun B, King NA. Resistance to exercise-induced weight loss: compensatory behavioral adaptations. Med Sci Sports Exerc 2013 Aug;45(8):1600-1609 [FREE Full text] [doi: 10.1249/MSS.0b013e31828ba942] [Medline: 23470300]
37. Blanchard CM, Reid RD, Morrin LI, Beaton LJ, Pipe A, Courneya KS, et al. Barrier self-efficacy and physical activity over a 12-month period in men and women who do and do not attend cardiac rehabilitation. Rehabil Psychol 2007;52(1):65-73 [FREE Full text] [doi: 10.1037/0090-5550.52.1.65]
38. Jenkins LS, Gortner SR. Correlates of self-efficacy expectation and prediction of walking behavior in cardiac surgery elders. Ann Behav Med 1998;20(2):99-103. [doi: 10.1007/BF02884455] [Medline: 9989315]
39. Fisher RJ. Social desirability bias and the validity of indirect questioning. J Consum Res 1993 Sep;20(2):303-315. [doi: 10.1086/209351]
40. Parsons HM. What happened at Hawthorne?: new evidence suggests the Hawthorne effect resulted from operant reinforcement contingencies. Science 1974 Mar 08;183(4128):922-932. [doi: 10.1126/science.183.4128.922] [Medline: 17756742]
41. 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]
42. 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]
43. Kolt GS, Rosenkranz RR, Vandelanotte C, Caperchione CM, Maeder AJ, Tague R, et al. Using Web 2.0 applications to promote health-related physical activity: findings from the WALK 2.0 randomised controlled trial. Br J Sports Med 2017 Oct;51(19):1433-1440 [FREE Full text] [doi: 10.1136/bjsports-2016-096890] [Medline: 28049624]
44. Lunde P, Bye A, Bergland A, Grimsmo J, Jarstad E, Nilsson BB. Long-term follow-up with a smartphone application improves exercise capacity post cardiac rehabilitation: a randomized controlled trial. Eur J Prev Cardiol 2020 Nov;27(16):1782-1792 [FREE Full text] [doi: 10.1177/2047487320905717] [Medline: 32106713]

Abbreviations

AE: adverse event
CLDA: constrained longitudinal data analysis
IPAQ: International Physical Activity Questionnaire
M0: month 0
M2: month 2
M6: month 6
M12: month 12
MET: metabolic equivalent of task
NCD: noncommunicable disease
OR: odds ratio
PA: physical activity
PAG: physical activity guidelines
PCS: physical component subscale
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture
SF-12: Short Form Health Survey-12

XSL•FO
RenderX

Edited by A Mavragani; submitted 17.04.21; peer-reviewed by R Gal, P Lunde; comments to author 12.05.21; revised version received 06.10.21; accepted 19.04.22; published 16.06.22.

Please cite as:
Fillol F, Paris L, Pascal S, Mulliez A, Roques CF, Rousset S, Duclos M
Possible Impact of a 12-Month Web- and Smartphone-Based Program to Improve Long-term Physical Activity in Patients Attending Spa Therapy: Randomized Controlled Trial
J Med Internet Res 2022;24(6):e29640
URL: https://www.jmir.org/2022/6/e29640
doi:10.2196/29640
PMID:35708743

©Florie Fillol, Ludivine Paris, Sébastien Pascal, Aurélien Mulliez, Christian-François Roques, Sylvie Rousset, Martine Duclos. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 16.06.2022. 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

To Screen or Not to Screen? At Which BMI Cut Point? Comment on “Obesity and BMI Cut Points for Associated Comorbidities: Electronic Health Record Study”

Chrissa Sioka[1], MD

Department of Nuclear Medicine, University Hospital of Ioannina, Ioannina, Greece

Corresponding Author:
Chrissa Sioka, MD
Department of Nuclear Medicine
University Hospital of Ioannina
1 Stavrou Niarhos Street
Ioannina, 45500
Greece
Phone: 30 2651099375
Email: csioka@yahoo.com

Related Articles:

Comment on: http://www.jmir.org/2021/8/e24017/

Comment in: http://www.jmir.org/2022/6/e39717/

(J Med Internet Res 2022;24(6):e37267) doi:10.2196/37267

KEYWORDS

obesity; body mass index; BMI; risk factors; screening; health services; chronic disease; heart disease; myocardial perfusion imaging; anxiety; depression

In their article, Liu et al [1] investigated if there are BMI cut points for various obesity-associated comorbidities. For this purpose, they evaluated 243,332 patients aged 18 to 75 years, documented in the electronic health record, who had at least 3 in-person clinical visits over a 2-year period. The authors reported a significant correlation and calculated cutoff points of BMI with 6 comorbidities, including coronary artery disease, hypertension, hyperlipidemia, obstructive sleep apnea, osteoarthritis, and type 2 diabetes mellitus [1]. Interestingly, no association was found with anxiety and depression.

In our recent study [2], although of a different concept, we prospectively evaluated a cohort of 80 patients who were subjected to myocardial perfusion imaging for myocardial ischemia evaluation. All patients in the study were additionally evaluated for the presence of anxiety and depression. Furthermore, cardiological risk factors, including obesity, were additionally assessed. Like the study by Liu et al [1], we found a positive association between obesity and myocardial ischemia. However, we also found a correlation between obesity and depression/anxiety and myocardial ischemia [2]. Other studies have also reported obesity and depression/anxiety as independent risk factors for acute coronary syndrome in young women [3] and that female patients had more central obesity and greater anxiety than male patients with coronary artery disease [4].

In any event, I agree with the conclusions of Liu et al [1] that additional studies may be needed to establish which comorbidities need to be screened in patients with overweight for appropriate management.

Conflicts of Interest

None declared.

References

1. Liu N, Birstler J, Venkatesh M, Hanrahan L, Chen G, Funk L. Obesity and BMI cut points for associated comorbidities: Electronic health record study. J Med Internet Res 2021 Aug 09;23(8):e24017 [FREE Full text] [doi: 10.2196/24017] [Medline: 34383661]

XSL•FO
RenderX

2. Fotopoulos A, Petrikis P, Iakovou I, Papadopoulos A, Sakelariou K, Gkika E, et al. The impact of depression and anxiety in prognosis of patients undergoing myocardial perfusion imaging with 99mTc tetrofosmin SPECT for evaluation of possible myocardial ischemia. Nucl Med Rev Cent East Eur 2020 Jul;23(2):58-62 [FREE Full text] [doi: 10.5603/NMR.a2020.0014] [Medline: 33007091]
3. Liu R, Xu F, Zhou Y, Liu T. The characteristics of risk factors in Chinese young women with acute coronary syndrome. BMC Cardiovasc Disord 2020 Jun 12;20(1):290 [FREE Full text] [doi: 10.1186/s12872-020-01577-z] [Medline: 32532208]
4. Setny M, Jankowski P, Kamiński K, Gąsior Z, Haberka M, Czarnecka D, et al. Secondary prevention of coronary heart disease in Poland: does sex matter? Results from the POLASPIRE survey. Pol Arch Intern Med 2022 Mar 30;132(3):22 [FREE Full text] [doi: 10.20452/pamw.16179] [Medline: 34935325]

Edited by T Leung; submitted 13.02.22; this is a non–peer-reviewed article;accepted 24.05.22; published 29.06.22.

Please cite as:
Sioka C
To Screen or Not to Screen? At Which BMI Cut Point? Comment on "Obesity and BMI Cut Points for Associated Comorbidities: Electronic Health Record Study"
J Med Internet Res 2022;24(6):e37267
URL: https://www.jmir.org/2022/6/e37267
doi:10.2196/37267
PMID:35767333

©Chrissa Sioka. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 29.06.2022. 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: To Screen or Not to Screen? At Which BMI Cut Point? Comment on "Obesity and BMI Cut Points for Associated Comorbidities: Electronic Health Record Study"

Luke Funk[1,2], MD, MPH; Natalie Liu[1], MD

[1]Department of Surgery, University of Wisconsin-Madison, Madison, WI, United States

[2]Department of Surgery, William S Middleton VA, Madison, WI, United States

Corresponding Author:
Luke Funk, MD, MPH
Department of Surgery
University of Wisconsin-Madison
600 Highland Ave
Madison, WI, 53597
United States
Phone: 1 608 263 1036
Email: funk@surgery.wisc.edu

Related Articles:

Comment on: http://www.jmir.org/2021/8/e24017/

Comment on: http://www.jmir.org/2022/6/e37267/

(J Med Internet Res 2022;24(6):e39717) doi:10.2196/39717

KEYWORDS

obesity; body mass index; BMI; risk factors; screening; health services; chronic disease; heart disease; myocardial perfusion imaging; anxiety; depression

We thank our colleague from Greece [1] for her interest in our article [2]. Similar to our study, Fotopoulos et al [3] found an association between obesity and coronary artery disease (CAD) in their analysis of patients undergoing cardiac stress tests. They also reported that the presence of obesity and depression together was associated with CAD. Similarly, the presence of obesity and anxiety together was associated with CAD. Our study did not explicitly measure associations between depression and CAD or anxiety and CAD, but Sioka [1] raised important points about how obesity, mental health, and heart disease may interact.

In the adjusted quantile regression analyses (Multimedia Appendix 4 of our paper), we found that patients with a diagnosis of anxiety had a similar BMI as those without anxiety. One systematic review of the literature suggested a positive association between obesity and anxiety, although a causal relationship has not been established [4]. Patients in our study who had a diagnosis of depression had a slightly higher median BMI than those without depression (0.74 BMI points, 95% CI 0.53-0.94). A meta-analysis of 15 longitudinal studies concluded that obesity increased the risk of depression, and depression was predictive of developing obesity [5].

Our study [2] and the study by Fotopoulos and colleagues [3] both reinforce the concept that obesity is associated with negative health outcomes that affect numerous body systems. Incorporating BMI into screening guidelines for conditions like CAD may help identify high-risk individuals so they can be intervened on earlier than current guidelines support.

Conflicts of Interest

None declared.

References

1. Sioka C. To screen or not to screen? At which BMI cut point? Comment on "Obesity and BMI cut points for associated comorbidities: Electronic health record study". J Med Internet Res 2022;24(6). [doi: 10.2196/37267]

XSL•FO
RenderX

2. Liu N, Birstler J, Venkatesh M, Hanrahan L, Chen G, Funk L. Obesity and BMI cut points for associated comorbidities: Electronic health record study. J Med Internet Res 2021 Aug 09;23(8):e24017 [FREE Full text] [doi: 10.2196/24017] [Medline: 34383661]
3. Fotopoulos A, Petrikis P, Iakovou I, Papadopoulos A, Sakelariou K, Gkika E, et al. The impact of depression and anxiety in prognosis of patients undergoing myocardial perfusion imaging with 99mTc tetrofosmin SPECT for evaluation of possible myocardial ischemia. Nucl Med Rev Cent East Eur 2020 Jul 31;23(2):58-62 [FREE Full text] [doi: 10.5603/NMR.a2020.0014] [Medline: 33007091]
4. Gariepy G, Nitka D, Schmitz N. The association between obesity and anxiety disorders in the population: a systematic review and meta-analysis. Int J Obes (Lond) 2010 Mar 8;34(3):407-419. [doi: 10.1038/ijo.2009.252] [Medline: 19997072]
5. Luppino FS, de Wit LM, Bouvy PF, Stijnen T, Cuijpers P, Penninx BW, et al. Overweight, obesity, and depression: A systematic review and meta-analysis of longitudinal studies. Arch Gen Psychiatry 2010 Mar;67(3):220-229. [doi: 10.1001/archgenpsychiatry.2010.2] [Medline: 20194822]

Abbreviations

CAD: coronary artery disease

Edited by T Leung; submitted 20.05.22; this is a non–peer-reviewed article;accepted 24.05.22; published 29.06.22.

Please cite as:
Funk L, Liu N
Authors' Reply to: To Screen or Not to Screen? At Which BMI Cut Point? Comment on "Obesity and BMI Cut Points for Associated Comorbidities: Electronic Health Record Study"
J Med Internet Res 2022;24(6):e39717
URL: https://www.jmir.org/2022/6/e39717
doi:10.2196/39717
PMID:35767330

©Luke Funk, Natalie Liu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 29.06.2022. 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: Predictors of Health Information–Seeking Behavior: Systematic Literature Review and Network Analysis

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

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

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

[*]these authors contributed equally

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

Related Article:

Correction of: https://www.jmir.org/2021/7/e21680

(J Med Internet Res 2022;24(6):e39705) doi:10.2196/39705

In "Predictors of Health Information–Seeking Behavior: Systematic Literature Review and Network Analysis" (J Med Internet Res 2021;23(7):e21680) the authors noted three errors.

1. In the originally published article, Reference 11 was incorrectly published as follows:

> *Cole C. Looking for information: a survey of research on information seeking, needs, and behavior (4th edition). Donald O. Case and Lisa M. Given. Bingley, UK: Emerald Group Publishing, 2016. 528 pp. $82.95 (hardcover). (ISBN: 9781785609688). J Asso Inf Sci Technol 2016 Dec 21;68(9):2284-2286. [doi: 10.1002/asi.23778]*

The correct reference is a book and has been updated as follows:

> *Case DO, Given LM. Looking for information: a survey of research on information seeking, needs, and behavior (4th edition). Bingley, UK: Emerald Group Publishing; 2016.*

2. The in-text citation for reference 11 incorrectly mentioned the year as 2002 in the following sentence:

> *Other known models from the information science perspective include the Comprehensive Model of Information Seeking, which looks at information carrier characteristics, antecedents, and information-seeking actions [10] and the book by Case in 2002 about the research on information-seeking needs and behaviors [11].*

The correct year for the current edition of the book is 2016 and the sentence has been updated as follows:

> *Other known models from the information science perspective include the Comprehensive Model of Information Seeking, which looks at information carrier characteristics, antecedents, and information-seeking actions [10] and the book by Case in 2016 about the research on information-seeking needs and behaviors [11].*

3. In affiliations 1 and 2 of the originally published article, the city was incorrectly mentioned as 'The University of Sydney'. This has been corrected to 'Sydney,' and the correct list of affiliations is as follows:

> *Ardalan Mirzaei[1*], BPharm, MPhil, GCertEdStud (Higher Ed); Parisa Aslani[1*], BPharm, MSc, PhD, GCertEdStud (Higher Ed); Edward Joseph Luca[2], BA, MBA; Carl Richard Schneider[1*], BN, BPharm, PhD, PGCert (Higher Ed)*
>
> *[1]School of Pharmacy, Faculty of Medicine and Health, University of Sydney, Sydney, Australia*
>
> *[2]The University of Sydney Library, The University of Sydney, Sydney, Australia*

XSL•FO
RenderX

*these authors contributed equally

The correction will appear in the online version of the paper on the JMIR Publications website on June 3, 2022, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 19.05.22; this is a non–peer-reviewed article;accepted 19.05.22; published 03.06.22.

Please cite as:
Mirzaei A, Aslani P, Luca EJ, Schneider CR
Correction: Predictors of Health Information–Seeking Behavior: Systematic Literature Review and Network Analysis
J Med Internet Res 2022;24(6):e39705
URL: https://www.jmir.org/2022/6/e39705
doi:10.2196/39705
PMID:35660657

©Ardalan Mirzaei, Parisa Aslani, Edward Joseph Luca, Carl Richard Schneider. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.06.2022. 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: 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:

Correction of: https://www.jmir.org/2022/5/e37480

(J Med Internet Res 2022;24(6):e40038) doi:10.2196/40038

In "Durability of the Treatment Effects of an 8-Week Self-administered Home-Based Virtual Reality Program for Chronic Low Back Pain: Follow-up Study of a Randomized Clinical Trial" (J Med Internet Res 2022;24(5):e37480) the authors made one clarification.

In the originally published paper, the title appeared as follows:

"Durability of the Treatment Effects of an 8-Week Self-administered Home-Based Virtual Reality Program for Chronic Low Back Pain: Follow-up Study of a Randomized Clinical Trial"

In the corrected version of the paper, the title has been changed to:

"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"

The correction will appear in the online version of the paper on the JMIR Publications website on June 8, 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.

XSL•FO
RenderX

Submitted 02.06.22; this is a non–peer-reviewed article;accepted 06.06.22; published 08.06.22.

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
Correction: 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(6):e40038
URL: https://www.jmir.org/2022/6/e40038
doi:10.2196/40038
PMID:35675658

©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), 08.06.2022. 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: Age- and Sex-Specific Differences in Multimorbidity Patterns and Temporal Trends on Assessing Hospital Discharge Records in Southwest China: Network-Based Study

Liya Wang[1*], MA; Hang Qiu[1,2*], PhD; Li Luo[3], PhD; Li Zhou[4], MPA

[1]Big Data Research Center, University of Electronic Science and Technology of China, Chengdu, China

[2]School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China

[3]Business School, Sichuan University, Chengdu, China

[4]Health Information Center of Sichuan Province, Chengdu, China

[*]these authors contributed equally

Corresponding Author:
Hang Qiu, PhD
School of Computer Science and Engineering
University of Electronic Science and Technology of China
No.2006, Xiyuan Ave, West Hi-Tech Zone
Chengdu, 611731
China
Phone: 86 28 61830278
Fax: 86 28 61830278
Email: qiuhang@uestc.edu.cn

Related Article:

Correction of: https://www.jmir.org/2022/2/e27146

(J Med Internet Res 2022;24(6):e39648) doi:10.2196/39648

In “Age- and Sex-Specific Differences in Multimorbidity Patterns and Temporal Trends on Assessing Hospital Discharge Records in Southwest China: Network-Based Study” (J Med Internet Res 2022; 24(2): e27146) the authors noted one error.

In the originally published paper, the denominator of Equation 2 was incorrect and should have been expressed as a square root.

In the corrected version of the paper, the denominator of Equation 2 has been revised to a square root and is provided below. The calculations in the original publication were according to the correct version of Equation 2. The error in Equation 2 did not influence the results of the originally published paper.

The correction will appear in the online version of the paper on the JMIR Publications website on June 16, 2022, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 17.05.22; this is a non–peer-reviewed article;accepted 17.05.22; published 16.06.22.

Please cite as:
Wang L, Qiu H, Luo L, Zhou L
Correction: Age- and Sex-Specific Differences in Multimorbidity Patterns and Temporal Trends on Assessing Hospital Discharge Records in Southwest China: Network-Based Study
J Med Internet Res 2022;24(6):e39648
URL: https://www.jmir.org/2022/6/e39648
doi:10.2196/39648
PMID:35709490

XSL•FO
RenderX

©Liya Wang, Hang Qiu, Li Luo, Li Zhou. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 16.06.2022. 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: Adherence to Telemonitoring Therapy for Medicaid Patients With Hypertension: Case Study

Sulki Park[1,2], MS; Hye-Chung Kum[1,2,3], MS, MSW, PhD; Michael A Morrisey[3], PhD; Qi Zheng[4], PhD; Mark A Lawley[1,2,4], MS, PhD

[1]Population Informatics Lab, Texas A&M University, College Station, TX, United States

[2]Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX, United States

[3]Department of Health Policy and Management, Texas A&M University, College Station, TX, United States

[4]Department of Epidemiology and Biostatistics, Texas A&M University, College Station, TX, United States

Corresponding Author:
Hye-Chung Kum, MS, MSW, PhD
Department of Health Policy and Management
Texas A&M University
212 Adriance Lab Rd
College Station, TX, 77843
United States
Phone: 1 979 436 9439
Email: kum@tamu.edu

Related Article:

Correction of: https://www.jmir.org/2021/9/e29018

(J Med Internet Res 2022;24(6):e39666) doi:10.2196/39666

In "Adherence to Telemonitoring Therapy for Medicaid Patients With Hypertension: Case Study" (J Med Internet Res 2021;23(9):e29018), the authors noticed one error.

In the originally published article, the Acknowledgments section inadvertantly missed a statement acknowledging the telemonitoring company that provided the real-world data for the study. In the corrected version of the article, the following sentence has been added to the Acknowledgments section:

Real-world telemonitoring data was provided by Coordination Centric.

The correction will appear in the online version of the paper on the JMIR Publications website on June 17, 2022, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 17.05.22; this is a non–peer-reviewed article;accepted 18.05.22; published 17.06.22.

Please cite as:
Park S, Kum HC, Morrisey MA, Zheng Q, Lawley MA
Correction: Adherence to Telemonitoring Therapy for Medicaid Patients With Hypertension: Case Study
J Med Internet Res 2022;24(6):e39666
URL: https://www.jmir.org/2022/6/e39666
doi:10.2196/39666
PMID:35714353

©Sulki Park, Hye-Chung Kum, Michael A Morrisey, Qi Zheng, Mark A Lawley. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 17.06.2022. 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

XSL•FO
RenderX

cited. The complete bibliographic information, a 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: Access to Electronic Personal Health Records Among Patients With Multiple Chronic Conditions: A Secondary Data Analysis

Alexandra J Greenberg[1], MPH, PhD; Angela L Falisi[2], MPH; Lila J Finney Rutten[1], MPH, PhD; Wen-Ying Sylvia Chou[2], MPH, PhD; Vaishali Patel[3], MPH, PhD; Richard P Moser[2], PhD; Bradford W Hesse[2], PhD

[1]Mayo Clinic, Rochester, MN, United States

[2]Behavioral Research Program, National Cancer Institute, Bethesda, MD, United States

[3]Office of the National Coordinator, US Department of Health and Human Services, Washington, DC, United States

Corresponding Author:
Alexandra J Greenberg, MPH, PhD
Mayo Clinic
200 First Street SW
Rochester, MN, 55905
United States
Phone: 1 507 538 7388
Fax: 1 507 555 1212
Email: worisek.alexandra@gmail.com

Related Article:

Correction of: https://www.jmir.org/2017/6/e188/

(J Med Internet Res 2022;24(6):e39719) doi:10.2196/39719

In "Access to Electronic Personal Health Records Among Patients With Multiple Chronic Conditions: A Secondary Data Analysis" (J Med Internet Res 2017;19(6):e188), the authors made the following updates.

The authors were notified of data errors in one of the Health Information National Trends Survey (HINTS) cycle datasets (HINTS 4, Cycle 4); the errors were in the weights provided for use in the analysis of these data [1]. Following the HINTS error notice [1], the authors reran analyses reported in Table 1 and Table 2. The originally published versions of these tables are in Multimedia Appendix 1 and Multimedia Appendix 2.

In rerunning analyses for Table 1, only one difference was found that resulted in a change in conclusion. Namely, the chi-squared analysis of "Confidence that PHI is safe" versus "Number of Chronic Conditions" was not significant in the updated analysis (P=.11). In the original analysis, the results were significant, with "Very Confident" more likely with two or more chronic conditions, and "Not Confident" more likely with no chronic conditions. Additionally, for "Accessed EHRs at least once" and "Frequency of EHR Access," the original table used column percentages instead of row percentages; this has been corrected in the updated table.

In section "Associations Between Patient Factors and Number of Chronic Conditions" in the *Results*, the fourth sentence in the first paragraph originally read as follows:

> *In general, having two or more chronic conditions was associated with being older, having health insurance, having a regular provider, being less confident in taking care of themselves, reporting fair to poor health, being less inclined to use the Internet or to use a mobile phone/tablet, and in feeling more confident that their PHI is safe and controllable (Table 1).*

It has been corrected as follows:

> *In general, having two or more chronic conditions was associated with being older, having health insurance, having a regular provider, being less confident in taking care of themselves, reporting fair to poor health, and being less inclined to use the Internet or to use a mobile phone/tablet (Table 1).*

The following text from the *Discussion* was no longer accurate and has been removed from the corrected version of the article:

> *Additionally, HINTS included items addressing concerns about safety and privacy of electronic health information, which revealed that those with MCC reported slightly higher frequencies of believing that they were "very confident" in having control of the privacy of their records and that their PHI was safe with their providers. This could provide one explanation for the increased use of ePHR among those with MCC.*

XSL•FO
RenderX

In rerunning analyses for Table 2, only one difference was found that resulted in a change in conclusion. Namely, for "Confidence that PHI is safe," a significant association was found for "Very Confident" (OR 2.00, 95% CI 1.21-3.31; *P*=.01) and "Somewhat Confident" (OR 1.99, 95% CI 1.25-3.17) as compared to the reference of "Not Confident." In the original analysis, neither was statistically significant. This change does not affect any of the text within the body of the manuscript. But it is a notable new conclusion, indicating that those who are more confident about the safety of their data are significantly more likely to use ePHR than those who are not confident about such security.

The corrected versions of Table 1 and Table 2 are below:

XSL•FO
RenderX

Table 1. Associations between patient characteristics, online characteristics, and attitudes with number of chronic conditions (N=3497).

Respondent characteristics	Number of chronic conditions, n (weighted %)[a]			χ^2 (*df*)	*P* value
	0	1	≥2		
Overall	1050 (39.8)	892 (26.0)	1555 (34.1)		
Sex				5.5 (2)	.007
Female	624 (36.2)	543 (27.3)	920 (36.5)		
Male	420 (43.5)	344 (25.2)	605 (31.3)		
Age (years)				59.5 (8)	<.001
18-34	283 (64.6)	117 (25.4)	60 (10.0)		
35-49	322 (48.1)	200 (24.2)	188 (27.8)		
50-64	272 (30.9)	299 (28.9)	556 (40.3)		
65-74	72 (14.9)	149 (27.0)	349 (58.1)		
≥75	28 (6.3)	74 (21.4)	274 (72.2)		
Race/ethnicity				4.1 (8)	.001
Hispanic	185 (41.9)	132 (28.5)	194 (29.6)		
Non-Hispanic White	555 (38.7)	496 (26.1)	844 (35.2)		
Non-Hispanic Black	142 (39.9)	123 (24.0)	252 (36.1)		
Non-Hispanic other	94 (57.4)	56 (22.5)	85 (20.0)		
Missing	74 (28.9)	85 (26.9)	180 (44.2)		
Education				12.1 (6)	<.001
Less than high school	59 (27.5)	65 (27.0)	163 (45.6)		
High school graduate	171 (38.1)	140 (23.0)	323 (38.9)		
Some college	257 (34.9)	282 (27.0)	511 (38.1)		
College graduate	534 (49.6)	377 (26.7)	500 (23.7)		
Income (US$)				8.3 (8)	<.001
<$20,000	163 (31.5)	182 (27.0)	442 (41.5)		
$20,000 to <$35,000	127 (29.1)	116 (22.1)	265 (48.8)		
$35,000 to <$50,000	145 (40.6)	139 (26.7)	220 (32.7)		
$50,000 to <$75,000	180 (38.2)	153 (26.7)	252 (35.0)		
≥$75,000	421 (48.7)	295 (26.5)	349 (24.9)		
Health insurance				9.3 (2)	<.001
Yes	872 (38.3)	768 (25.8)	1397 (35.9)		
No	168 (51.6)	110 (27.2)	130 (21.2)		
Regular provider				50.1 (2)	<.001
Yes	548 (32.4)	612 (25.1)	1256 (42.5)		
No	494 (54.3)	266 (27.7)	268 (18.0)		
Self-reported ability to take care of own health				9.7 (4)	<.001
Completely confident/very confident	787 (42.9)	629 (27.2)	890 (29.8)		
Somewhat confident	224 (34.8)	231 (24.8)	518 (40.4)		
A little confident/not at all confident	36 (25.9)	29 (19.2)	137 (54.9)		
Self-reported general health				52.0 (4)	<.001
Excellent/very good	675 (53.3)	443 (26.4)	427 (20.3)		
Good	301 (31.5)	345 (28.8)	672 (39.7)		
Fair/Poor	69 (17.9)	97 (17.5)	443 (64.7)		

XSL•FO
RenderX

Respondent characteristics	Number of chronic conditions, n (weighted %)[a]			χ^2 (*df*)	*P* value
	0	1	≥2		
Regular Internet use				35.9 (2)	<.001
Yes	923 (42.9)	709 (26.1)	1077 (31.0)		
No	123 (25.3)	173 (25.4)	455 (49.2)		
Accessed EHRs at least once				0.5 (2)	.61
Yes	284 (28.0)	250 (26.7)	371 (25.5)		
No	757 (72.0)	630 (73.3)	1158 (74.5)		
Frequency of EHR access				5.6 (8)	<.001
Never	757 (72.0)	630 (73.3)	1158 (74.5)		
1-2 times	158 (15.4)	124 (14.2)	153 (9.1)		
3-5 times	74 (7.5)	78 (7.4)	101 (7.6)		
6-9 times	24 (2.2)	30 (3.4)	57 (3.7)		
≥10 times	28 (2.9)	18 (1.6)	60 (5.1)		
Use a mobile phone or tablet				36.7 (2)	<.001
Yes	848 (44.9)	610 (26.0)	854 (29.1)		
No	185 (25.7)	256 (26.2)	638 (48.1)		
Use health-related mobile phone/tablet apps				0.7 (2)	.49
Yes	297 (46.0)	204 (24.4)	295 (29.6)		
No	522 (44.2)	388 (27.9)	516 (27.9)		
Exchanged emails with provider(s)				0.3 (2)	.76
Yes	246 (42.0)	206 (25.7)	331 (32.3)		
No	791 (39.6)	662 (26.1)	1179 (34.3)		
Confidence that PHI is safe				2.0 (2)	.11
Very confident	207 (38.9)	178 (23.7)	389 (37.5)		
Somewhat confident	534 (38.3)	473 (27.4)	809 (34.3)		
Not confident	295 (44.4)	221 (24.9)	324 (30.8)		
Control privacy of records				3.3 (4)	.02
Very confident	255 (34.5)	246 (26.8)	487 (38.7)		
Somewhat confident	479 (39.2)	420 (26.0)	733 (34.7)		
Not confident	307 (47.6)	215 (25.2)	302 (27.1)		
Ever withheld information due to privacy concern				0.4 (2)	.66
Yes	160 (43.0)	128 (24.4)	222 (32.6)		
No	882 (39.4)	754 (26.3)	1306 (34.2)		
Concerned about security of information when sent between providers				1.2 (4)	.32
Very concerned	226 (41.9)	191 (25.9)	338 (32.3)		
Somewhat concerned	510 (40.4)	431 (24.4)	756 (35.3)		
Not concerned	305 (37.9)	259 (28.9)	433 (33.2)		

[a] Percentages are weighted.

XSL•FO
RenderX

Table 2. Weighted multivariate logistic regression model of predictors of using electronic personal health records among those reporting having Internet access or who own a mobile phone (n=2941).

Predictors of use of electronic personal health records	OR (95% CI)	Beta (SE)	Adj Wald *F* (*df*)	*P* value
Number of chronic conditions			4.51 (2)	.02
0	Ref	Ref		
1	0.98 (0.60-1.59)	-0.02 (0.24)		
≥2	1.88 (1.09-3.24)	0.63 (0.27)		
Sex			0.13 (1)	.72
Male	Ref	Ref		
Female	1.06 (0.77-1.45)	0.16 (0.16)		
Age (years)			2.05 (4)	.10
≥75	Ref	Ref		
65-74	1.80 (0.69-4.66)	0.59 (0.48)		
50-64	2.39 (1.01-5.67)	0.87 (0.43)		
35-49	2.68 (1.13-6.36)	0.98 (0.43)		
18-34	3.23 (1.24-8.41)	1.17 (0.47)		
Race/ethnicity			0.98 (4)	.43
Non-Hispanic White	Ref	Ref		
Hispanic	0.62 (0.31-1.26)	-0.47 (0.35)		
Non-Hispanic Black	0.90 (0.57-1.42)	-0.11 (0.23)		
Non-Hispanic other	1.34 (0.70-2.55)	0.29 (0.32)		
Missing	0.47 (0.14-1.54)	-0.76 (0.59)		
Education			1.35 (3)	.27
Less than high school	Ref	Ref		
High school graduate	1.22 (0.25-5.88)	0.20 (0.78)		
Some college	1.51 (0.35-6.52)	0.41 (0.73)		
College graduate	1.85 (0.41-8.31)	0.61 (0.75)		
Income (US$)			3.04 (4)	.03
<$20,000	Ref	Ref		
$20,000 to <$35,000	1.90 (0.81-4.47)	0.42 (-0.21)		
$35,000 to <$50,000	2.75 (1.25-6.08)	0.39 (0.22)		
$50,000 to <$75,000	1.89 (0.85-4.23)	0.40 (-0.16)		
≥$75,000	3.17 (1.50-6.71)	0.37 (0.41)		
Health insurance			1.71 (1)	.20
No	Ref	Ref		
Yes	1.48 (0.81-2.71)	0.30 (-0.21)		
Regular provider			7.43 (1)	.01
No	Ref	Ref		
Yes	1.84 (1.17-2.88)	0.61 (0.22)		
Self-reported ability to take care of own health			0.21 (2)	.81
A little confident/not at all confident	Ref	Ref		
Somewhat confident	0.97 (0.40-2.34)	-0.03 (0.44)		
Completely confident/very confident	1.14 (0.54-2.39)	0.13 (0.37)		
Self-reported general health			1.71 (2)	.19

XSL•FO
RenderX

Predictors of use of electronic personal health records	OR (95% CI)	Beta (SE)	Adj Wald *F* (*df*)	*P* value
Excellent/very good	Ref	Ref		
Good	1.40 (0.94-2.09)	0.34 (0.20)		
Fair/Poor	1.04 (0.52-2.10)	0.04 (0.35)		
Confidence that PHI is safe			5.24 (2)	.01
Not confident	Ref	Ref		
Somewhat confident	1.99 (1.25-3.17)	0.69 (0.23)		
Very confident	2.00 (1.21-3.31)	0.69 (0.25)		

In addition, the corresponding author's email address has been changed to worisek.alexandra@gmail.com, as the author is no longer affiliated with Mayo Clinic College of Medicine and Science.

The correction will appear in the online version of the paper on the JMIR Publications website on June 20, 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 1.
[PDF File (Adobe PDF File), 69 KB - jmir_v24i6e39719_app1.pdf]

Multimedia Appendix 2
Originally published Table 2.
[PDF File (Adobe PDF File), 47 KB - jmir_v24i6e39719_app2.pdf]

Reference

1. HINTS Data Errors, Remediation, and Recommendations. Health Information National Trends Survev. URL: https://hints.cancer.gov/data/data-remediation.aspx [accessed 2022-05-19]

Abbreviations

HINTS: Health Information National Trends Survey

Submitted 20.05.22; this is a non–peer-reviewed article;accepted 20.05.22; published 20.06.22.

Please cite as:
Greenberg AJ, Falisi AL, Finney Rutten LJ, Chou WYS, Patel V, Moser RP, Hesse BW
Correction: Access to Electronic Personal Health Records Among Patients With Multiple Chronic Conditions: A Secondary Data Analysis
J Med Internet Res 2022;24(6):e39719
URL: https://www.jmir.org/2022/6/e39719
doi:10.2196/39719
PMID:35724378

©Alexandra J Greenberg, Angela L Falisi, Lila J Finney Rutten, Wen-Ying Sylvia Chou, Vaishali Patel, Richard P Moser, Bradford W Hesse. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.06.2022. 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

Media Use During the COVID-19 Pandemic: Cross-sectional Study

Marjolaine Rivest-Beauregard[1,2], MSc; Justine Fortin[2,3], MSc; Connie Guo[2]; Sabrina Cipolletta[4], PhD; Ram P Sapkota[1], PhD; Michelle Lonergan[2,5], PhD; Alain Brunet[1,2], PhD

[1]Department of Psychiatry, McGill University, Montréal, QC, Canada

[2]Division of Psychosocial Research, Douglas Mental Health University Institute, Montréal, QC, Canada

[3]Department of Psychology, University of Québec in Montréal, Montréal, QC, Canada

[4]Department of Psychology, University of Padua, Padua, Italy

[5]School of Psychology, University of Ottawa, Ottawa, ON, Canada

Corresponding Author:
Marjolaine Rivest-Beauregard, MSc
Division of Psychosocial Research
Douglas Mental Health University Institute
Perry Building, Third Floor
6875 bd LaSalle
Montréal, QC, H4H 1R3
Canada
Phone: 1 514 761 6131 ext 3473
Email: marjolaine.rivest-beauregard@mail.mcgill.ca

Abstract

Background: Throughout the pandemic, the general population was encouraged to use media to be kept informed about sanitary measures while staying connected with others to obtain social support. However, due to mixed findings in the literature, it is not clear whether media use in such a context would be pathogenic or salutogenic.

Objective: Therefore, the associations between COVID-19–related stressors and frequency of media use for information-seeking on trauma- and stressor-related (TSR) symptoms were examined while also investigating how social media use for support-seeking and peritraumatic distress interact with those variables.

Methods: A path model was tested in a sample of 5913 adults who completed an online survey.

Results: The number of COVID-19–related stressors (β=.25; *P*<.001) and extent of information-seeking through media (β=.24; *P*=.006) were significantly associated with the severity of TSR symptoms in bivariate comparisons. Associations between levels of peritraumatic distress and both COVID-19–related stressors and information-seeking through media, and social media use for support- and information-seeking through media were found ($\beta_{\text{COVID-19 stressors: Peritraumatic Distress Inventory}}$=.49, *P*<.001; $\beta_{\text{seeking information: Peritraumatic Distress Inventory}}$=.70, *P*<.001; $\beta_{\text{seeking information–seeking support}}$=.04, *P*<.001).

Conclusions: Results suggest that exposure to COVID-19–related stressors and seeking COVID-19–related information through the media are associated with higher levels of peritraumatic distress that, in turn, lead to higher levels of TSR symptoms. Although exposure to the stress of the COVID-19 pandemic may be unavoidable, the frequency of COVID-19–related information consumption through various media should be approached with caution.

(J Med Internet Res 2022;24(6):e33011) doi:10.2196/33011

KEYWORDS

media use; support; information-seeking behaviors; trauma- and stressor-related symptoms; COVID-19; media; information-seeking; behavior; trauma; stress; symptom; frequency; risk; distress

XSL•FO
RenderX

Introduction

The COVID-19 Pandemic, Media Use, and Mental Health

The COVID-19 pandemic has resulted in global destabilization and repeated confinement of billions of people to their homes, and has made *physical distancing* a new way of life. In the early days of the pandemic and for quite some time, governmental authorities have broadcasted bleak news updates through the media daily (eg, [1,2]) and have urged people to stay connected to each other virtually to offset the absence of social contacts (eg, [3]). While some literature suggests that using social media platforms for support is salutogenic (ie, supporting health and well-being) [4,5], others suggest that media consumption involves repeated exposure to aversive content inducing pathogenic effects, including trauma- and stressor-related (TSR) symptoms typically associated with adjustment disorder or posttraumatic stress disorder (PTSD), as per the *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition) [6-8].

The reliance on media (eg, social media, television, or news reports) during a global crisis like the COVID-19 pandemic has multifaceted benefits, including the dissemination of urgent information and public health guidelines. In previous infectious disease outbreaks, the provision of information through the media has fostered preventive behaviors in the general population, such as wearing a mask, avoiding crowded public spaces, and handwashing [9]. However, this media presence may also come at the cost of experiencing psychological distress, including depression [10], PTSD symptoms [10-13], and anger [9], particularly if the media disseminate bleak or sensationalistic information or misinformation on a large scale [14], a phenomenon known as an "infodemic" [15,16]. That being said, the silver lining is that social media can also be used for various other purposes, such as seeking social connection and support [17-20], which can be associated with better mental health outcomes, including lower levels of anxiety, depression, and stress [5,18].

Empirical findings on the relation between different forms of media use and mental health have been mixed (eg, [21,22]), highlighting the complex nature of this association. In the context of a mass disaster such as the COVID-19 pandemic [23] and given the stressful and traumatic nature of this event for some [6,24], one must consider an array of trauma-related variables to better understand how media use is associated with TSR outcomes. Indeed, peritraumatic distress is a strong predictor of TSR symptoms [25], which could, in turn, be further exacerbated by media use when exposed to more gory details of the events [26]. Such associations should be taken into consideration when contemplating media use in a stressful or traumatic context.

Additionally, much of the current evidence regarding the pathogenic effects of media use is based on social media such as Facebook (eg, [21,22]). However, media use goes far beyond social media and is a constantly evolving domain. We use the term "media" to refer to traditional and new media. The concept, as used in this paper, includes all classical mass media such as newspapers, magazines, radio, and television, and their online versions (the latter belong to mainstream news media, even if they use an electronic format). "Social media" is used to designate various forms of consumer-generated content such as blogs, social network sites, forums, virtual communities, online newspaper reader comments, and media files shared on sites such as YouTube. Additionally, several studies have distinguished between passive (involves information consumption) and active (involves connectiveness) media use (eg, [27,28]), suggesting that active use is associated with positive outcomes on one's mental health. However, the literature on passive media use still presents mixed results [27], highlighting the need for more fine-grained research. In this manuscript, *media use* is defined as media engagement for informative purposes and media engagement for connectiveness and support purposes.

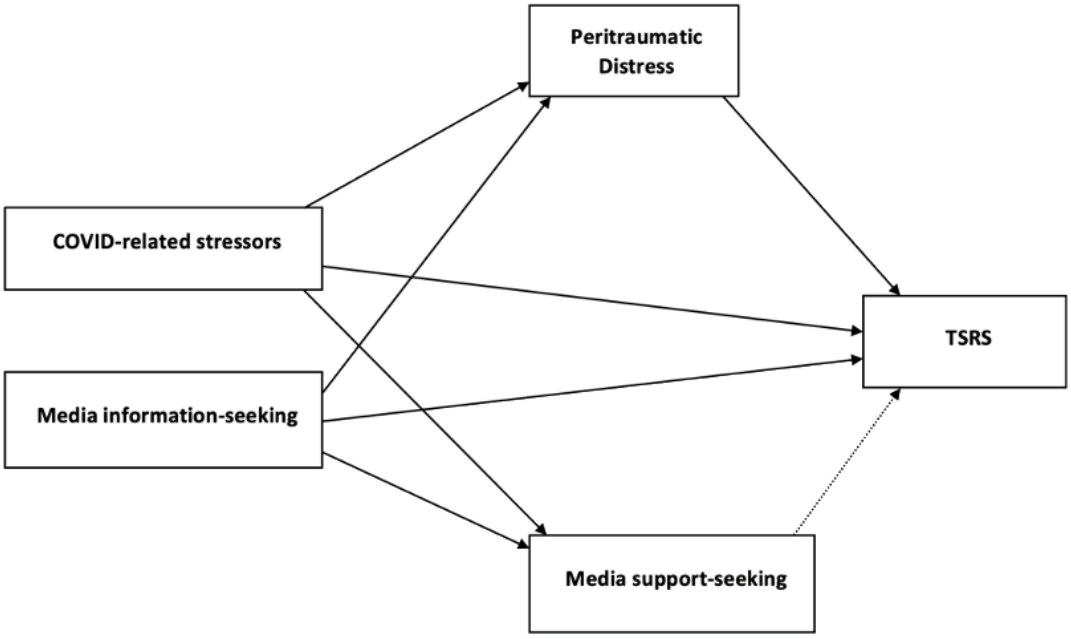
The differential susceptibility to media effects model (DSMM) is a framework that ties together media- and non-media–related variables so as to better understand the complexity of the associations between media use and mental health [29]. According to the DSMM [29], many variables, such as social context and individual differences including cognitive, affective, behavioral, and physiological factors, moderate the associations between media use and mental health. For instance, differential-susceptibility variables, whether dispositional (eg, gender, motivations, or values), developmental (eg, cognitive, social, or emotional development), or social (eg, peers, family, or work context), can be moderators of the association with mental health. However, the DSMM also suggests that the association between media use and mental health are indirect and can be mediated by various cognitive and emotional factors [29]. Previous research has used this model to investigate the pathogenic effects of media exposure following other mass disasters such as the 9/11 attacks and the Iraq war [13,30]. More recently, this model has been applied in the context of the COVID-19 pandemic [31], where the authors investigated the mediating role of negative affect on the association between social media use and TSR symptoms; however, the authors did not include a measure of peritraumatic distress, one of the strongest predictors of TSR symptoms [25].

Proposed Model and Hypotheses

We operationalized media use as defined by the authors of the DSMM: the broad use of various media types such as social networks, virtual environments, and traditional media (eg, newspapers and televisions) [29]. We also further divided media use according to the intended use (ie, information-seeking behaviors and support-seeking behaviors). Thus, in this research, we aimed to explore the associations between COVID-19–related stressors and media use for information-seeking with TSR symptoms, and the role of media use for support-seeking and peritraumatic distress and this association. We hypothesized that COVID-19–related stressors and using media for COVID-19–related information-seeking would be positively associated with peritraumatic distress, using media connection and support-seeking, and TSR symptoms. Additionally, we hypothesized that peritraumatic distress will be positively associated with TSR symptoms, while support-seeking behaviors will be negatively associated with TSR symptoms (see Figure 1).

XSL•FO
RenderX

Figure 1. Hypothetical path model. Solid lines represent positive relationships, while the dotted line represents a negative relationship. TSRS: trauma- and stressor-related symptoms.



Methods

Sample and Recruitment

A convenience sample of 5913 adults from Canada, France, Italy, the United States, and China took part in a web-based survey in April/May 2020 on the psychosocial effects of the COVID-19 pandemic. Participants were recruited through the *snowball technique*. Email invitations were sent to various professional and student associations and to individuals, and an advertisement was shared on social media (ie, Facebook and Twitter).

Survey and Procedure

The survey was hosted on *SurveyMonkey*. Potential participants were reached through e-mails or social media platforms and were invited to click on the online survey link. Participants were next directed to an online consent form. Before participating in the survey (in one sitting), participants were consented and confirmed that they were 18 years or older. The survey completion rate was very high (n=5913, 92%) among those who clicked on the survey hyperlink. The proportion who received the survey but did not click on the hyperlink remains unknown, however.

Ethics Approval

Ethics approval was obtained from the Douglas Mental Health University Institute (IUSMD-20-13).

Measures

Media Use

Frequency of media use was assessed with two self-report questions rated on a 5-point scale ranging from 0 (never) to 4 (very often), as similarly done by others [7] in the context of the COVID-19 pandemic [32]: "I maintain closeness and receive the support that I need through social networks and messaging apps" (support-seeking item), and "I looked for and shared information and news on COVID-19 on traditional media, on the internet, or on social networks" (information-seeking item).

TSR Symptoms

Severity of TSR symptoms over the previous 7 days was measured using the abridged (6-item) self-report Impact of Event Scale–Revised (IES-6) [33]. The IES-6 measures the severity of intrusions, avoidance, and hypervigilance in response to a stressor or a traumatic event. Items are rated on a 5-point scale, ranging from 0 (not at all) to 4 (extremely), and higher scores (sum of all items; range 0-24) indicate greater symptom severity. A cutoff score of 10 was found to identify clinically significant levels of symptoms [33]. This measure was selected for its comparative performance with the 22-item original scale (r=0.95) [33].

Peritraumatic Distress

Peritraumatic distress was assessed using the 13-item Peritraumatic Distress Inventory (PDI) [34]. The PDI self-report measures the level of perceived life threat, as well as fear, helplessness, and horror that can occur during and immediately after exposure to a stressor or a traumatic event. Items are rated on a 5-point scale ranging from 0 (not at all) to 4 (extremely true). Higher scores (sum of all items; range 0-52) indicate greater levels of distress. A score of 13 or above indicates clinically significant distress experienced during or shortly after a negative life event [35].

COVID-19 Stressors

Exposure to COVID-19–related stressors was measured using a set of 19 dichotomous (1=yes and 0=no) items developed by the team that mapped onto previously used similar questionnaires [31,36]. Specific items included stressors related to the illness itself (eg, have you been diagnosed with COVID-19?), the risk of illness (eg, are you part of an at-risk

XSL•FO
RenderX

group?), and other stressors (eg, have you lost your job?). A severity score was calculated by summing all items, resulting in a score ranging between 0 and 19, with higher scores indicating a greater exposure to COVID-19–related stressors.

Statistical Analyses

All analyses were performed using SPSS (version 27; IBM Corp) and AMOS (version 27; IBM Corp) except for the missing data imputation. A path analysis using a specification search was performed to identify the best fit for the model illustrated in Figure 1. The maximum likelihood method was used to estimate the parameters of the model. Multiple fit indexes were used to evaluate model fit [37]: the chi-square test of absolute fit, the comparative index fit (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR).

Missing Data

A total of 6409 respondents opened the survey link, from which 496 were removed from the database for not taking part in the study or not responding to any survey question. This led to a final sample size of 5913, in which 65% of the variables had less than 5% of missing values. According to Little's [38] missing completely at random test (χ^2_{1236}=2135.8; P<.001), data were not missing completely at random. Therefore, missing data was imputed using the k-nearest neighbor imputation method with k=5 using the VIM package for R (R Foundation for Statistical Computing) [39].

Results

Sample

The sample was composed of 5913 adults, most of whom identified as female (n=4681, 79.2%) and resided in Canada (n=1946, 32.9%), the United States (n=1302, 22%), Italy (n=1094, 18.5%), France (n=1036, 17.5%), and China (n=336, 5.7%). Almost half of the sample were essential workers (n=2717, 45.9%), the rest being nonessential workers (n=1578, 26.7%), stay-at-home occupations (ie, students, unemployed, and retired; n=676, 11.4%), and other uncategorized occupations (n=942, 15.9%). Additionally, the sample was, on average, aged 42 (SD 15.24) years. Table 1 presents the sociodemographic information of the sample.

Table 1. Sociodemographic, media, and clinical variables.

Sociodemographic, media, and clinical characteristics	Participants (N=5913)
Marital status, n (%)	
Single	1391 (23.52)
Dating/cohabiting/married	4043 (68.37)
Separated/divorced/widowed	479 (8.10)
Ethnicity, n (%)	
First Nations	142 (2.40)
Caucasian	4306 (72.82)
Black	68 (1.15)
Latino	316 (5.34)
Asian	617 (10.43)
Mixed	123 (2.08)
Other	341 (5.77)
Education, n (%)	
Preuniversity	837 (14.16)
Undergraduate level	2064 (34.91)
Graduate level	3012 (50.94)
Media variables, mean (SD)	
Support-seeking media use	2.88 (1.03)
Information-seeking media use	2.65 (1.10)
Clinical variables, mean (SD)	
Trauma- and stress-related symptoms[a]	11.24 (5.85)
Peritraumatic distress[b]	17.53 (10.56)

[a]Abridged 6-item Impact of Event Scale–Revised. A score of 10 or more denotes the presence of clinically significant symptoms [33].

[b]Peritraumatic Distress Inventory. A score of 13 or more is indicative of clinically significant symptoms [35].

XSL•FO
RenderX

Path Analysis: COVID-19–Related Stressors, Media Use, and TSR Symptoms

The path model was specified as hypothesized in Figure 1. Exposure to COVID-19 stressors and frequency of media use for COVID-19–related information (media information-seeking) were entered as predictors. Peritraumatic distress and frequency of media use for support and connection (media support-seeking) were entered as the intermediary variables, and TSR symptoms were entered as the outcome. As the specified path model was saturated (*df*=0), the nonsignificant direct association between COVID-19 stressors and media support-seeking (β=.004, SE 0.006; CR=0.58; *P*=.57) was removed from the final model. All endogenous variables in the model were ordered categorical with four or more categories. The multivariate kurtosis score of 9.94 suggested that the variables in the final model violated assumptions of multivariate normality [40]. Therefore, estimates and SEs were recalculated using bias-corrected machine learning–bootstrapping (5000 bootstrapped resamples; see Table 2). The lowest percentage of variance explained by the final model was for media support-seeking (R^2=0.09), while the highest percentage of variance explained was for TSR symptoms (R^2=0.57); the model explained 11% of the variance in peritraumatic distress. The final model demonstrated good to excellent indexes of fit (χ^2_1=0.3; *P*=.57; CFI=1; RMSEA<0.001; SRMR=0.002) [41,42].

Results from the path analysis revealed that COVID-19 stressors and media information-seeking were significantly and positively associated with peritraumatic distress ($P_{\text{COVID-19 stressors}}$<.001; $P_{\text{seeking information}}$<.001). TSR symptoms were weakly but positively associated with COVID-19 stressors (*P*<.001) and media information-seeking (*P*<.001), while peritraumatic distress was strongly and positively associated with TSR symptoms (*P*<.001). Media information-seeking (*P*<.001) and peritraumatic distress (*P*<.001) were both positively associated with media support-seeking. In contrast with our predictions, media support-seeking was significantly and positively associated with TSR symptoms (*P*=.006), although the effect was very small. Standardized parameter estimates of the direct associations can be found in Figure 1, and the unstandardized parameters estimates can be found in Table 2.

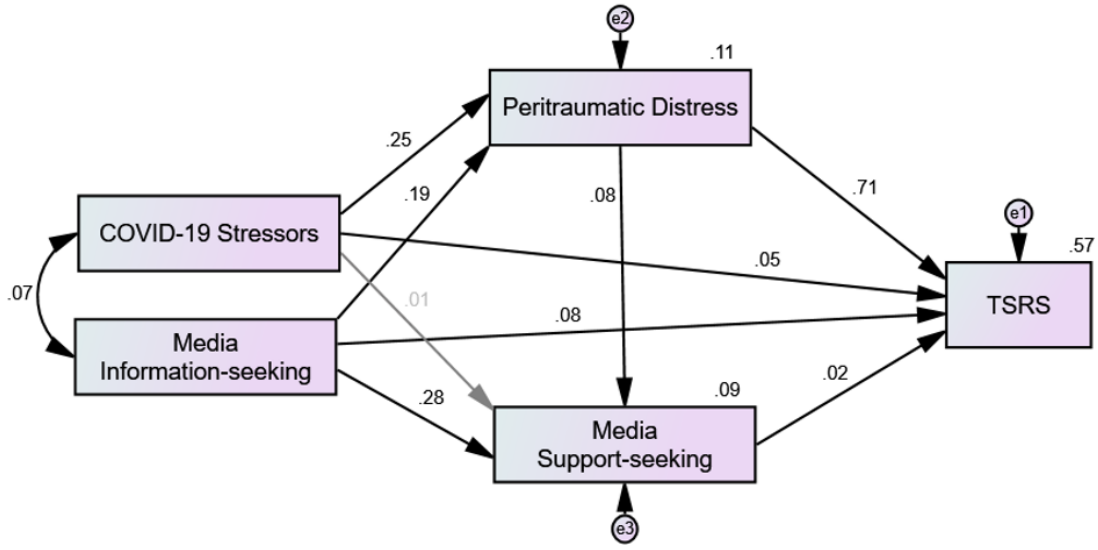
Black lines represent statistically significant associations, while the gray ones represent nonsignificant associations that were removed from the final model. Numbers above the lines represent standardized beta coefficients, which were used to interpret the strength of the associations. The explained variance (R^2) is indicated on the upper-right corner of the intermediary and outcomes variables. Error terms are represented by the e-labelled circles above the intermediary and outcome variables (Figure 2).

Table 2. Standardized parameter estimates of direct effects.

Independent variables	Dependent variables	Noncorrected estimates, B (SE)	Bias-corrected estimates		
			B (SE)	95% CI	*P* value
COVID-19 stressors	Peritraumatic distress	1.24 (0.06)	1.24 (0.08)	1.10-1.39	<.001
Media information	Peritraumatic distress	1.78 (0.12)	1.79 (0.12)	1.55-2.01	<.001
Peritraumatic distress	Media support	0.01 (<0.01)	0.01 (<0.01)	0.01-0.01	<.001
Media information	Media support	0.26 (0.01)	0.26 (0.01)	0.23-0.29	<.001
Media information	Distress	0.45 (0.05)	0.45 (0.05)	0.35-0.54	<.001
Media support	Distress	0.14 (0.05)	0.14 (0.05)	0.03-0.24	.008
Peritraumatic distress	Distress	0.39 (0.01)	0.39 (0.01)	0.38-0.40	<.001
COVID-19 stressors	Distress	0.14 (0.02)	0.14 (0.02)	0.09-0.18	<.001

XSL•FO
RenderX

Figure 2. Path analysis and results. TSRS: trauma- and stressor-related symptoms.



Analysis of Indirect Associations

The path analysis revealed that the associations between the independent variables (exposure to COVID-19–related stressors and media information-seeking) and the outcome variable (TSR symptoms) passed through levels of peritraumatic distress and media support-seeking. As per Table 3, both peritraumatic distress and support-seeking were intermediary variables, as they were both significantly correlated with at least one independent variable and the outcome [43].

As recommended by Preacher and Hayes [44], the 95% bias-corrected bootstrap CIs were computed to examine whether the indirect associations of the predictors on the outcome were significantly different from zero when using 5000 bootstrapped samples. The bootstrapped indirect associations of exposure to COVID-19 stressors and media information-seeking through peritraumatic distress on TSR symptoms were statistically significant ($\beta_{\text{COVID-19 stressors}}$=.49, 95% CI 0.43-0.55; $\beta_{\text{seeking information}}$=.70, 95% CI 0.61-0.79). Additionally, the indirect associations of media information-seeking through media support-seeking on TSR symptoms was also statistically significant (β=.04, 95% CI 0.01-0.06), although this association was negligible.

Table 3. Pearson correlation coefficients between variables included in the path analysis (N=5913)[a].

Variables	COVID[b]	IM[c]	SM[d]	PDI[e]	TSR[f]
COVID	—[g]				
IM	0.07	—			
SM	0.05	0.30	—		
PDI	0.28	0.20	0.14	—	
TSR	0.24	0.24	0.15	0.75	—

[a]*P*<.001 for all correlation coefficients. The correlation is significant at the .001 level.

[b]COVID: COVID-19–related stressors.

[c]IM: information-seeking media use.

[d]SM: support-seeking media use.

[e]PDI: peritraumatic distress.

[f]TSR: trauma- and stress-related symptoms.

[g]Not applicable.

Discussion

Media Use and Trauma- and Stressor-Related Symptoms

The goal of this study was to understand the combined associations between exposure to COVID-19–related stressors and media use for COVID-19–related information on the development of TSR symptoms while considering the role that peritraumatic distress and media use for support-seeking might play. In line with our main hypothesis, exposure to COVID-19 stressors and media use for obtaining COVID-19–related information led to elevated peritraumatic distress that, in turn, was associated with higher TSR symptoms. However, contrary to predictions, media use for support and connection did not

XSL•FO
RenderX

appear to be related with the association between exposure to COVID-19 stressors and TSR symptoms or between the use of media for obtaining COVID-19–related information and TSR symptoms. In fact, the association between media use for support and TSR symptoms was positive, rather than negative, suggesting that using social media for support, alone, may not be sufficient to protect against the mental health effects of exposure to the pandemic-related stressors. However, such an association may also be attributable to endogeneity issues, which can occur when observed associations are due to correlations between the model and error terms rather than the variables of interest [45]. Further longitudinal research is, therefore, necessary to better understand the associations and potential causality links between media use for support-seeking, peritraumatic distress, and TSR symptoms.

Peritraumatic Distress, Media Use, and Trauma- and Stressor-Related Symptoms

Peritraumatic distress was strongly associated not only with severity of TSR symptoms but also with exposure to COVID-19 stressors and seeking information through media on trauma-related symptoms. Such findings are consistent with prior literature, as the predictive power of peritraumatic distress on the later development and severity of TSR symptoms and, more precisely, of PTSD is well documented [25,46]. Moreover, in this study, both exposure to COVID-19–related stressors and media use for seeking COVID-19–related information were significantly and similarly associated with peritraumatic distress. Although this was slightly stronger for exposure to COVID-19 stressors, these results suggest that the frequency with which individuals consume COVID-19–related information through the media may be related with significant peritraumatic distress reactions, which in turn, can increase the severity of TSR symptoms. This is consistent with current literature suggesting that media exposure during the pandemic is associated with increased anxiety, depression, and secondary trauma symptoms across various populations [47-50]. Additionally, a statistically significant association was found between peritraumatic distress and media use for seeking support, suggesting that individuals who are more distressed may seek more support through social media. Further investigation of this issue is warranted, as it is not possible to verify the direction of the association with the current data.

Contrary to predictions, findings from this study suggest that media use for support and connection may not be sufficient to reduce the pathogenic effects of the pandemic. Considering that TSR symptoms result from exposure to environmental stressors, it is possible that individuals who frequently use media to actively seek support and connection during the pandemic are simultaneously exposed to overwhelming information about the pandemic [17], resulting in an increased risk of TSR symptoms. Importantly, however, the examination of different mental health outcomes may yield different results. For instance, some studies have suggested that, during the pandemic, individuals who sought support through social media reported lower levels of loneliness [51,52], while others have found that this behavior may not negatively affect levels of anxiety and mental health [18]. Thus, future studies may incorporate additional measures of mental health in the examination of the association between media use for support-seeking and mental health.

Limitations and Future Directions

The cross-sectional nature of the data prevents making causal inferences. For instance, the small association between support-seeking through media and TSR symptoms may be partly explained by the level of distress experienced by individuals [53]. Additionally, the media variables used in this study were constructed of single items, which may have affected the stability of the model. However, the large sample size and variability in the data allowed testing of a strong model and, consequently, increased confidence in the results. Importantly, though, given that the sample is composed of a higher proportion of female respondents, caution must be considered when generalizing the results. In this study, TSR symptoms were measured using a self-reported questionnaire, which may yield a somewhat higher symptom severity (eg, [54]). Further investigations will need to address the effect of media use on TSR symptoms longitudinally to understand this association overtime while also evaluating TSR symptoms using a clinician-administered measure. Finally, given that media use may affect other indicators of mental health, such as anxiety or depression, future research may wish to include additional mental health outcomes.

This paper highlights important considerations from both empirical and societal perspectives. First, future research should explore the effects of media on mental health according to how it is used, as results from this study emphasize the differences between the use of media to seek information and support on TSR symptoms. Second, considering the recommendations stemming from national and international guidelines [1-3], there is a need to inform the population on how the use of media impacts various aspects of mental health. Although it is important to remain connected and informed during these unprecedented times, too much use of traditional and social media may negatively impact mental health. Educating the public on how to use media in a crisis to remain informed without jeopardizing their mental health should be prioritized, as this could help prevent detrimental mental health consequences such as TSR symptoms.

Acknowledgments

The authors would like to thank the research team of the Psychological Trauma Laboratory at the Douglas Mental Health University Institute: Erika Di Paola, Cristina Bove, Romain Hassan Omar, Rubina Kasparian, and its collaborators who were involved in data collection.

XSL•FO
RenderX

AB received funding from the Internal Social Sciences and Humanities Development Grant from McGill University to conduct this study. Additionally, both MR-B and JF are financially supported by the Fonds de Recherche du Québec.

Data Availability

Data access can be granted upon request.

Conflicts of Interest

None declared.

References

1. COVID-19: outbreak update. Government of Canada. 2021. URL: https://tinyurl.com/53y9w2ph [accessed 2021-02-04]
2. COVID-19. Centers for Disease Control and Prevention. 2021. URL: https://www.cdc.gov/coronavirus/2019-ncov/index.html [accessed 2021-02-04]
3. Mental health and psychosocial considerations during the COVID-19 outbreak, 18 March 2020. World Health Organization. 2020 Mar 18. URL: https://apps.who.int/iris/handle/10665/331490 [accessed 2022-05-13]
4. Naslund JA, Aschbrenner KA, Marsch LA, Bartels SJ. The future of mental health care: peer-to-peer support and social media. Epidemiol Psychiatr Sci 2016 Apr;25(2):113-122 [FREE Full text] [doi: 10.1017/S2045796015001067] [Medline: 26744309]
5. Prescott J, Rathbone AL, Brown G. Online peer to peer support: qualitative analysis of UK and US open mental health Facebook groups. Digit Health 2020;6:2055207620979209 [FREE Full text] [doi: 10.1177/2055207620979209] [Medline: 33354335]
6. Bridgland VME, Moeck EK, Green DM, Swain TL, Nayda DM, Matson LA, et al. Why the COVID-19 pandemic is a traumatic stressor. PLoS One 2021;16(1):e0240146 [FREE Full text] [doi: 10.1371/journal.pone.0240146] [Medline: 33428630]
7. Chao M, Xue D, Liu T, Yang H, Hall BJ. Media use and acute psychological outcomes during COVID-19 outbreak in China. J Anxiety Disord 2020 Aug;74:102248 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102248] [Medline: 32505918]
8. Levaot Y, Greene T, Palgi Y. Making and receiving offers of help on social media following disaster predict posttraumatic growth but not posttraumatic stress. Disaster Med Public Health Prep 2021 Aug;15(4):484-490 [FREE Full text] [doi: 10.1017/dmp.2020.43] [Medline: 32349839]
9. Oh S, Lee SY, Han C. The effects of social media use on preventive behaviors during infectious disease outbreaks: the mediating role of self-relevant emotions and Public Risk Perception. Health Commun 2021 Jul;36(8):972-981. [doi: 10.1080/10410236.2020.1724639] [Medline: 32064932]
10. Ahern J, Galea S, Resnick H, Kilpatrick D, Bucuvalas M, Gold J, et al. Television images and psychological symptoms after the September 11 terrorist attacks. Psychiatry 2002;65(4):289-300. [doi: 10.1521/psyc.65.4.289.20240] [Medline: 12530330]
11. Deroma V, Saylor C, Swickert R, Sinisi C, Marable TB, Vickery P. College students' PTSD symptoms, coping, and perceived benefits following media exposure to 9/11. J Coll Student Psychother 2003 Nov 10;18(1):49-64. [doi: 10.1300/j035v18n01_05]
12. Goodwin R, Lemola S, Ben-Ezra M. Media use and insomnia after terror attacks in France. J Psychiatr Res 2018 Mar;98:47-50. [doi: 10.1016/j.jpsychires.2017.12.006] [Medline: 29276963]
13. Holman EA, Garfin DR, Lubens P, Silver RC. Media exposure to collective trauma, mental health, and functioning: does it matter what you see? Clin Psychological Sci 2019 Oct 08;8(1):111-124. [doi: 10.1177/2167702619858300]
14. Basch CH, Hillyer GC, Erwin ZM, Mohlman J, Cosgrove A, Quinones N. News coverage of the COVID-19 pandemic: missed opportunities to promote health sustaining behaviors. Infect Dis Health 2020 Aug;25(3):205-209 [FREE Full text] [doi: 10.1016/j.idh.2020.05.001] [Medline: 32426559]
15. Novel Coronavirus(2019-nCoV): situation report - 13. World Health Organization. 2020. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf?sfvrsn=195f4010_6 [accessed 2022-05-16]
16. 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]
17. Saud M, Mashud M, Ida R. Usage of social media during the pandemic: seeking support and awareness about COVID-19 through social media platforms. J Public Aff 2020 Sep 15:e02417. [doi: 10.1002/pa.2417]
18. Kaya T. The changes in the effects of social media use of Cypriots due to COVID-19 pandemic. Technol Soc 2020 Nov;63:101380 [FREE Full text] [doi: 10.1016/j.techsoc.2020.101380] [Medline: 33518848]
19. Pahayahay A, Khalili-Mahani N. What media helps, what media hurts: a mixed methods survey study of coping with COVID-19 using the media repertoire framework and the appraisal theory of stress. J Med Internet Res 2020 Aug 06;22(8):e20186 [FREE Full text] [doi: 10.2196/20186] [Medline: 32701459]

XSL•FO
RenderX

20. Carlsen HB, Toubøl J, Brincker B. On solidarity and volunteering during the COVID-19 crisis in Denmark: the impact of social networks and social media groups on the distribution of support. Eur Soc 2020 Sep 14;23(sup1):S122-S140. [doi: 10.1080/14616696.2020.1818270]
21. Clark DL, Raphael JL, McGuire AL. HEADS: social media screening in adolescent primary care. Pediatrics 2018 Jun;141(6):e20173655. [doi: 10.1542/peds.2017-3655] [Medline: 29716979]
22. Kross E, Verduyn P, Sheppes G, Costello CK, Jonides J, Ybarra O. Social media and well-being: pitfalls, progress, and next steps. Trends Cogn Sci 2021 Jan;25(1):55-66 [FREE Full text] [doi: 10.1016/j.tics.2020.10.005] [Medline: 33187873]
23. Thomas DS, Jang S, Scandlyn J. The CHASMS conceptual model of cascading disasters and social vulnerability: the COVID-19 case example. Int J Disaster Risk Reduct 2020 Dec;51:101828 [FREE Full text] [doi: 10.1016/j.ijdrr.2020.101828] [Medline: 32895627]
24. Kira IA, Shuwiekh HA, Ashby JS, Elwakeel SA, Alhuwailah A, Sous MSF, et al. The impact of COVID-19 traumatic stressors on mental health: is COVID-19 a new trauma type. Int J Ment Health Addict 2021 Jul 06:1-20 [FREE Full text] [doi: 10.1007/s11469-021-00577-0] [Medline: 34248442]
25. Thomas É, Saumier D, Brunet A. Peritraumatic distress and the course of posttraumatic stress disorder symptoms: a meta-analysis. Can J Psychiatry 2012 Feb;57(2):122-129. [doi: 10.1177/070674371205700209] [Medline: 22340152]
26. Holman EA, Garfin DR, Silver RC. Media's role in broadcasting acute stress following the Boston Marathon bombings. Proc Natl Acad Sci U S A 2014 Jan 07;111(1):93-98 [FREE Full text] [doi: 10.1073/pnas.1316265110] [Medline: 24324161]
27. Verduyn P, Ybarra O, Résibois M, Jonides J, Kross E. Do social network sites enhance or undermine subjective well-being? A critical review. Soc Issues Policy Rev 2017 Jan 13;11(1):274-302. [doi: 10.1111/sipr.12033]
28. Yue Z, Zhang R, Xiao J. Passive social media use and psychological well-being during the COVID-19 pandemic: the role of social comparison and emotion regulation. Comput Human Behav 2022 Feb;127:107050 [FREE Full text] [doi: 10.1016/j.chb.2021.107050] [Medline: 34646057]
29. Valkenburg PM, Peter J. The differential susceptibility to media effects model. J Commun 2013 Mar 07;63(2):221-243. [doi: 10.1111/jcom.12024]
30. Silver RC, Holman EA, Andersen JP, Poulin M, McIntosh DN, Gil-Rivas V. Mental- and physical-health effects of acute exposure to media images of the September 11, 2001, attacks and the Iraq War. Psychol Sci 2013 Sep;24(9):1623-1634. [doi: 10.1177/0956797612460406] [Medline: 23907546]
31. Zhao N, Zhou G. Social media use and mental health during the COVID-19 pandemic: moderator role of disaster stressor and mediator role of negative affect. Appl Psychol Health Well Being 2020 Dec;12(4):1019-1038 [FREE Full text] [doi: 10.1111/aphw.12226] [Medline: 32945123]
32. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. PLoS One 2020;15(4):e0231924 [FREE Full text] [doi: 10.1371/journal.pone.0231924] [Medline: 32298385]
33. Thoresen S, Tambs K, Hussain A, Heir T, Johansen VA, Bisson JI. Brief measure of posttraumatic stress reactions: impact of Event Scale-6. Soc Psychiatry Psychiatr Epidemiol 2010 Mar;45(3):405-412. [doi: 10.1007/s00127-009-0073-x] [Medline: 19479171]
34. Brunet A, Weiss DS, Metzler TJ, Best SR, Neylan TC, Rogers C, et al. The Peritraumatic Distress Inventory: a proposed measure of PTSD criterion A2. Am J Psychiatry 2001 Sep;158(9):1480-1485. [doi: 10.1176/appi.ajp.158.9.1480] [Medline: 11532735]
35. Guardia D, Brunet A, Duhamel A, Ducrocq F, Demarty A, Vaiva G. Prediction of trauma-related disorders: a proposed cutoff score for the peritraumatic distress inventory. Prim Care Companion CNS Disord 2013;15(1):PCC.12l01406 [FREE Full text] [doi: 10.4088/PCC.12l01406] [Medline: 23724345]
36. Main A, Zhou Q, Ma Y, Luecken LJ, Liu X. Relations of SARS-related stressors and coping to Chinese college students' psychological adjustment during the 2003 Beijing SARS epidemic. J Couns Psychol 2011 Jul;58(3):410-423. [doi: 10.1037/a0023632] [Medline: 21574694]
37. Goodboy AK, Kline RB. Statistical and practical concerns with published communication research featuring structural equation modeling. Commun Res Rep 2016 Dec 16;34(1):68-77. [doi: 10.1080/08824096.2016.1214121]
38. Little RJA. A test of missing completely at random for multivariate data with missing values. J Am Stat Assoc 1988 Dec;83(404):1198-1202. [doi: 10.1080/01621459.1988.10478722]
39. Huang J, Keung JW, Sarro F, Li Y, Yu Y, Chan W, et al. Cross-validation based K nearest neighbor imputation for software quality datasets: an empirical study. J Syst Software 2017 Oct;132:226-252. [doi: 10.1016/j.jss.2017.07.012]
40. Byrne BM. Structural Equation Modeling With AMOS: Basic Concepts, Applications, and Programming, 3rd Ed. New York: Routledge; 2016.
41. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Structural Equation Modeling Multidisciplinary J 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
42. Kline RB. Principles and Practice of Structural Equation Modeling, 4th ed. New York: Guilford Press; 2016.
43. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol 1986 Dec;51(6):1173-1182. [doi: 10.1037//0022-3514.51.6.1173] [Medline: 3806354]

XSL•FO
RenderX

44. Preacher KJ, Hayes AF. SPSS and SAS procedures for estimating indirect effects in simple mediation models. Behav Res Methods Instrum Comput 2004 Nov;36(4):717-731. [doi: 10.3758/bf03206553] [Medline: 15641418]
45. Abdallah W, Goergen M, O'Sullivan N. Endogeneity: how failure to correct for it can cause wrong inferences and some remedies. Br J Manage 2015 Jul 16;26(4):791-804. [doi: 10.1111/1467-8551.12113]
46. Bunnell BE, Davidson TM, Ruggiero KJ. The Peritraumatic Distress Inventory: factor structure and predictive validity in traumatically injured patients admitted through a Level I trauma center. J Anxiety Disord 2018 Apr;55:8-13 [FREE Full text] [doi: 10.1016/j.janxdis.2018.03.002] [Medline: 29549879]
47. Drouin M, McDaniel BT, Pater J, Toscos T. How parents and their children used social media and technology at the beginning of the COVID-19 pandemic and associations with anxiety. Cyberpsychol Behav Soc Netw 2020 Nov;23(11):727-736. [doi: 10.1089/cyber.2020.0284] [Medline: 32726144]
48. Hossain MT, Ahammed B, Chanda SK, Jahan N, Ela MZ, Islam MN. Social and electronic media exposure and generalized anxiety disorder among people during COVID-19 outbreak in Bangladesh: A preliminary observation. PLoS One 2020;15(9):e0238974 [FREE Full text] [doi: 10.1371/journal.pone.0238974] [Medline: 32916691]
49. Bendau A, Petzold MB, Pyrkosch L, Mascarell Maricic L, Betzler F, Rogoll J, et al. Associations between COVID-19 related media consumption and symptoms of anxiety, depression and COVID-19 related fear in the general population in Germany. Eur Arch Psychiatry Clin Neurosci 2021 Mar;271(2):283-291 [FREE Full text] [doi: 10.1007/s00406-020-01171-6] [Medline: 32691135]
50. Zhong B, Huang Y, Liu Q. Mental health toll from the coronavirus: social media usage reveals Wuhan residents' depression and secondary trauma in the COVID-19 outbreak. Comput Human Behav 2021 Jan;114:106524 [FREE Full text] [doi: 10.1016/j.chb.2020.106524] [Medline: 32836728]
51. Lisitsa E, Benjamin KS, Chun SK, Skalisky J, Hammond LE, Mezulis AH. Loneliness among young adults during COVID-19 pandemic: the mediational roles of social media use and social support seeking. J Soc Clin Psychol 2020 Oct;39(8):708-726. [doi: 10.1521/jscp.2020.39.8.708]
52. Lin S, Liu D, Niu G, Longobardi C. Active social network sites use and loneliness: the mediating role of social support and self-esteem. Curr Psychol 2020 Feb 13;41(3):1279-1286. [doi: 10.1007/s12144-020-00658-8]
53. Cramer KM. Psychological antecedents to help-seeking behavior: a reanalysis using path modeling structures. J Counseling Psychol 1999 Jul;46(3):381-387. [doi: 10.1037/0022-0167.46.3.381]
54. Monson CM, Gradus JL, Young-Xu Y, Schnurr PP, Price JL, Schumm JA. Change in posttraumatic stress disorder symptoms: do clinicians and patients agree? Psychol Assess 2008 Jun;20(2):131-138. [doi: 10.1037/1040-3590.20.2.131] [Medline: 18557690]

Abbreviations

CFI: comparative index fit
DSMM: differential susceptibility to media effects model
IES-6: Impact of Event Scale–Revised
PDI: Peritraumatic Distress Inventory
PTSD: posttraumatic stress disorder
RMSEA: root mean square error of approximation
SRMR: standardized root mean square residual
TSR: trauma- and stressor-related

Edited by C Basch; submitted 18.08.21; peer-reviewed by Z Jin, Z Yue; comments to author 08.12.21; revised version received 24.03.22; accepted 18.04.22; published 07.06.22.

Please cite as:
Rivest-Beauregard M, Fortin J, Guo C, Cipolletta S, Sapkota RP, Lonergan M, Brunet A
Media Use During the COVID-19 Pandemic: Cross-sectional Study
J Med Internet Res 2022;24(6):e33011
URL: https://www.jmir.org/2022/6/e33011
doi:10.2196/33011
PMID:3553703

©Marjolaine Rivest-Beauregard, Justine Fortin, Connie Guo, Sabrina Cipolletta, Ram P Sapkota, Michelle Lonergan, Alain Brunet. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.06.2022. 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

XSL•FO
RenderX

Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Nursing Workflow Change in a COVID-19 Inpatient Unit Following the Deployment of Inpatient Telehealth: Observational Study Using a Real-Time Locating System

Stacie Vilendrer[1], MS, MBA, MD; Mary E Lough[1,2], PhD; Donn W Garvert[1], MS; Monique H Lambert[2], PhD; Jonathan Hsijing Lu[3], MSE; Birju Patel[3], MD, MPH; Nigam H Shah[3], MBBS, PhD; Michelle Y Williams[1,2], RN, PhD; Samantha M R Kling[1], PhD

[1]Evaluation Sciences Unit, Division of Primary Care and Population Health, Department of Medicine, Stanford University School of Medicine, Stanford, CA, United States

[2]Office of Research Patient Care Services, Stanford Health Care, Palo Alto, CA, United States

[3]Center for Biomedical Informatics Research, Stanford University School of Medicine, Stanford, CA, United States

Corresponding Author:
Stacie Vilendrer, MS, MBA, MD
Evaluation Sciences Unit, Division of Primary Care and Population Health
Department of Medicine
Stanford University School of Medicine
1265 Welch Road
Stanford, CA, 94305
United States
Phone: 1 587 206 4043
Email: staciev@stanford.edu

Abstract

Background: The COVID-19 pandemic prompted widespread implementation of telehealth, including in the inpatient setting, with the goals to reduce potential pathogen exposure events and personal protective equipment (PPE) utilization. Nursing workflow adaptations in these novel environments are of particular interest given the association between nursing time at the bedside and patient safety. Understanding the frequency and duration of nurse-patient encounters following the introduction of a novel telehealth platform in the context of COVID-19 may therefore provide insight into downstream impacts on patient safety, pathogen exposure, and PPE utilization.

Objective: The aim of this study was to evaluate changes in nursing workflow relative to prepandemic levels using a real-time locating system (RTLS) following the deployment of inpatient telehealth on a COVID-19 unit.

Methods: In March 2020, telehealth was installed in patient rooms in a COVID-19 unit and on movable carts in 3 comparison units. The existing RTLS captured nurse movement during 1 pre- and 5 postpandemic stages (January-December 2020). Change in direct nurse-patient encounters, time spent in patient rooms per encounter, and total time spent with patients per shift relative to baseline were calculated. Generalized linear models assessed difference-in-differences in outcomes between COVID-19 and comparison units. Telehealth adoption was captured and reported at the unit level.

Results: Change in frequency of encounters and time spent per encounter from baseline differed between the COVID-19 and comparison units at all stages of the pandemic (all $P<.001$). Frequency of encounters decreased (difference-in-differences range –6.6 to –14.1 encounters) and duration of encounters increased (difference-in-differences range 1.8 to 6.2 minutes) from baseline to a greater extent in the COVID-19 units relative to the comparison units. At most stages of the pandemic, the change in total time nurses spent in patient rooms per patient per shift from baseline did not differ between the COVID-19 and comparison units (all $P>.17$). The primary COVID-19 unit quickly adopted telehealth technology during the observation period, initiating 15,088 encounters that averaged 6.6 minutes (SD 13.6) each.

Conclusions: RTLS movement data suggest that total nursing time at the bedside remained unchanged following the deployment of inpatient telehealth in a COVID-19 unit. Compared to other units with shared mobile telehealth units, the frequency of nurse-patient in-person encounters decreased and the duration lengthened on a COVID-19 unit with in-room telehealth availability, indicating “batched” redistribution of work to maintain total time at bedside relative to prepandemic periods. The simultaneous

XSL•FO
RenderX

adoption of telehealth suggests that virtual care was a complement to, rather than a replacement for, in-person care. However, study limitations preclude our ability to draw a causal link between nursing workflow change and telehealth adoption. Thus, further evaluation is needed to determine potential downstream implications on disease transmission, PPE utilization, and patient safety.

(J Med Internet Res 2022;24(6):e36882) doi:10.2196/36882

KEYWORDS

telemedicine; telehealth; informatics; real-time locating system; COVID-19; pandemic; nursing; patient safety; PPE; virtual care; nurses; patient outcomes; pathogen exposure; health risk; health care staff; health care professional

Introduction

The COVID-19 pandemic prompted the widespread implementation of telehealth throughout the health care sector to protect patients and health care workers by reducing the risk of infection [1,2]. In the inpatient setting, telehealth was previously used to connect hospitalized patients in rural settings with remote specialists [3], but has recently expanded to facilitate digital communication between patients and on-site clinicians [4-6]. The impact of inpatient telehealth on infection reduction [7], clinical workflows [8,9], patient safety [10], and personal protective equipment (PPE) utilization [10,11] is still under investigation. In the context of a pandemic, telehealth's impact on infection control and clinical care is of particular interest, although existing evaluations have relied on PPE inventory data and patient and provider satisfaction surveys [10,11]. However, the expanding use of real-time locator systems (RTLSs), as an objective source of location and time data, in inpatient settings offers a unique opportunity to understand how clinical workflows adapt to these novel circumstances.

An RTLS captures the time spent in specific locations, providing a direct measure of workflow for health care professionals as well as a proxy for outcomes associated with staff movement, such the use of PPE upon each entrance into a room [12-17]. In the context of infection control, RTLS data have been used to identify possible transmission: one study demonstrated higher sensitivity and specificity using RTLS-based contact tracing than an audit log capturing electronic health record (EHR) logins at diverse workstations [7]. Availability of an RTLS provides an opportunity to evaluate the impact of new technologies or clinical circumstances (eg, infectious outbreak) on clinical workflows. A recent study analyzing RTLS data in the emergency setting provided an analytical framework to understand possible clinician workflow adaptations, although no change was ultimately detected in this setting [9].

The impact of telehealth on nursing workflows is of particular interest, as nurses spend approximately 6-fold more time at the bedsides of hospitalized patients than attending physicians [18]. More time in direct nurse-patient encounters has been associated with improved patient safety [19] and satisfaction [17]. The manner in which nurses structure in-person encounters with patients depends on local hospital guidelines [19], but can also vary between individuals, according to the level of training, and time of day [20]. Bedside encounters for inpatients under isolation precautions (such as is required during COVID-19 treatment) appear to be reduced relative to other inpatients [21]. This reduction may contribute to isolated patients receiving substandard care [22-24]. Understanding direct nurse-patient care following the introduction of a novel telehealth platform in the context of COVID-19 may provide insight into the downstream impacts on patient safety, pathogen exposure, and PPE utilization.

Thus, we aimed to evaluate changes in nursing workflow relative to prepandemic levels using an RTLS following the deployment of inpatient telehealth on a COVID-19 unit. Given isolation precautions and ready availability of telehealth equipment on the COVID-19 unit, we hypothesized a reduction in the frequency and duration of in-person nurse-patient encounters in the COVID-19 unit relative to prepandemic comparator units.

Methods

Design

Telehealth was implemented throughout an acute care academic hospital in response to the COVID-19 pandemic in March of 2020. The setup of telehealth differed between the hospital's primary COVID-19 unit and comparison units. Our primary aim was to explore changes in nursing workflows in these novel circumstances through a retrospective, observational evaluation using RTLS data to capture the frequency of direct nurse-patient encounters at the bedside, time nurses spent in patient rooms per encounter, and total time nurses spent with each patient in the patient room per shift. For each outcome, the change from the prepandemic level was calculated, and difference-in-differences analyses were used to determine if changes in nurse movement differed between the COVID-19 unit and comparison units. The simultaneous adoption of telehealth in terms of video call frequency and duration was captured and is reported at the unit level.

Setting

The academic acute care hospital is in a major metropolitan area in the western United States and serves a diverse population. The four inpatient units in this evaluation were identical in size and layout with 22-bed single-room capacity. Prepandemic, these four units focused on inpatient general medicine populations. Other units serving primarily surgical, oncological, and intensive care patients were excluded. At the beginning of the pandemic, one of the units became the primary COVID-19 unit for the hospital and was therefore compared to the other three units. Hospital administration reported that standard registered nurse-to-patient ratio on the four units ranged from 1:4 to 1:3, varying with disease acuity based on state law

XSL•FO
RenderX

[25]. The registered nurse-to-patient ratio on the COVID-19 unit shifted from 1:4 to 1:3 by April 2020.

Timeline

The first patient tested positive for SARS-CoV-2 in the hospital's emergency department on March 2, 2020, and local stay-at-home orders were announced on March 16, 2020 [26]. For the purposes of this evaluation, data from all data sources were collected from January 1, 2020, through December 27, 2020. These data captured 6 stages of the pandemic that were defined based on local case rates [27] (for criteria and specific dates, see section A1 of Multimedia Appendix 1): (1) prepandemic, (2) telehealth rollout, (3) nonsurge #1, (4) surge #1, (5) nonsurge #2, and (6) surge #2.

Telehealth Deployment

In response to the pandemic, telehealth was rapidly implemented throughout inpatient settings in mid-March 2020, as previously reported [4] and described in section A2 of Multimedia Appendix 1. In the COVID-19 inpatient unit, telehealth hardware with a video tablet was permanently installed in each patient room. However, in non-COVID units, shared telehealth video tablets were available only on mobile carts, which could be transported into a patient room as needed. A member of the clinical team was required to roll the cart into the patient room prior to each use, and these units were sometimes unavailable for patients if they were already in use by another patient.

All clinical team members, including hospitalist and specialist physicians, nurses, respiratory therapists, trainees, and clinical researchers, received instructions on and were encouraged to incorporate telehealth into their clinical or research activities. Patients received incoming calls passively given the default automatic turn-on feature of the video tablet, but received no other instruction or guidance on its use [28].

Data Sources and Processing

RTLS Nurse Movement Data

The primary data source for this evaluation was extracted from the existing RTLS platform (Midmark, Dayton, OH). The RTLS captured movement of nurses into and out of patient rooms on the selected units. These data were used to calculate the following 3 outcomes related to nurses' movement and direct patient care: (1) number of nurse-patient direct encounters within patient rooms, (2) time nurses spent in patient room per encounter, and (3) total time nurses spent in the patient room with each patient per shift.

The key components of the RTLS are infrared and radiofrequency sensors installed in each room and staff badges worn alongside their name badge [18,29]. Line of sight between the room sensor and staff member's badge triggered the system to record an event such as entry of a nurse into a patient room. The badge emitted a ping every 1 to 3 seconds to indicate presence in the room. Only events longer than 5 seconds were recorded in the system. At installation, sensitivity settings were optimized based on the geometric configuration and construction materials, but were not reassessed for this evaluation [9,18].

Since the direct line of sight between the room sensor and nurse badge could be briefly interrupted (eg, by turning away from the sensor), a single nurse-patient encounter could appear as several short, consecutive events. Thus, to identify unique direct nurse-patient encounters and calculate the time nurses spent in a patient room, multiple successive RTLS events that occurred within 30 seconds of another and were associated with an individual nurse in a single patient room were collapsed into a single direct nurse-patient encounter. The duration of individual encounters was summed to calculate the total time nurses spent with each patient per shift using the first and last timestamp of each encounter associated with a unique nurse in a specific patient room. Since nursing needs and thus workflows may differ between shifts [20], movement from 7:00 AM to 6:59 PM (morning shift) is presented separately from movement from 7:00 PM to 6:59 AM (night shift).

Staff RTLS badges were linked within the system to an employee identifier and role. Nurses categorized as "nurse" or "float nurse" within the RTLS were included in the analysis. Hospital administration and managers encourage all nurses to wear the badges to utilize the system's beneficial features, including automatic silencing of patient room alarms when a nurse enters a room and a discreet button to call security. Compliance with badge wearing is near universal among nursing staff compared to other members of the clinical team (eg, physicians) [9]. However, the evaluation team could not and did not confirm that all nurses were compliant during the observation period. The process to access these data is briefly described in section A3 of Multimedia Appendix 1.

EHR Patient Data

Data were extracted from the EHR (Epic, Verona, WI, USA) to determine the presence of a patient in each hospital room in the unit at midnight, and the results of the patient's most recent COVID-19 test result (positive or negative) within the prior 14 days or from hospital admission. Besides this information, no other patient-level data such as identifiers or clinical characteristics were obtained. The RTLS and EHR data were then merged by patient room and time to identify direct encounters between nurses and patients.

Telehealth Utilization Data

Data were extracted from Zoom video conferencing software and included the unit associated with the host (originating hardware) user ID, call start time, call end time, and number of participants. Only calls lasting between 30 seconds and 2 hours with 2 or more participants were included in analysis. Patients were not instructed or encouraged to initiate calls to clinical staff or family members themselves during the observation period, although anecdotal reports suggest that nurses occasionally set up calls between patients and their families. This data platform did not link a telehealth encounter to individual physicians, staff members (nurses), or patients. Consequently, telehealth use described all possible use cases at the unit level.

Data Analysis

Descriptive statistics were generated to describe patient census by COVID-19 status and telehealth utilization in the COVID-19

XSL•FO
RenderX

and comparison units prepandemic during the 5 stages of the pandemic. Telehealth utilization is expressed as the number of telehealth calls per patient (using the midnight patient census) per unit for all case uses such as clinical encounters, research activities, and patient-family connections. Since telehealth events could not be linked to individual users (patients or health care workers), no additional analysis was performed with these data.

To investigate differences in the three primary outcomes based on RTLS data between the COVID-19 and comparison units, a difference-in-differences approach was applied. Change in each outcome was calculated for each of the 5 stages of the pandemic, as defined by local case rates, relative to prepandemic levels. To determine if the change from the prepandemic stage differed between the COVID-19 and comparison units, difference-in-differences was determined using a generalized linear model in SAS (version 9.4, SAS Institute Inc) for each of the three outcomes.

Means, standard deviations, and ranges are reported where appropriate. For all models, $P<.05$ was considered statistically significant. When multiple comparisons were made, P values were adjusted using an adaptive, two-stage linear setup procedure to control the false discovery rate [30].

Ethics Considerations

Data were obtained from multiple sources, including RTLS movement data at the individual nurse level, EHR data at the patient and thus room level, and telehealth log data at the unit level. All data were deidentified and data points that did not occur in patient rooms were removed before being sent to the analytics team to protect the anonymity of the workforce. The evaluation was exempt per the Stanford University's institutional review board (protocol 55927).

Results

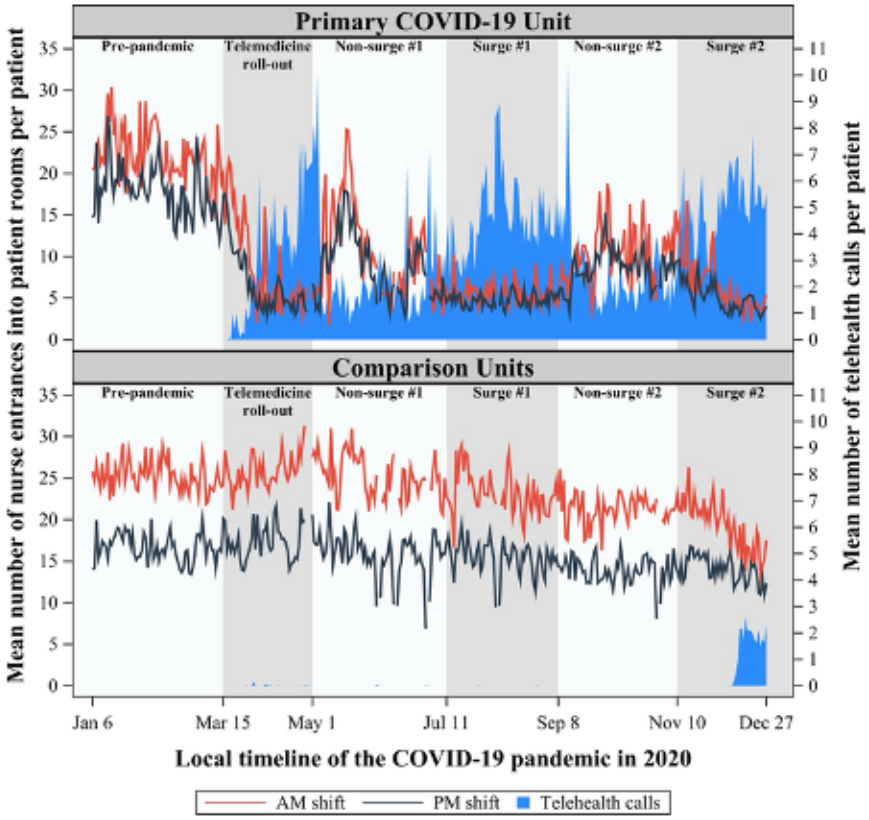
Overview

The COVID-19 unit was the main care location for hospitalized patients diagnosed with COVID-19, but it was also the care location for some patients without COVID-19 (see Table S1 in Multimedia Appendix 2). In contrast, the comparison units cared for very few patients diagnosed with COVID-19 prior to the final evaluation stage, surge #2, when the number of COVID-19 patients increased throughout the hospital.

The comparison units demonstrated little adoption of telehealth throughout most of the pandemic, whereas the primary COVID-19 unit quickly adopted this technology (Figures 1-3), initiating 15,088 inpatient telehealth video encounters, resulting in a cumulative duration of 1660 hours throughout the observation period. On average, encounters were 6.6 (SD 13.6) minutes per telehealth call.

Results are presented in order of the three primary outcomes, including the change in the frequency nurses entered patient rooms, time nurses spent in patient rooms per entry, and total time nurses spent in each patient room per shift.

Figure 1. Daily mean number of times nurses entered patient rooms by shift on a COVID-19 unit and three comparison units during a telehealth implementation in the context of the SARS-CoV-2 pandemic.



XSL•FO
RenderX

Figure 2. Median time (minutes) nurses spent in patient rooms per encounter by shift on a primary COVID-19 unit and three comparison units during a telehealth implementation in the context of the SARS-CoV-2 pandemic.

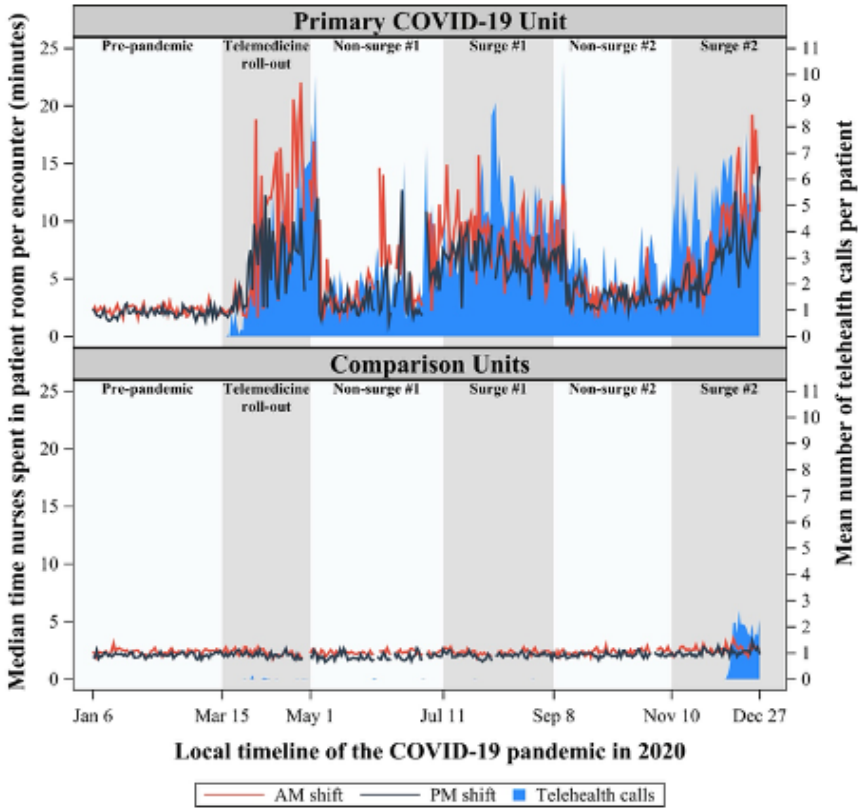
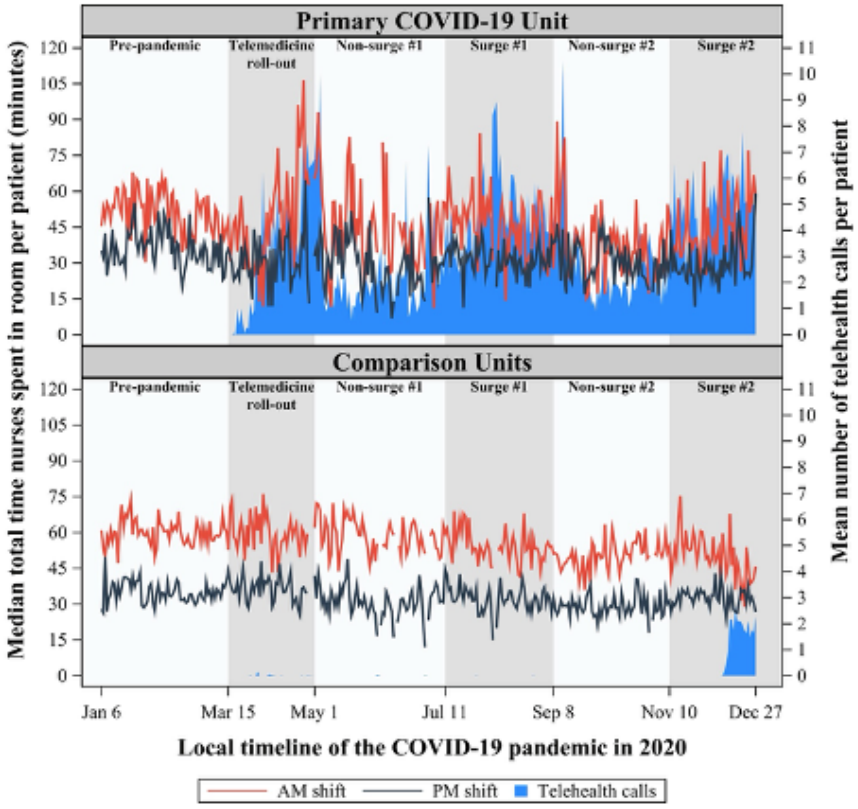


Figure 3. Median total time (minutes) nurses spent in patient rooms per patient per shift on a primary COVID-19 unit and three comparison units during a telehealth implementation in the context of the SARS-CoV-2 pandemic.



XSL•FO
RenderX

Direct Nurse-Patient Encounters

Across both the COVID-19 and comparison units, 876,177 unique direct nurse-patient encounters were identified, including 226,326 encounters prepandemic and 649,791 during the pandemic. Of these, 18,011 encounters were linked to a patient who was positive for COVID-19.

The daily mean number of nurse entries in patient rooms by shift (morning and night shifts) for the primary COVID-19 unit and comparison units is shown in Figure 1. In the prepandemic stage, nurses in the primary COVID-19 unit entered patient rooms less frequently than nurses in the comparison units for the morning and more frequently for the night shift (Table 1). Because the COVID-19 versus comparator units differed in the number of prepandemic nurse-patient encounters, a difference-in-differences analysis was applied to determine if change in this outcome relative to baseline differed between the COVID-19 and comparison units.

For each stage of the pandemic, change in number of times nurses entered patient rooms relative to that prepandemic are shown in Table 1 for the COVID-19 unit and comparison units. In the COVID-19 unit, the decrease in encounters relative to prepandemic levels ranged from –9.2 to –16.8 nurse-patient encounters per shift, whereas in the comparison units, these fluctuations were less pronounced (range –5.9 to +0.8 nurse-patient encounters per shift). At all stages of the pandemic, for both morning and night shifts, the number of times nurses entered patient rooms decreased from prepandemic levels to a greater extent for nurses on the COVID-19 unit than for nurses in the comparison units (all P<.001). Nurses in the COVID-19 unit entered patient rooms less frequently during each surge period when compared to the nonsurge period just before; this pattern was seen for both the morning and night shifts (all P<.001; estimates range from 3.4 to 4.8 additional entries).

Table 1. Number of direct nurse-patient encounters per shift, difference in number of encounters from prepandemic stage, and difference-in-differences between a primary COVID-19 unit and three comparison units in the context of the SARS-CoV-2 pandemic.

Stage	Primary COVID-19 unit		Comparison units		Difference in differences	P value[a]
	Number of direct nurse-patient encounters, mean (SD)	Difference from prepandemic	Number of direct nurse-patient encounters, mean (SD)	Difference from prepandemic		
Prepandemic baseline						
AM shift	22.4 (3.2)	N/A[b]	25.3 (1.8)	N/A	N/A	N/A
PM shift	17.8 (3.1)	N/A	16.6 (1.7)	N/A	N/A	N/A
Telehealth rollout						
AM shift	8.9 (5.7)	–13.5	25.8 (2.0)	0.5	–14.1	<.001
PM shift	6.8 (3.9)	–11.0	17.3 (2.1)	0.8	–11.8	<.001
Nonsurge #1						
AM shift	10.5 (5.5)	–11.9	25.3 (2.5)	0.0	–11.9	<.001
PM shift	8.6 (3.8)	–9.2	15.9 (2.6)	–0.7	–8.5	<.001
Surge #1						
AM shift	5.6 (1.4)	–16.8	23.2 (2.6)	–2.1	–14.7	<.001
PM shift	4.7 (0.8)	–13.1	15.5 (2.0)	–1.1	–12.0	<.001
Nonsurge #2						
AM shift	10.2 (3.4)	–12.2	21.3 (1.8)	–4.0	–8.2	<.001
PM shift	8.5 (2.2)	–9.3	13.9 (1.6)	–2.7	–6.6	<.001
Surge #2						
AM shift	6.6 (3.1)	–15.8	19.4 (3.0)	–5.9	–9.9	<.001
PM shift	5.1 (1.8)	–12.7	14.0 (1.6)	–2.6	–10.1	<.001

[a]Difference-in-differences was statistically tested using a generalized linear model.

[b]N/A: not applicable.

Duration of Nurse-Patient Encounters

The daily median time nurses spent in patient rooms per shift (morning and night) for the primary COVID-19 unit and comparison units is shown in Figure 2. The downward trend in encounters in patient rooms in the COVID-19 unit (Figure 1, Table 1) corresponded with an upward shift in the amount of time nurses spent in the patient rooms per entry (Figure 2), and this pattern remained consistent throughout the pandemic.

Time nurses spent in patient rooms per entry did not differ between the COVID-19 unit and comparison units during the prepandemic stages for both the morning (P=.79) and night

XSL•FO
RenderX

(*P*=.81) shifts (Table 2). For each stage of the pandemic, change in time nurses spent in patient rooms per encounter relative to prepandemic is shown in Table 2 for the primary COVID-19 unit and comparison units. In the COVID-19 unit, the duration of encounters increased in all phases of the pandemic, although at different rates, with mean increases of 1.8 to 6.2 minutes compared to prepandemic levels (Table 2). In the comparison units, these fluctuations were less pronounced, with no change during several phases and a range of –0.2 to 0.2 minutes per encounter (Table 2). At all stages of the pandemic, for both the morning and night shifts, the duration of nurse-patient encounters increased from prepandemic levels to a greater extent for nurses in the COVID-19 unit than for nurses in the comparison units (all *P*<.001). Nurses in the COVID-19 unit spent more time in patient rooms per entry during each surge period when compared to that in the nonsurge period just before; this pattern was seen for both the morning and night shifts (all *P*<.001; estimates range from 2.6 to 4.1 fewer minutes per entry).

Table 2. Duration (minutes) each nurse spent in patient rooms per encounter by shift, difference in duration of encounters, and difference-in-differences between the primary COVID-19 unit and three comparison units during a telehealth implementation in the context of the SARS-CoV-2 pandemic.

Period	Primary COVID-19 unit		Comparison units		Differences in differences	*P* value[a]
	Duration (minutes) nurse spent in patient room per encounter, mean[b] (SD)	Difference from prepandemic	Duration (minutes) nurse spent in patient room per encounter, mean[b] (SD)	Difference from prepandemic		
Prepandemic baseline						
AM shift	2.3 (0.3)	N/A[c]	2.4 (0.2)	N/A	N/A	N/A
PM shift	2.0 (0.4)	N/A	2.1 (0.2)	N/A	N/A	N/A
Telehealth rollout						
AM shift	8.5 (6.3)	6.2	2.3 (0.3)	–0.1	6.2	<.001
PM shift	5.7 (3.0)	3.7	2.1 (0.2)	0.0	3.7	<.001
Nonsurge #1						
AM shift	5.5 (3.6)	3.2	2.3 (0.2)	–0.1	3.3	<.001
PM shift	4.0 (2.5)	2.0	1.9 (0.3)	–0.2	2.1	<.001
Surge #1						
AM shift	8.5 (2.6)	6.2	2.3 (0.2)	0.0	6.2	<.001
PM shift	6.8 (1.5)	4.8	2.1 (0.2)	0.0	4.8	<.001
Nonsurge #2						
AM shift	4.1 (2.2)	1.8	2.4 (0.2)	0.0	1.8	<.001
PM shift	3.8 (1.5)	1.8	2.1 (0.2)	0.0	1.8	<.001
Surge #2						
AM shift	8.2 (4.2)	5.9	2.6 (0.3)	0.2	5.7	<.001
PM shift	6.4 (2.7)	4.4	2.3 (0.3)	0.2	4.2	<.001

[a]Difference in differences was statistically tested using a generalized linear model.

[b]To determine the duration a nurse spent in the patient rooms per encounter, the median value per shift was calculated. In this table, the mean represents the mean of these median values.

[c]N/A: not applicable.

Total Time of Direct Nurse-Patient Care

The daily median total time nurses spent in patient rooms per patient per shift (morning and night) for the primary COVID-19 unit and comparison units is shown in Figure 3. During the prepandemic stage, nurses in the primary COVID-19 unit spent less total time per patient per shift than nurses in the comparison units during the morning shift (*P*<.001) but not the night shift (*P*=.57), as shown in Table 3.

In each stage of the pandemic, the change in the total time nurses were in a patient room per patient per shift relative to baseline is shown in Table 3 for the primary COVID-19 unit and comparison units. For the COVID-19 unit, the mean change in the total time nurses spent in patient rooms per patient per shift ranged from –2.8 to –13.0 minutes relative to the prepandemic times. In comparison units, the mean change relative to prepandemic was less pronounced and ranged from –9.9 to +1.6 minutes (Table 3). At most stages of the pandemic, change in total time in the patient room from prepandemic did not differ

XSL•FO
RenderX

between the primary COVID-19 unit and comparison units (all $P>.17$), with the exception of the telehealth rollout phase.

Table 3. Total time (minutes) all nurses spent in a patient room by shift, difference in total time relative to baseline, and difference-in-differences between the primary COVID-19 unit and three comparison units during a telehealth implementation in the context of the SARS-CoV-2 pandemic.

Period	Primary COVID-19 unit		Comparison units		Differences in differences	*P* value[a]
	Total time (minutes) nurses spent in patient room per patient per shift, mean[b] (SD)	Difference from prepandemic	Total time (minutes) nurse spent in patient room per patient per shift, mean[b] (SD)	Difference from prepandemic		
Prepandemic baseline						
AM shift	51.0 (9.1)	N/A[c]	60.0 (5.1)	N/A	N/A	N/A
PM shift	35.8 (7.6)	N/A	34.9 (4.4)	N/A	N/A	N/A
Telehealth rollout						
AM shift	48.2 (20.3)	–2.8	59.4 (7.3)	–0.6	–2.2	.48
PM shift	30.7 (10.1)	–5.1	36.5 (5.0)	1.6	–6.8	.04
Nonsurge #1						
AM shift	44.8 (17.1)	–6.2	58.7 (6.9)	–1.3	-4.9	.17
PM shift	28.9 (10.6)	–6.9	30.9 (6.3)	–4.0	–3.0	.26
Surge #1						
AM shift	46.2 (12.8)	–4.8	54.4 (6.5)	–5.7	0.8	.72
PM shift	31.4 (6.2)	–4.4	32.2 (5.3)	–2.7	–1.6	.50
Nonsurge #2						
AM shift	38.0 (14.4)	–13.0	50.4 (5.6)	–9.6	–3.4	.24
PM shift	30.4 (8.1)	–5.4	29.1 (4.3)	–5.8	0.4	.87
Surge #2						
AM shift	44.9 (13.6)	–6.1	50.1 (9.2)	–9.9	3.8	.24
PM shift	29.8 (8.8)	–6.0	32.5 (3.9)	–2.4	–3.7	.26

[a]Difference-in-differences was statistically tested using a generalized linear model.

[b]To determine the total time nurses spent in patient rooms, the median value per shift was calculated. In this table, the mean represents the mean of these median values.

[c]N/A: not applicable.

Discussion

Principal Findings

Evaluating changes in the frequency and duration of direct nurse-patient encounters using an RTLS following an inpatient telehealth deployment during the COVID-19 pandemic was feasible and provided novel insights into nursing workflow redistribution in this setting. Relative to the prepandemic stage, nurses in a COVID-19 unit with in-room ready access to telehealth decreased the frequency of entries into patient rooms to a greater extent than that for nurses in other units with shared mobile telehealth units. Counter to our hypothesis, the average in-person encounter length increased proportionally, such that the total in-person time nurses spent with patients on the COVID-19 unit did not significantly differ from that in prepandemic comparator units. The simultaneous adoption of telehealth, presented at the unit level, suggests it was used as a complement to, rather than a replacement for, in-person care.

To put the above findings into context, the average decrease in encounters weighted by time period in the COVID-19 unit relative to other units was 11.25 and 9.13 encounters per patient per morning and night shift, respectively. Assuming full capacity on the 22-bed COVID-19 unit and PPE use for a quarter of such encounters (as isolation precautions were not required for every patient on the unit), these data suggest workflow adaptations saved approximately 785 PPE units over the course of a week. Similarly, given the increased time burden required in caring for isolated patients—approximately 4 minutes to don and 3 minutes to doff PPE outside the patient room [31]—workflow adaptations saved nurses an extra 78 and 64 minutes per morning and night shift, respectively.

The change in total time spent at patient bedside per shift from baseline did not differ significantly between the COVID-19 and comparison units, given the increased duration of each encounter. Past work suggests that this “batching” or “clustering” of bedside work includes performing physical assessments, administering medications, and delivering a food

XSL•FO
RenderX

tray all in one bedside encounter [8,32]. The impact of workflow change to fewer, longer encounters on patient safety and satisfaction is an area for future research, although recent qualitative work suggests that COVID-19 patients overall did not feel their care was compromised and accepted the technology given the need for isolation precautions [28]. Further, the novel finding that total in-person nursing time at the bedside is unchanged following a telehealth deployment may be favorable, particularly given the known positive link between nursing time spent at the bedside and patient safety [19].

Notably, Figure 1 suggests an inverse relation between the mean number of direct care events and the mean number of telehealth calls during surge versus nonsurge periods. While nurses shifted their practice patterns to fewer, longer encounters during surges, it appears they also increased telehealth use before reverting back to standard practice patterns during nonsurge periods. This simultaneous adoption of telehealth on the COVID-19 unit suggests that virtual care was an additive complement to, rather than a replacement for, in-person care in terms of total time spent with the patient. Our findings therefore point to the possible role inpatient telehealth could play in improving patient safety for isolated patients. Specifically, understanding the optimal ratio for in-person to virtual encounters as well as patient and clinician triggers for each type of encounter may further inform telehealth use across varied inpatient settings.

The COVID-19 unit and comparison units had important differences that likely influenced the adoption of telehealth. In the COVID-19 unit, hardware was readily available in each patient room, whereas the comparison units had a limited number of carts in a central location on the unit; the hardware had to be retrieved and then set up in the patient rooms prior to use. In addition to the ready availability of technology, the increased threat of pathogen exposure on the COVID-19 unit likely further promoted rapid staff uptake relative to comparison units, which did not adopt telehealth until late in the observation period (surge #2) when the number of COVID-19 patients on those units increased. These factors appeared to help health care workers overcome typical barriers to telehealth adoption and integrate the technology into their regular clinical routine [33,34].

Beyond the introduction of telehealth, the observed changes in nurse-patient encounters could also be due to unobserved differences between the primary COVID-19 unit and comparator units during this real-world health crisis. Aspects of nursing care were changing, in tandem with standards for the management of COVID-19, PPE availability, and infection control recommendations [32,35,36]. Telehealth encounters may have served as a replacement for nurse-to-patient calls on the bedside phone, which were not measured in this evaluation. Further, as the hospital was responding to the pandemic surges, fluctuations in patient acuity in all units impacted nurse-to-patient ratios. As the use of inpatient telehealth continues to evolve, evaluating the long-term use and sustainability outside of the pandemic setting may be an area for future research.

This retrospective, observational study utilized readily available data in a real-world setting and thus certain limitations exist. RTLS badges were worn consistently among nurses; however, limited compliance among other health care professionals such as physicians preclude analysis of other roles [9]. Only nurses with the role of “nurse” or “float nurse” who wore the RTLS badge were included; thus, health care workers identified under a different role (eg, certified nursing assistant) or nurses not wearing a badge (eg, broken badge) were not captured in this study. Further, since the RTLS sensors are linked to room numbers, RTLS-based badge data were merged with patient-level EHR data, which were limited to the presence of a patient in a patient room and their most recent COVID-19 status based on midnight census. Therefore, changes in patient census or location that occurred throughout the day were not captured. In addition, the telehealth platform captured all telehealth encounters that occurred within the four units. However, the purpose of the call, and the identity and role of participants (eg, nurse, patient, other health care worker, family) were not captured by the telehealth platform, nor were data on the quantity and duration of calls using the bedside phone captured. This eliminated the possibility of analyzing the purpose of the telehealth communication at the patient or nurse level, which is an area for future work.

Despite these limitations, an RTLS offers an alternative to other high-burden data sources of interest, such as ethnographic observation, to provide novel insights that may not otherwise be available. The strength of this evaluation is that movement data are available for nurses and linked with individual patient encounters in this novel clinical context.

Conclusions

Assessment of nursing workflow change following the deployment of inpatient telehealth in the context of the COVID-9 pandemic was feasible utilizing RTLS data in combination with EHR data. Compared with those of other units with shared mobile health units, direct nurse-patient encounters on a COVID-19 unit with in-room ready access to telehealth decreased in frequency and increased in duration, leading to a redistribution of work that did not impact total time at the bedside relative to prepandemic periods. The simultaneous adoption of telehealth suggests virtual care complemented, rather than replaced, in-person care in this setting. Study limitations, including the lack of telehealth utilization data at the nurse or patient level and multiple differences between the COVID-19 and comparator units (ease of telehealth availability and proportion of COVID-19 patients), preclude our ability to draw a causal link between nursing workflow change and telehealth adoption. Further evaluation is needed to determine potential downstream implications on unmeasured outcomes such as disease transmission, PPE utilization, and patient safety.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank Rajiv Ramdeo, Ann-Marie Yap, Kevin Yuen, Derek Nielsen, Christopher Sharp, Gretchen Brown, and Steven Asch for their contributions to this project. This research received institutional support through the Stanford Health Care Improvement Capability Development Program. SV and BP were funded through the Stanford-Intermountain Population Health Fellowship and the Clinical Informatics Fellowship, respectively.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Methodological details, including telemedicine deployment and time intervals.
[DOCX File , 22 KB - jmir_v24i6e36882_app1.docx]

Multimedia Appendix 2
Average inpatient census by COVID-19 status (Table S1).
[DOCX File , 18 KB - jmir_v24i6e36882_app2.docx]

References

1. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
2. 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]
3. Moeckli J, Gutierrez J, Kaboli PJ. Perceived need and potential applications of a telehospitalist service in rural areas. Telemed J E Health 2021 Jan 01;27(1):90-95. [doi: 10.1089/tmj.2020.0018] [Medline: 32316876]
4. Vilendrer S, Patel B, Chadwick W, Hwa M, Asch S, Pageler N, et al. Rapid deployment of inpatient telemedicine in response to COVID-19 across three health systems. J Am Med Inform Assoc 2020 Jul 01;27(7):1102-1109 [FREE Full text] [doi: 10.1093/jamia/ocaa077] [Medline: 32495830]
5. Keller JJ, Johnson JP, Latour E. Inpatient teledermatology: diagnostic and therapeutic concordance among a hospitalist, dermatologist, and teledermatologist using store-and-forward teledermatology. J Am Acad Dermatol 2020 May;82(5):1262-1267. [doi: 10.1016/j.jaad.2020.01.030] [Medline: 31972258]
6. Pilosof NP, Barrett M, Oborn E, Barkai G, Pessach IM, Zimlichman E. Telemedicine implementation in COVID-19 ICU: balancing physical and virtual forms of visibility. HERD 2021 Jul 02;14(3):34-48 [FREE Full text] [doi: 10.1177/19375867211009225] [Medline: 34075789]
7. Ho HJ, Zhang ZX, Huang Z, Aung AH, Lim W, Chow A. Use of a real-time locating system for contact tracing of health care workers during the COVID-19 pandemic at an infectious disease center in Singapore: validation study. J Med Internet Res 2020 May 26;22(5):e19437 [FREE Full text] [doi: 10.2196/19437] [Medline: 32412416]
8. Safaeinili N, Vilendrer S, Williamson E, Zhao Z, Brown-Johnson C, Asch SM, et al. Inpatient telemedicine implementation as an infection control response to COVID-19: qualitative process evaluation study. JMIR Form Res 2021 Jun 16;5(6):e26452 [FREE Full text] [doi: 10.2196/26452] [Medline: 34033576]
9. Patel B, Vilendrer S, Kling SMR, Brown I, Ribeira R, Eisenberg M, et al. Using a real-time locating system to evaluate the impact of telemedicine in an emergency department during COVID-19: observational study. J Med Internet Res 2021 Jul 26;23(7):e29240 [FREE Full text] [doi: 10.2196/29240] [Medline: 34236993]
10. Legler S, Diehl M, Hilliard B, Olson A, Markowitz R, Tignanelli C, et al. Evaluation of an intrahospital telemedicine program for patients admitted with COVID-19: mixed methods study. J Med Internet Res 2021 Apr 29;23(4):e25987 [FREE Full text] [doi: 10.2196/25987] [Medline: 33872187]
11. Halabi R, Smith G, Sylwestrzak M, Clay B, Longhurst CA, Lander L. The impact of inpatient telemedicine on personal protective equipment savings during the COVID-19 pandemic: cross-sectional study. J Med Internet Res 2021 May 19;23(5):e28845 [FREE Full text] [doi: 10.2196/28845] [Medline: 33945494]
12. Okoniewska B, Graham A, Gavrilova M, Wah D, Gilgen J, Coke J, Ward of the 21st Century team. Multidimensional evaluation of a radio frequency identification wi-fi location tracking system in an acute-care hospital setting. J Am Med Inform Assoc 2012 Jul 01;19(4):674-679 [FREE Full text] [doi: 10.1136/amiajnl-2011-000560] [Medline: 22298566]
13. Jones TL, Schlegel C. Can real time location system technology (RTLS) provide useful estimates of time use by nursing personnel? Res Nurs Health 2014 Feb 11;37(1):75-84 [FREE Full text] [doi: 10.1002/nur.21578] [Medline: 24338915]
14. Berg B, Longley G, Dunitz J. Improving clinic operational efficiency and utilization with RTLS. J Med Syst 2019 Jan 30;43(3):56. [doi: 10.1007/s10916-019-1174-z] [Medline: 30701407]

XSL•FO
RenderX

15. Heaton HA, Nestler DM, Sadosty AT, Ernste VK, Boggust A, Clements CM, et al. R-EME: RTLS-event mapping engine applications in emergency medicine. Am J Emerg Med 2018 Feb;36(2):324-325. [doi: 10.1016/j.ajem.2017.07.055] [Medline: 28739390]
16. Jones TL. Radiofrequency identification: exploiting an old technology for measuring nurse time and motion. Comput Inform Nurs 2012 Sep;30(9):463-472 [FREE Full text] [doi: 10.1097/NXN.0b013e3182545418] [Medline: 22592451]
17. Shelley KA. Leveraging technology to measure, evaluate, and adjust nursing interventions. Nurs Manage 2020 Jan;51(1):26-33. [doi: 10.1097/01.NUMA.0000617012.12271.57] [Medline: 31880617]
18. Li RC, Marafino BJ, Nielsen D, Baiocchi M, Shieh L. Assessment of a real-time locator system to identify physician and nurse work locations. JAMA Netw Open 2020 Feb 05;3(2):e1920352 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.20352] [Medline: 32022876]
19. Mitchell MD, Lavenberg JG, Trotta RL, Umscheid CA. Hourly rounding to improve nursing responsiveness: a systematic review. J Nurs Adm 2014 Sep;44(9):462-472 [FREE Full text] [doi: 10.1097/NNA.0000000000000101] [Medline: 25148400]
20. Fahey L, Dunn Lopez K, Storfjell J, Keenan G. Expanding potential of radiofrequency nurse call systems to measure nursing time in patient rooms. J Nurs Adm 2013 May;43(5):302-307. [doi: 10.1097/NNA.0b013e31828eebe1] [Medline: 23615373]
21. Lucet J, Laouenan C, Chelius G, Veziris N, Lepelletier D, Friggeri A, et al. Electronic sensors for assessing interactions between healthcare workers and patients under airborne precautions. PLoS One 2012 May 25;7(5):e37893 [FREE Full text] [doi: 10.1371/journal.pone.0037893] [Medline: 22662245]
22. Abad C, Fearday A, Safdar N. Adverse effects of isolation in hospitalised patients: a systematic review. J Hosp Infect 2010 Oct;76(2):97-102 [FREE Full text] [doi: 10.1016/j.jhin.2010.04.027] [Medline: 20619929]
23. Siddiqui ZK, Conway SJ, Abusamaan M, Bertram A, Berry SA, Allen L, et al. Patient isolation for infection control and patient experience. Infect Control Hosp Epidemiol 2019 Feb 18;40(2):194-199. [doi: 10.1017/ice.2018.324] [Medline: 30560748]
24. Nair R, Perencevich E, Goto M, Livorsi D, Balkenende E, Kiscaden E, et al. Patient care experience with utilization of isolation precautions: systematic literature review and meta-analysis. Clin Microbiol Infect 2020 Jun;26(6):684-695 [FREE Full text] [doi: 10.1016/j.cmi.2020.01.022] [Medline: 32006691]
25. Kasprak J. California RN staffing ratio law. OLR Research Report. 2004 Feb 10. URL: https://www.cga.ct.gov/2004/rpt/2004-R-0212.htm [accessed 2021-08-24]
26. Order of the Health Officer of the County of Santa Clara Directing All Individuals Living in the County to Shelter in Place of Residence. County of Santa Clara Public Health Department. 2020 Mar 16. URL: https://covid19.sccgov.org/sites/g/files/exjcpb766/files/03-16-20-Health-Officer-Order-to-Shelter-in-Place.pdf [accessed 2021-08-24]
27. COVID-19 Risk tier categories. California for ALL. URL: https://covid19.ca.gov/ [accessed 2021-02-08]
28. Vilendrer S, Sackeyfio S, Akinbami E, Ghosh R, Luu JH, Pathak D, et al. Patient perspectives of inpatient telemedicine during the COVID-19 pandemic: qualitative assessment. JMIR Form Res 2022 Mar 30;6(3):e32933 [FREE Full text] [doi: 10.2196/32933] [Medline: 35147510]
29. Sang AX, Tisdale RL, Nielsen D, Loica-Mersa S, Miller T, Chong I, et al. How much time are physicians and nurses spending together at the patient bedside? J Hosp Med 2019 Aug 01;14(8):468-473. [doi: 10.12788/jhm.3204] [Medline: 31112496]
30. Benjamini Y, Krieger AM, Yekutieli D. Adaptive linear step-up procedures that control the false discovery rate. Biometrika 2006;93(3):491-507. [doi: 10.1093/biomet/93.3.491]
31. Lim SM, Cha WC, Chae MK, Jo IJ. Contamination during doffing of personal protective equipment by healthcare providers. Clin Exp Emerg Med 2015 Sep 30;2(3):162-167. [doi: 10.15441/ceem.15.019] [Medline: 27752591]
32. Schroeder K, Norful AA, Travers J, Aliyu S. Nursing perspectives on care delivery during the early stages of the covid-19 pandemic: A qualitative study. Int J Nurs Stud Adv 2020 Nov;2:100006 [FREE Full text] [doi: 10.1016/j.ijnsa.2020.100006] [Medline: 32864632]
33. 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]
34. May CR, Mair F, Finch T, MacFarlane A, Dowrick C, Treweek S, et al. Development of a theory of implementation and integration: normalization process theory. Implement Sci 2009 May 21;4:29 [FREE Full text] [doi: 10.1186/1748-5908-4-29] [Medline: 19460163]
35. Iacobucci G. Covid-19: Doctors still at "considerable risk" from lack of PPE, BMA warns. BMJ 2020 Mar 31;368:m1316. [doi: 10.1136/bmj.m1316] [Medline: 32234713]
36. Mahase E. Covid-19: What do we know about "long covid"? BMJ 2020 Jul 14;370:m2815. [doi: 10.1136/bmj.m2815] [Medline: 32665317]

Abbreviations

EHR: electronic health record
PPE: personal protective equipment

XSL•FO
RenderX

RTLS: real-time locating system

Edited by C Basch; submitted 31.01.22; peer-reviewed by A Neuraz, W Zhang, S Fenton; comments to author 23.03.22; revised version received 13.04.22; accepted 11.05.22; published 17.06.22.

Please cite as:
Vilendrer S, Lough ME, Garvert DW, Lambert MH, Lu JH, Patel B, Shah NH, Williams MY, Kling SMR
Nursing Workflow Change in a COVID-19 Inpatient Unit Following the Deployment of Inpatient Telehealth: Observational Study Using a Real-Time Locating System
J Med Internet Res 2022;24(6):e36882
URL: https://www.jmir.org/2022/6/e36882
doi:10.2196/36882
PMID:35635840

©Stacie Vilendrer, Mary E Lough, Donn W Garvert, Monique H Lambert, Jonathan Hsijing Lu, Birju Patel, Nigam H Shah, Michelle Y Williams, Samantha M R Kling. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 17.06.2022. 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 Longitudinal Cough, Breath, and Voice Data for COVID-19 Progression Prediction via Sequential Deep Learning: Model Development and Validation

Ting Dang[1], PhD; Jing Han[1*], PhD; Tong Xia[1*], MPhil; Dimitris Spathis[1], PhD; Erika Bondareva[1], MRES; Chloë Siegele-Brown[1], PhD; Jagmohan Chauhan[1,2], PhD; Andreas Grammenos[1], PhD; Apinan Hasthanasombat[1], MPhil; R Andres Floto[1], PhD; Pietro Cicuta[1], PhD; Cecilia Mascolo[1], PhD

[1]Department of Computer Science and Technology, University of Cambridge, Cambridge, United Kingdom

[2]Electronics and Computer Science, University of Southampton, Southampton, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Ting Dang, PhD
Department of Computer Science and Technology
University of Cambridge
15 JJ Thomson Ave
Cambridge, CB3 0FD
United Kingdom
Phone: 44 7895587796
Email: td464@cam.ac.uk

Abstract

Background: Recent work has shown the potential of using audio data (eg, cough, breathing, and voice) in the screening for COVID-19. However, these approaches only focus on one-off detection and detect the infection, given the current audio sample, but do not monitor disease progression in COVID-19. Limited exploration has been put forward to continuously monitor COVID-19 progression, especially recovery, through longitudinal audio data. Tracking disease progression characteristics and patterns of recovery could bring insights and lead to more timely treatment or treatment adjustment, as well as better resource management in health care systems.

Objective: The primary objective of this study is to explore the potential of longitudinal audio samples over time for COVID-19 progression prediction and, especially, recovery trend prediction using sequential deep learning techniques.

Methods: Crowdsourced respiratory audio data, including breathing, cough, and voice samples, from 212 individuals over 5-385 days were analyzed, alongside their self-reported COVID-19 test results. We developed and validated a deep learning–enabled tracking tool using gated recurrent units (GRUs) to detect COVID-19 progression by exploring the audio dynamics of the individuals' historical audio biomarkers. The investigation comprised 2 parts: (1) COVID-19 detection in terms of positive and negative (healthy) tests using sequential audio signals, which was primarily assessed in terms of the area under the receiver operating characteristic curve (AUROC), sensitivity, and specificity, with 95% CIs, and (2) longitudinal disease progression prediction over time in terms of probability of positive tests, which was evaluated using the correlation between the predicted probability trajectory and self-reported labels.

Results: We first explored the benefits of capturing longitudinal dynamics of audio biomarkers for COVID-19 detection. The strong performance, yielding an AUROC of 0.79, a sensitivity of 0.75, and a specificity of 0.71 supported the effectiveness of the approach compared to methods that do not leverage longitudinal dynamics. We further examined the predicted disease progression trajectory, which displayed high consistency with longitudinal test results with a correlation of 0.75 in the test cohort and 0.86 in a subset of the test cohort with 12 (57.1%) of 21 COVID-19–positive participants who reported disease recovery. Our findings suggest that monitoring COVID-19 evolution via longitudinal audio data has potential in the tracking of individuals' disease progression and recovery.

Conclusions: An audio-based COVID-19 progression monitoring system was developed using deep learning techniques, with strong performance showing high consistency between the predicted trajectory and the test results over time, especially for recovery trend predictions. This has good potential in the postpeak and postpandemic era that can help guide medical treatment

XSL•FO
RenderX

and optimize hospital resource allocations. The changes in longitudinal audio samples, referred to as audio dynamics, are associated with COVID-19 progression; thus, modeling the audio dynamics can potentially capture the underlying disease progression process and further aid COVID-19 progression prediction. This framework provides a flexible, affordable, and timely tool for COVID-19 tracking, and more importantly, it also provides a proof of concept of how telemonitoring could be applicable to respiratory diseases monitoring, in general.

(J Med Internet Res 2022;24(6):e37004) doi:10.2196/37004

KEYWORDS

COVID-19; audio; COVID-19 progression; deep learning; mobile health; longitudinal study

Introduction

Background

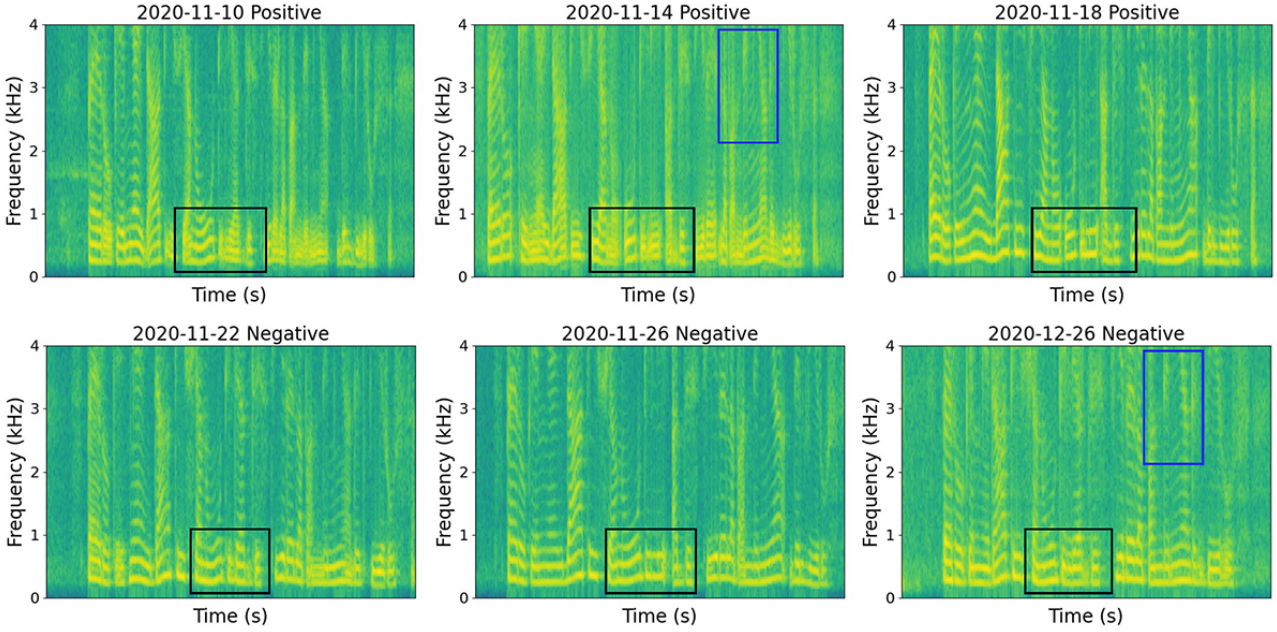
Since the beginning of the SARS-CoV-2 pandemic in January 2020, different methods have been developed and used for diagnostic testing and screening. In addition to the most commonly adopted laboratory tests via reverse transcription polymerase chain reaction (RT-PCR) [1,2] or chest computed tomography (CT) scans [3] for diagnosis, a variety of digital technologies, often using artificial intelligence, have also been investigated for COVID-19 screening [4-7]. Among these, automatic audio-based COVID-19 detection has drawn increasing attention due to its numerous advantages, including its flexible, affordable, scalable, noninvasive, and sustainable data collection methods.

The existing literature has mainly investigated the information content of different audio modalities (eg, cough, breathing, and voice) [8-12] and the power of various machine learning techniques, especially deep learning for COVID-19 detection [10,13-19]. Although success has been witnessed recently in COVID-19 detection from audio signals through machine learning techniques [10], there is still a paucity of work on continuous monitoring of COVID-19 progression. This could provide individual-specific, timely indication of disease development at scale, guide personalized medical treatments, potentially capture disease onset to curb transmission, and estimate the recovery rate, which plays a key role in determining quarantine rules during the current postpeak and postpandemic times. It would also allow better resource management in health care systems, while remotely monitoring patients and bringing them to the hospital (only) when necessary. Evidence suggests that COVID-19 progression varies among individuals, with the mean disease duration ranging from 11 to 21 days, depending on gender and age, comorbidities, variants of SARS-CoV-2, and time receiving treatment [20-25]. By continuously monitoring patients' disease progression, individual-specific information could be captured to benefit both patients and doctors. In addition, compared to the commonly adopted uncomfortable diagnostic methods of RT-PCR tests and radiation-intensive CT scans that are conducted on hospital sites, audio-based monitoring of disease progression can be nonintrusively repeated on a daily basis and for prolonged periods, proving a good fit for longitudinal remote monitoring.

Recent work has recognized the use of a 3-escalating-phase description of COVID-19 progression [26], including early infection, pulmonary involvement, and systemic hyperinflammation, which demonstrates commonality in disease progression. We hypothesize that this could be captured longitudinally via audio signals in automatic monitoring systems. Though the participants in our study may not experience all 3 stages, it is assumed that audio characteristics are affected during clinical progression of the disease. Figure 1 shows spectrograms of 1 participant reading the same sentence over a 43-day period who reported COVID-19 infection followed by recovery. As indicated in the black box, the fundamental frequency and its harmonics were not clearly separable when the participant tested positive (top row), especially on November 14, 2020, indicating a lack of control of vibration of the vocal cords. The separability increased after recovery. This matches the observed clinical course of COVID-19 progression [22], with the least separability 5 days after the first positive test result (November 14, 2020) and an increase in separability 9 days after infection (November 18, 2020). Similar patterns were also observed for the harmonics in the high-frequency range from 2 to 4 kHz (blue box). There was no obvious difference in the spectrogram patterns between November 18 (positive) and November 22 (negative), reflecting the difficulty of the COVID-19 detection task in general. Overall, this evolution of the spectrograms with disease progression shows that COVID-19 infection can manifest as changes in acoustic representations. As disease progression varies among individuals (eg, different recovery times or different severity levels), longitudinal audio changes could vary among individuals. However, it is common that longitudinal audio changes present with COVID-19, and modeling them can potentially benefit COVID-19 progression prediction.

XSL•FO
RenderX

Figure 1. Sound recordings have distinct features during disease progression. This is evident here in the spectrograms of 1 participant who repeated the same sentence on 6 different days. The participant reported positive test results from November 10 to 18, 2020, and reported negative test results from November 22 to December 26, 2020, indicating a recovery trend. The fundamental frequency and its harmonics (black box) for positive recordings demonstrate a lack of control in vocal cords, indicated by their nonseparability. An increasing separability can be seen from positive to negative recordings over time, suggesting the recovery of voice characteristics. Similarly, the harmonics in the frequency range (2-4 kHz, blue box) manifest increasing separability, also reflecting the recovery trend.



Furthermore, audio characteristics vary among individuals (eg, one COVID-19–*positive* participant may produce a similar spectrogram as another COVID-19–*negative* participant). This is not considered in most conventional audio-based COVID-19 detection systems, which so far have only used a single audio sample rather than sequences. This makes automatic detection a difficult task and may lead to wrong predictions. While longitudinally modeling the evolution of spectrograms, the individual's past audio signal can serve as a baseline and the predictions can be corrected, given this reference. Additionally, the spectrogram for each individual when healthy can be used as a reference for its own infected status, and longitudinally modeling the relative changes in the audio sequences is likely to be more accurate for COVID-19 detection. In a much generalized sense, the mean and SD of the audio recordings in an individual's healthy state are both personalized. This provides a good threshold for the nonhealthy states and benefits COVID-19 detection. Motivated by these advantages, we explored the potential of longitudinally modeling sequential audio recordings as biomarkers of disease progression, focusing on how best to capture dynamic changes in the audio sequences over time and aiming to demonstrate predictive power.

Objective

In this study, we developed an audio-based COVID-19 progression prediction model using longitudinal audio data. We adopted sequential deep learning models to capture longitudinal audio dynamics and to make predictions of disease progression over time. First, we examined whether modeling audio dynamics could aid in COVID-19 detection. This showed strong performance compared to conventional models using a single audio sample. We then evaluated our model's performance in predicting disease progression trajectories: our predictions successfully tracked test results and also matched the statistical analysis in the timeline and duration of COVID-19 progression. In particular, we explored the use of audio signals for recovery prediction, as this may be useful in relation to setting home quarantine requirements. From a public health perspective, an approach such as the one we propose has potential implications for how infected individuals are monitored, namely it could allow more fine-grained remote tracking and hence more efficient management of health system resources by keeping individuals out of the hospital as much as possible.

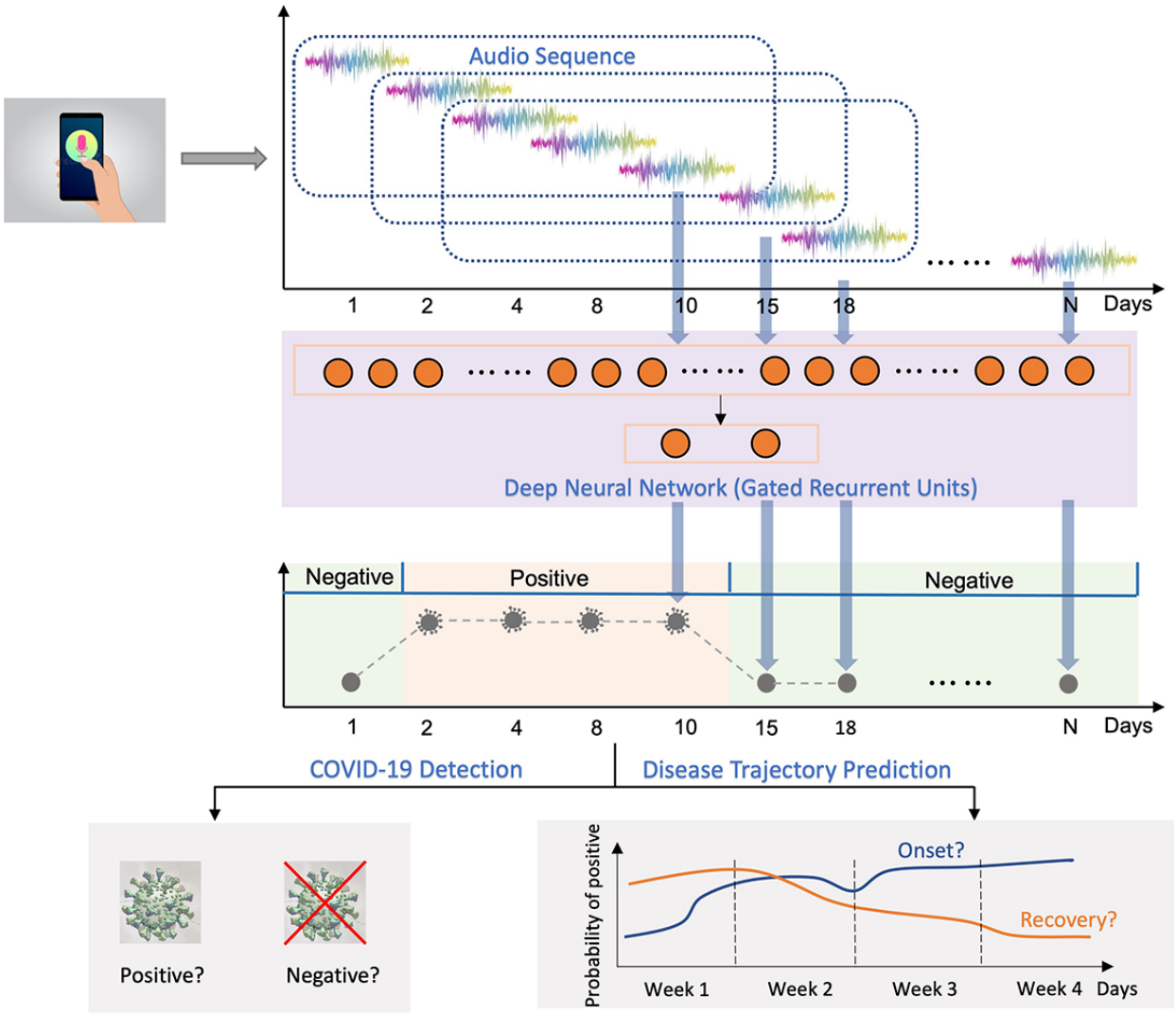
Methods

Study Design and Overview

We investigated whether longitudinally modeling audio biomarkers (cough, breath, and voice) can benefit COVID-19 detection and whether it could be used to monitor disease progression accurately and in a timely manner (Figure 2). The audio sequences were modeled by recurrent neural networks with gated recurrent units (GRUs) to consider audio dynamics, which reflect disease progression. The investigation was divided into 2 subtasks: one concerning COVID-19 detection by predicting audio biomarkers as positive and negative and the other concerning disease progression trajectory monitoring, examining the predicted probability of being positive over time. For instance, a decrease in the probability of being positive over time indicates a recovery trend, while an increase indicates a worsening trend. The first subtask aimed to assess whether modeling past audio biomarkers in the input space benefits COVID-19 detection in general, while the second subtask focused on longitudinal analysis of disease progression in the output space.

XSL•FO
RenderX

Figure 2. Overview of study design: COVID-19 detection and progression were assessed from audio data. Voice, cough, and breathing sound recordings were collected from each participant over a period, together with self-reported COVID-19 test results. During model development, audio recordings were chunked into segments consisting of 5 samples covering a few days and processed using sequential modeling techniques (GRUs) for COVID-19 monitoring. Two subtasks were evaluated: (1) COVID-19 detection (positive vs negative) and (2) disease progression monitoring. GRU: gated recurrent unit.



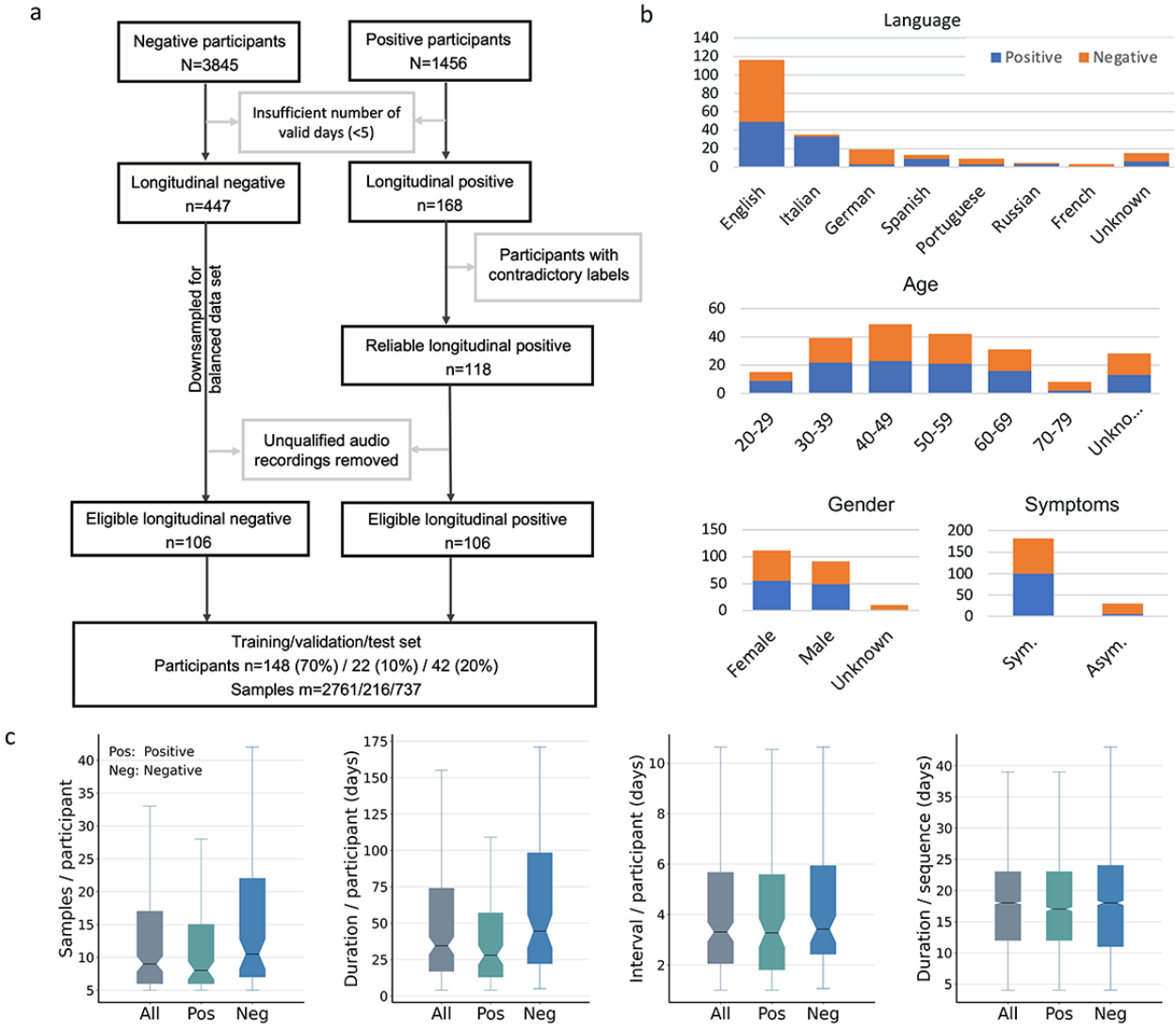
Data Set Preparation and Statistics

A mobile app [27] was developed and released in April 2020 for data gathering, aiming to crowdsource participants' audio recordings, COVID-19 test results, and demographics, medical history, and COVID-19–related symptoms. Each participant was asked to record 3 different audio sounds on a single day, including a cough recording with 3 voluntary cough sounds, a breathing recording with 3-5 breathing sounds, and a voice recording where each participant was asked to read a short phrase displayed on the screen, 3 times. The COVID-19 test results were self-reported, chosen from a positive report, a negative one, or not having been tested. No specific methods for the diagnosis were required, which can be a lateral flow test result, an RT-PCR test result, or a CT scan result. Participants were encouraged to provide data regularly. More details can be found in Multimedia Appendix 1 and Xia et al [28].

From April 2020 to April 2021 (Figure 3a), 3845 healthy participants (COVID-19–negative) and 1456 COVID-19–positive participants contributed audio samples with positive or negative clinical self-reported test results for at least 1 day. We used these participants' data if 5 or more samples were provided, resulting in 447 (11.6%) and 168 (11.5%) negatively and positively tested participants, respectively. Label quality was manually checked to remove unreliable users, and audio recording quality was examined using Yet Another Mobile Network (YAMNet) [29] to filter out corrupted and noisy samples, leaving 106 (63.1%) COVID-19–positive participants. To generate a balanced data set, a cohort of 212 longitudinal users in total (106, 50%, COVID-19–positive and 106, 50%, COVID-19–negative) was selected across different countries.

XSL•FO
RenderX

Figure 3. Data flow diagram and demographic statistics. Large data sets were required to identify and avoid biases. (a) Data selection process. (b) Demographic statistics of eligible participants, including language, gender, age, and symptoms. English was the dominant language, comprising 54.2% (n=115) of the cohort. Age and gender were relatively balanced between positive and negative groups. In addition, 100 (94.3%) COVID-19–positive participants and 82 (77.4%) COVID-19–negative participants reported COVID-19 symptoms. (c) Duration and reporting intervals in terms of days and number of samples. The median number of samples was 9 (left), corresponding to a time span of 35 days (middle left). COVID-19–negative participants reported for a longer period compared to COVID-19–positive participants. The median reporting interval for the cohort was 3 days (middle right), validating the effective temporal dependencies of the audio data. The median duration after augmentation was 17 and 18 days for COVID-19–positive and COVID-19–negative participants, respectively (right), showing that the augmentation eliminated the confounding effects of the different duration for the 2 subgroups.



The mobile app is a multilanguage platform, and the cohort consisted of 8 different languages, with 115 (54.2%) English users as the dominant subgroup (Figure 3b). Age and gender were relatively balanced between COVID-19–positive and COVID-19–negative groups, with 110 (51.9%) female participants and 142 (67.0%) participants aged between 30 and 59 years. In addition, 100 (94.3%) of 106 COVID-19–positive participants and 82 (77.4%) of 106 COVID-19–negative participants reported COVID-19 symptoms, such as loss of taste or smell, and fever. The median number of samples for each user was 9 (Figure 3c, left), corresponding to a period of 35 days (Figure 3c, middle left). The median duration for COVID-19–positive participants was smaller than for COVID-19–negative participants, namely 28 and 45 days, respectively. This duration is expected to contain adequate audio dynamics associated with disease progression and to cover the complete course of disease progression for most participants [21]. In addition, the reporting interval for each participant was also computed as the average of the time intervals between 2 consecutive samples, and the median value was 3 days for both COVID-19–positive and COVID-19–negative groups (Figure 3c, middle right), validating the temporal dependencies of the data. To develop machine learning models, data augmentation was carried out, and the duration after augmentation for COVID-19–positive and COVID-19–negative participants was balanced, with a median value of 17 and 18 days, respectively (Figure 3c, right). This duration aligns with the disease progression duration that is generally from 11-21 days [20-24]. The similar duration for COVID-19–positive and COVID-19–negative participants after augmentation also helped to eliminate the confounding effect of the original different duration for the 2 groups in model development (Figure 3c,

XSL•FO
RenderX

left). The data were split into training, validation, and test sets, with 148 (70%), 22 (10%), and 42 (20%) balanced COVID-19–positive and COVID-19–negative participants, respectively, as well as the relatively balanced languages and genders (see Multimedia Appendix 2).

Data Processing

To effectively develop the deep learning model, we ensured a sufficient amount of appropriately processed data available for modeling by performing audio preprocessing, sequence generation, and data augmentation.

Audio Preprocessing

Audio recordings were first resampled to 16 kHz and converted to a mono channel. These audio recordings were then preprocessed by removing the silence periods at the beginning and end of the recordings, normalized to a maximum amplitude of 1.

Sequence Generation

To increase the number of audio sequences for model development, the audio samples for each participant were chunked into short sequences with a fixed number of 5 samples. To ensure the 5 samples within an audio sequence contained effective and adequate audio dynamics in COVID-19 progression, a further constraint was applied, limiting the maximum time gap between 2 subsequent samples to 14 days. Any sequences that violated this constraint were removed. This resulted in sequence lengths ranging from 5 to 56 days, covering the time span of disease progression.

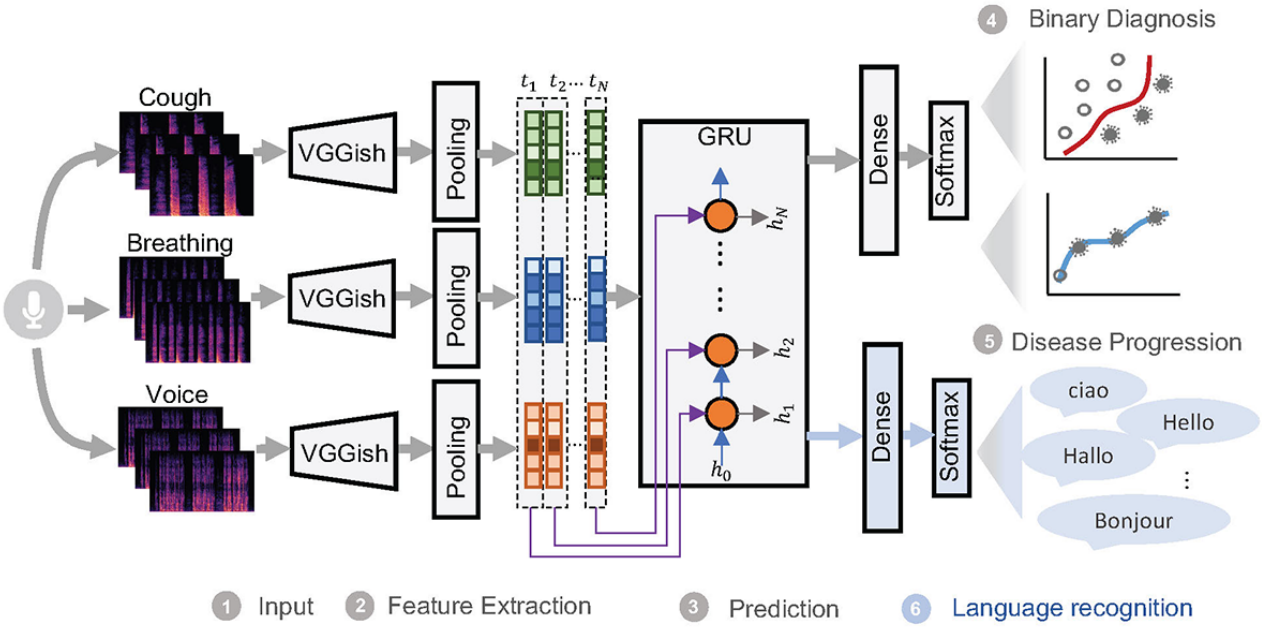
Data Augmentation

Though the data set was selected to balance the number of participants for COVID-19–positive and COVID-19–negative groups, the number of samples for each participant was still different (refer to Figure 3c). The COVID-19–negative participants provided more samples than COVID-19–positive participants, and the COVID-19–positive participants also provided negative samples after recovery, leading to a significantly larger number of COVID-19–negative samples than COVID-19–positive samples in the cohort. Further, the data set was relatively small. Therefore, 3 augmentation techniques were used to increase the data size and to balance the COVID-19–positive and COVID-19–negative samples (see Multimedia Appendix 1).

Model Architecture

The proposed model consisted of a pretrained network of VGGish (Google) for feature extraction and a recurrent neural network of GRUs for disease prediction (Figure 4). Three different modalities (breathing, cough, and voice recordings) were adopted as the input. For each modality, audio recordings were first converted to spectrograms and then fed into a VGGish pretrained network for higher-level feature representation, which could help leverage and transfer the knowledge learned from an external massive general audio data set [29]. The embeddings converted by VGGish from the 3 modalities were concatenated to form a multimodal input vector for the subsequent GRU-based prediction network. The reason for choosing GRUs over a long short-term memory neural network is fewer parameters in limited data size regimes. GRUs also use less memory and execute faster, which would be a benefit during potential model deployment. The outputs from the GRUs were evaluated for 2 different tasks: one estimating the model capability in binary diagnosis by taking the binary output of the model and the other predicting the disease progression trajectory by utilizing the probabilities of positive predictions. Further details can be found in Multimedia Appendix 1.

Figure 4. Model structure. A pretrained convolutional neural network (CNN)–based model VGGish was used as the feature extractor, and GRUs were used as a classifier, followed by dense layers, to account for longitudinal audio dynamics. This is a multitask learning framework, with COVID-19 detection as the main task and language detection as an auxiliary task to avoid language bias. h_i, i ∑ [1,2,…N] represents the hidden vectors in the GRUs for time step t_i. The reverse layer is used for the language task, as shown in Multimedia Appendix 1, Equation (5). GRU: gated recurrent unit.



XSL•FO
RenderX

Since the data set contained voices in 8 different languages and the disease prevalence for each language was different (Figure 3b), a language bias may have been introduced in the machine learning models, leading to the model recognizing the language instead of COVID-19–related information (eg, classifying Italian speakers as COVID-19–positive and English speakers as COVID-19–negative owing to the higher prevalence of the disease in Italian-speaking users and lower prevalence in English speakers). To reduce the language impact, a multitask learning framework was proposed to include the auxiliary task of language recognition simultaneously with the COVID-19 prediction task (see Multimedia Appendix 1).

Performance Evaluation

As the sequence length in the training data varied within the range of 5-56 days, the model was able to capture longitudinal audio dynamics with varying length. Therefore, during the inference stage, the prediction was made, given all the past audio recordings with no fixed number of samples. This was slightly different from the training phase but was more practical in real applications. To maintain the effective temporal dependencies of the audio recordings and match the training, a time constraint was applied to account only for the past audio recordings within 56 days before the current day, the maximum duration in the training phase. Further, we evaluated the model performance from the second sample of each participant to make sure the predictions captured the longitudinal audio dynamics.

COVID-19 Detection

For COVID-19 detection, the performance was measured using the AUROC, sensitivity, and specificity. The AUROC illustrated the diagnostic ability of the binary classifier. Sensitivity showed the model's ability to identify correctly COVID-19–positive samples, while specificity showed the model's ability to correctly identify samples without the disease.

Disease Progression Prediction

For the disease trajectory prediction, model performance was evaluated for each participant. There were 2 different metrics used based on individuals' test labels. For participants who recorded any transitions between positive and negative test results during the reported period, we adopted the point-biserial correlation coefficient γ_{pb} to measure the correlation between the predicted probability of positive test results and test labels. For participants who reported positive and negative test results consistently over the reported period, it was not possible to compute the correlation between continuous predictions and test labels. Therefore, we adopted the accuracy γ computed as the ratio of the correctly predicted samples over the total number of samples (see Multimedia Appendix 1). Though γ_{pb} ranges between –1 and 1 and γ is in the range of 0-1, the higher the value, the better the predictions for both metrics. It is also expected that γ_{pb} achieves a positive correlation from 0 to 1 for a good model. Therefore, we reported performance by combining these 2 measures of γ_{pb} and γ.

Recovery Trajectory Prediction

We further examined model performance for the recovery subgroup in the test cohort, where 12 (57.1%) of 21 COVID-19–positive participants reported a recovery trend in their test results. We adopted γ_{pb} as the evaluation metric. As we noted that there may be a delay in participants taking clinical tests or reporting results, negative predictions could be earlier than self-reported negative test results, which is acceptable. We further aligned predictions and test results temporally using dynamic time warping (DTW) and computed γ_{pb} with aligned predictions to account for the delay.

Latent Space Visualization

To gain an in-depth understanding of the model, we aimed to compare the intermediate audio representations of the model for different participants, including a participant who reported infection followed by recovery, a participant who continuously reported positive test results, and a healthy participant. The intermediate audio representations were taken as latent vectors from the last hidden layer of the model and further projected to a 4-dimension latent space using principal component analysis.

Statistical Analysis

Disease Progression With Symptoms

We studied the correlations between the predicted probability and symptoms for COVID-19–positive and COVID-19–negative participants, respectively. For COVID-19–positive participants, we assumed that the number of symptoms is correlated with the severity of illness; thus, a high probability of positive predictions was expected for the audio recordings reported alongside more symptoms. This led to a high correlation between the predicted probability and the number of symptoms. On the contrary, for the healthy participants, the number of symptoms was not correlated with the severity of illness; thus, no correlation was expected. This can validate whether the model is capable of learning COVID-19–related information instead of symptom-related information.

Disease Progression in the First 7 Days

Evidence on chest X-ray or CT showed that 56 (22.6%) patients presenting with disease experienced resolution 7 days after symptom onset, 30 (12.1%) showed a stable condition, and the remaining 162 (65.3%) patients worsened within 7 days from symptom onset [26]. We analyzed the predicted trajectories over a similar period, namely the first 7 days, to compare the statistics of the predicted trend with the reported ones. Though many participants reported symptoms on the first day they started recording, which may not be the first day they experienced symptoms, the increasing trend in the first few days could still suggest initial worsening of the patients' condition. We defined the 7-day window to be from either the first day of reported symptoms or the day of the first positive test if no symptoms were reported before that.

Ethical Considerations

The study was approved by the ethics committee of the Department of Computer Science at the University of Cambridge (with ID #722). Our mobile app displays a consent screen, where we ask the user’s permission to participate in the study by using the app.

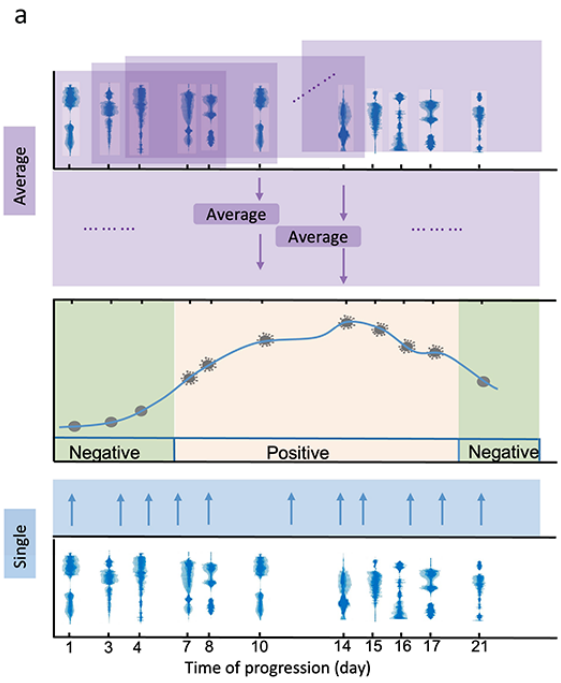
XSL•FO
RenderX

Results

COVID-19 Detection

To determine whether considering audio dynamics via the sequential modeling techniques of GRUs is effective in detecting COVID-19, the performance was compared against 2 benchmarks that do not capture audio dynamics (Figure 5): one uses only audio biomarkers of the same day for prediction, while the other uses the average feature representation of audio sequences in the prior days for the last day (Figure 5a). The proposed system (Figure 5b) outperforms the 2 benchmarks with the highest AUROC of 0.79 (0.74-0.84), a sensitivity of 0.75 (0.67-0.82), and a specificity of 0.71 (0.67-0.75), yielding 19.7% and 31.7% relative improvement in terms of the AUROC over the 2 benchmarks and demonstrating the effectiveness of modeling longitudinal audio biomarkers. We used the 1-tailed *z* test to validate the significance of the performance improvement in terms of the AUROC of the proposed approach over 2 baselines. We found that the proposed approach significantly improved the performance over "single" (P=.09) and "average" (P=.03).

Figure 5. The proposed sequential model shows superior performance in COVID-19 detection compared to benchmarks leveraging only 1 isolated audio data point per user. (a) "Average" means using the average of feature representations within the sequence for prediction, and "Single" means using only the feature representation on the same day for prediction. None of these systems capture longitudinal voice dynamics. (b) The proposed sequential modeling outperformed 2 benchmarks, suggesting the advantages of capturing disease progression via voice dynamics. (c) Individual-level accuracy for 42 participants in the test cohort.

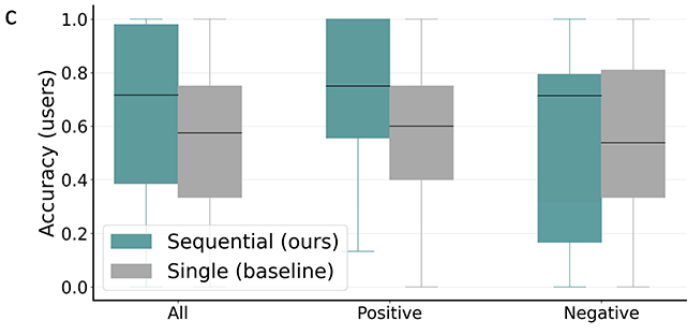


b

Table 1: COVID-19 detection performance for the test cohort

Input	AUROC	Sensitivity	Specificity
Single	0.66(0.61-0.72)	0.64(0.55-0.73)	0.61(0.57-0.66)
Average	0.60(0.53-0.66)	0.62(0.53-0.70)	0.55(0.50-0.59)
Sequential (ours)	0.79(0.75-0.84)	0.75(0.67-0.82)	0.71(0.67-0.75)

(-) is the 95% CI.



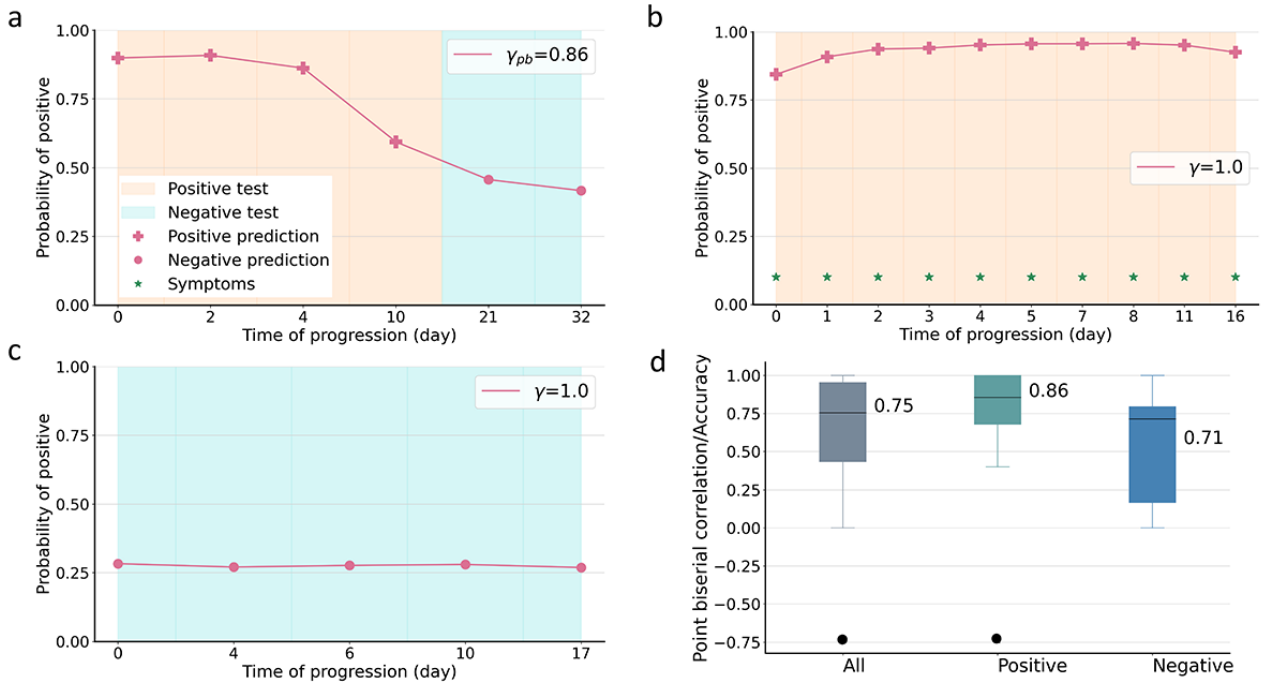
For further assessment of whether sequential modeling using past audio biomarkers could provide an adjustable baseline for each individual, the prediction accuracy for each participant was evaluated and compared to the "single" benchmark, defined as the ratio of correctly predicted samples over the total number of samples for each participant (Figure 5c). We observed that the proposed sequential modeling outperformed the "single" benchmark. The performance range of the sequential model for negative participants was larger than the benchmark, due to worse performance on 2 individuals.

Disease Progression Prediction

We analyzed the predicted disease progression trajectory by comparing it with the test results. The predicted progression trajectory is represented by the probability of positive prediction within the range of 0-1 over time, with a higher value indicating a high possibility of positive test results (Figure 6). Three different disease progression trajectories are shown in Figure 6a, Figure 6b, and Figure 6c, respectively. For the recovering participant P1 (Figure 6a), a high probability was observed when P1 tested positive and a low probability when P1 tested negative. The model performance was evaluated using the point-biserial correlation coefficients γ_{pb} that measured the correlation between the predicted trajectory and the test results. Our model achieved γ_{pb}=0.86 for P1, demonstrating a strong capability to predict disease progression. We further categorized the probability over 0.5 as a positive prediction and below 0.5 as a negative prediction. The predictions also matched the test results.

XSL•FO
RenderX

Figure 6. Our approach enabled prediction of disease progression. Orange and cyan indicate positive and negative test results, and + and • represent positive and negative predictions, respectively. The green star indicates the presence of symptoms. (a) Disease progression of recovering participant P1. (b) Disease progression of COVID-19–positive participant P2. (c) Disease progression of COVID-19–negative participant P3. (d) Overall performance for the test cohort in terms of the point-biserial correlation coefficient γ_{pb}/accuracy γ.



For the COVID-19–positive participant P2 who consistently reported positive test results (Figure 6b), our model outputs positive predictions matching the test results. It should be noted that γ_{pb} is not applicable to participants who report consistently positive or negative test results; therefore, accuracy γ was instead used, which computes the ratio of correct predictions in terms of positive or negative predictions over the total number of samples. Participant P2 had γ=1 since all the predictions were positive. Further, the probability of positive predictions increased from symptom onset day 0 to day 8 and decreased slightly to day 16, which matches the clinical course in general.

For the COVID-19–negative participant P3 (Figure 6c), the predicted probability was consistently below 0.5, corresponding to negative predictions that align with the test results. One type of disease progression that transits from negative to positive was not included due to the limited number of participants in the cohort. Though we adopted time-inverse augmentation (see Multimedia Appendix 1) that reverses the audio biomarkers and their corresponding labels in time to enrich the different disease progressions, especially the negative-to-positive transitions, the time-reversed audio biomarkers and disease progression may still not match the actual progression and cannot be well captured in the model.

Figure 6d demonstrates the model performance in disease progression monitoring for the entire test cohort, where γ_{pb} and γ were adopted for different participants. Our model achieved 0.75 for all the 42 (100%) test participants and 0.86 and 0.71 for COVID-19–positive and COVID-19–negative participants, respectively. Some more examples are given in Multimedia Appendix 3 and Multimedia Appendix 4.

Recovery Trajectory Prediction

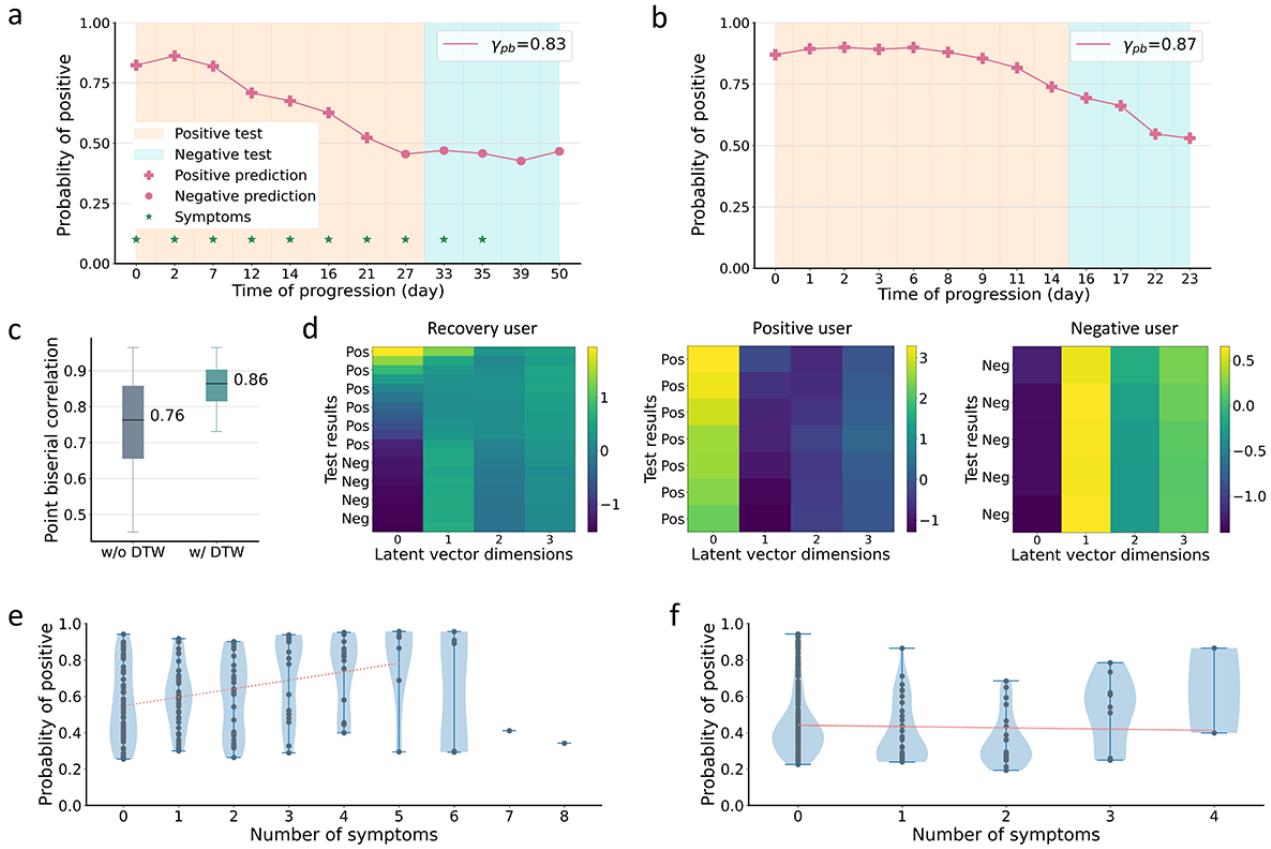
The predicted recovery trajectory of 2 randomly selected participants P4 and P5 are presented in Figure 7a and Figure 7b. For participant P4, a slight increase in probability was observed from day 0 to day 2, suggesting an increase in the severity of illness. The predicted probability decreased from day 2, showing a recovery trend. The categorized positive and negative predictions also matched the test results, except for day 27, with our model predicting negative result and the test result still being positive. This is possibly due to a delay in participants taking clinical tests or reporting results; therefore, earlier negative predictions are acceptable. This also suggests the advantages of our audio-based data, which are precisely timed and can be instantly analyzed.

For a weakly predicted trajectory of participant P5 in Figure 7b, a predicted recovery trend with decreasing probability was clearly observed. However, the probabilities were all categorized as positive predictions even after the user tested negative from day 16 to day 23. This is possibly due to (1) individual differences in audio characteristics or (2) an asymptomatic participant exhibiting minor changes in audio characteristics, thus leading to a slowly recovering trend prediction.

The overall performance for all 12 (100%) recovering participants is reported in Figure 7c, with γ_{pb}=0.76. As negative predictions with a time shift from negative test results are acceptable (as shown in Figure 7a), we aligned predictions and test results temporally using DTW. The model further achieved γ_{pb}=0.86, demonstrating a strong capability to monitor recovery. Some more examples can be found in Multimedia Appendix 3 and Multimedia Appendix 4.

XSL•FO
RenderX

Figure 7. Recovery trends can be predicted. The orange and cyan areas indicate positive and negative test results, respectively. The predictions above 0.5 were categorized as positive predictions (+) and below 0.5 as negative predictions (•). (a,b) Recovery predictions for 2 different participants P4 and P5, respectively. (c) Overall performance for recovery participants in the test cohort with and without DWT, which calculates the optimal match between the predicted recovery trajectory and test results. (d) Projection of latent vectors learnt by the model for 3 different participants. The y axis from top to bottom indicates the test results over time. A clear change in each latent vector dimension transitioning from positive to negative can be observed (recovering user), and consistent and different patterns can be observed for COVID-19–positive and COVID-19–negative participants. (e,f) Scatter plot of symptoms vs probability of positive predictions for COVID-19–positive (e) and COVID-19–negative (f) participants, with a high correlation observed for COVID-19–positive participants and no correlation for COVID-19–negative participants. DWT; dynamic time warping.



Latent Space Visualization

Figure 7d shows the latent space visualization for 3 different participants. For the recovery user, a clear transition was observed for the first 3 dimensions when the participant recovered, while consistent but different patterns were observed for the COVID-19–positive and COVID-19–negative participants. This validates that the model can capture the different disease progression of different participants.

Statistical Analysis

Figure 7e and Figure 7f display the correlations between the predicted probability and symptoms for COVID-19–positive and COVID-19–negative participants, respectively. With an increase in the number of symptoms, we observed a general increase in the predicted probability (Figure 7e). We further fit a line (red) for these participants, excluding those with more than 5 symptoms due to the limited number of samples. We observed a clear positive correlation. Conversely, for COVID-19–negative participants (Figure 7f), we observed no correlation between the probability and number of symptoms, suggesting that our model does not capture symptoms but information related to COVID-19.

Further analysis of the model predictions in the first 7 days was carried out on 12 (28.6%) eligible participants in the test cohort, where we found that 8 (66.7%) of 12 participants show an increase in predicted probability, similar to the statistical analysis (n=162, 65.3%). We hypothesized that the predicted progression in the first 7 days could provide a prompter indication of an individual's recovery rate.

Discussion

Principal Results

From a crowdsourced audio data set, we studied 212 longitudinal participants and developed a deep learning model for COVID-19 progression monitoring via audio signals. We showed that modeling audio dynamics longitudinally shows benefit for COVID-19 detection. Individual-level performance also displayed a significant improvement over baseline. The model capability to predict disease progression was validated. Successful tracking of reported test labels showed strong performance in disease progression with γ_{pb}/γ=0.76. We specifically focused on recovery prediction, with a correlation γ_{pb} of 0.86 between the predicted progression trajectory and test results.

XSL•FO
RenderX

Individuals experience different disease progression trajectories, and our model can capture this variability among individuals. For the COVID-19–positive user P6 (Multimedia Appendix 3, Figure A3c), our model demonstrated a decrease in the predicted probability from day 0 to day 3, and this was followed by a slight increase from day 3 to day 6. For participant P7 (Multimedia Appendix 3, Figure A3d), our model predicted a continuous decrease from day 0 to 11. This suggests the effectiveness in predicting individual-specific disease progression trajectories. Though there is no reported severity of illness to validate the predicted probabilities, symptoms can be used as a reference. For participant P6, the number of symptoms increased from 3 to 6 at day 3 and decreased to 3 after. Therefore, it is reasonable to assume a worsening condition and an increase in the predicted probability.

Similarly, the model can also predict individual-specific recovery trajectories. A sharper recovery trend was observed for participant P1, with a 49.2% relative decrease in 21 days (Figure 6a) than for participants P4 (Figure 7a) and P5 (Figure 7b), with a 36.6% and 37.1% relative decrease in 21 and 22 days, respectively. This is consistent with the evidence that recovery tends to be faster in younger people [22], as participant P1 was in the age group of 20-29 years, while participants P4 and P5 were aged between 30-39 and 40-49 years, respectively. Though it is difficult to draw statistical conclusions due to the limited number of participants, these results still suggest differences in the predicted recovery rate for different individuals. In terms of practical applications, individual-specific recovery monitoring may be beneficial in providing prompt feedback to self-isolating patients and, more importantly, can provide treatment guidance for doctors according to each individual's recovery status. Specifically, when a sharp decrease in the predicted probability is observed, it indicates that the individual is recovering well. Conversely, no decrease in the predicted probability over a long period may require further or more effective treatment. Additionally, the predicted recovery trend could also be used to some extent for risk assessment of COVID-19 patients.

As our model showed strong performance in using longitudinal audio biomarkers, another important factor in the model deployment is the impact of sequence length, which was also analyzed to obtain insights into how many samples are needed for reliable predictions (Multimedia Appendix 5). The cumulative histogram suggests that the longer the sequence, the better the performance. For sequence lengths with more than 2 samples (Multimedia Appendix 5, Figure A2a) or around and more than 4 days (Multimedia Appendix 5, Figure A2b), the model can produce reasonably good predictions. For telemonitoring purposes, the use of audio recordings from the past 4 days would offer a more reliable prediction.

Limitations

Our study also had several limitations. First, the testing cohort was relatively small with only 21 participants for the COVID-19–positive and COVID-19–negative groups. This may not comprehensively represent the target population. In addition, the self-reported test results may inevitably be noisy to some extent, where a mismatch between audio recordings and test results may exist. This is due to possible delays in participants reporting the test results. This mismatch introduced confounding variability into the model development that was not fully considered.

The other limitation in our study is the limited control over confounding factors. The age and gender groups were relatively balanced within and across the training, validation, and test sets, while language was only balanced between 3 partitions but still unbalanced within each data partition. Our model using a multitask framework mitigated the language impact, but some language bias might remain due to the limited number of samples of some language subgroups.

We also acknowledged that changes in voice may be attributed to not only COVID-19 infection but also other factors (eg, mental state or other respiratory infections, such as influenza). To validate whether the model captures changes caused by COVID-19 instead of other factors, a large amount of longitudinal data that have the corresponding labels (eg, emotion state, influenza) is required to develop and evaluate the model. Collecting such data is difficult and time-consuming, which is our long-term goal.

It is also worth noting that the predicted disease progression trend matches the test results, but for some users, probabilities may be overall high or low over the course of COVID-19 progression. This suggests individual differences in audio characteristics. Though our model resolves this better than simple sample-based models by capturing past audio signals, it is a universal model and therefore still imprecise. The development of participant-specific models is on our future agenda, but more data needs to be collected for this purpose.

Conclusion

In conclusion, by modeling audio biomarkers longitudinally with sequential machine learning techniques, we proposed audio-based diagnostics with longitudinal data as a robust technique for COVID-19 progression tracking. We showed that our system is able to monitor disease progression, especially the recovery trajectory of individuals. This work not only provides a flexible, affordable, and timely tool for COVID-19 tracking but also provides a proof of concept of how telemonitoring could be applicable to respiratory diseases monitoring in general.

Acknowledgments

This work was supported by European Research Council (ERC) Project 833296 (EAR). We thank everyone who volunteered their data.

XSL•FO
RenderX

Data Availability

The data were sensitive as voice sounds can be deanonymized. Anonymized data will be made available for academic research upon requests directed to the corresponding author. Institutions need to sign a data transfer agreement with the University of Cambridge to obtain the data. A copy of the data will be transferred to the institution requesting the data. We already have the data transfer agreement in place. Python code and parameters used for training of neural networks will be available on GitHub for reproducibility purposes.

Authors' Contributions

AF, CM, and PC designed the study. AH, AG, CS-B, DS, and JC designed and implemented the mobile app to collect the sample data. AG designed and implemented the server infrastructure. JH, TD, and TX selected the data for analysis. DS, TD, and TX developed the neural network models. TD conducted the experiments, performed the statistical analysis, wrote the main draft of the manuscript, and generated tables and figures. JH and TX cowrote the manuscript. All authors vouch for the data, analyses, and interpretations. All authors critically reviewed, contributed to the preparation of the manuscript, and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Details of model development and validation, including data collection, data selection, data augmentation, model architecture, and model training and evaluation.
[DOCX File , 25 KB - jmir_v24i6e37004_app1.docx]

Multimedia Appendix 2
Data statistics in terms of gender, age, and language.
[DOCX File , 574 KB - jmir_v24i6e37004_app2.docx]

Multimedia Appendix 3
Additional examples of strong disease progression predictions.
[DOCX File , 242 KB - jmir_v24i6e37004_app3.docx]

Multimedia Appendix 4
Additional examples of weak disease progression predictions.
[DOCX File , 419 KB - jmir_v24i6e37004_app4.docx]

Multimedia Appendix 5
Detailed analysis on the impact of sequence length for COVID-19 detection.
[DOCX File , 957 KB - jmir_v24i6e37004_app5.docx]

References

1. Vogels CBF, Brito AF, Wyllie AL, Fauver JR, Ott IM, Kalinich CC, et al. Analytical sensitivity and efficiency comparisons of SARS-CoV-2 RT-qPCR primer-probe sets. Nat Microbiol 2020 Oct;5(10):1299-1305. [doi: 10.1038/s41564-020-0761-6] [Medline: 32651556]
2. Cevik M, Kuppalli K, Kindrachuk J, Peiris M. Virology, transmission, and pathogenesis of SARS-CoV-2. BMJ 2020 Oct 23;371:m3862. [doi: 10.1136/bmj.m3862] [Medline: 33097561]
3. Fan L, Liu S. CT and COVID-19: Chinese experience and recommendations concerning detection, staging and follow-up. Eur Radiol 2020 May 06;30(9):5214-5216. [doi: 10.1007/s00330-020-06898-3]
4. Deshpande G, Batliner A, Schuller BW. AI-Based human audio processing for COVID-19: a comprehensive overview. Pattern Recognit 2022 Feb;122:108289. [doi: 10.1016/j.patcog.2021.108289] [Medline: 34483372]
5. Ates HC, Yetisen AK, Güder F, Dincer C. Wearable devices for the detection of COVID-19. Nat Electron 2021 Jan 25;4(1):13-14. [doi: 10.1038/s41928-020-00533-1]
6. Channa A, Popescu N, Skibinska J, Burget R. The rise of wearable devices during the COVID-19 pandemic: a systematic review. Sensors (Basel) 2021 Aug 28;21(17):5787 [FREE Full text] [doi: 10.3390/s21175787] [Medline: 34502679]
7. Barr PJ, Ryan J, Jacobson NC. Precision assessment of COVID-19 phenotypes using large-scale clinic visit audio recordings: harnessing the power of patient voice. J Med Internet Res 2021 Feb 19;23(2):e20545. [doi: 10.2196/20545]

XSL•FO
RenderX

8. Stasak B, Huang Z, Razavi S, Joachim D, Epps J. Automatic detection of COVID-19 based on short-duration acoustic smartphone speech analysis. J Healthc Inform Res 2021 Mar 11;5(2):201-217. [doi: 10.1007/s41666-020-00090-4] [Medline: 33723525]
9. Miranda ID, Diacon AH, Niesler NR. A comparative study of features for acoustic cough detection using deep architectures. 2019 Presented at: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); July 23-27, 2019; Berlin, Germany p. 2601-2605. [doi: 10.1109/embc.2019.8856412]
10. Al Ismail M, Deshmukh S, Singh R. Detection of COVID-19 through the analysis of vocal fold oscillations. 2021 Presented at: 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); June 6-11, 2021; Toronto, Ontario, Canada p. 1035-1039. [doi: 10.1109/icassp39728.2021.9414201]
11. Deshmukh S, Al Ismail M, Singh R. Interpreting glottal flow dynamics for detecting COVID-19 from voice. 2021 Presented at: 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); June 6-11, 2021; Toronto, Ontario, Canada p. 1055-1059. [doi: 10.1109/icassp39728.2021.9414530]
12. Deshpande G, Schuller BW. Audio, Speech, Language, and Signal Processing for COVID-19: A Comprehensive Overview. arXiv Preprint posted online November 29, 2020. [doi: 10.48550/arXiv.2011.14445]
13. Laguarta J, Hueto F, Subirana B. COVID-19 artificial intelligence diagnosis using only cough recordings. IEEE Open J Eng Med Biol 2020;1:275-281. [doi: 10.1109/ojemb.2020.3026928]
14. Imran A, Posokhova I, Qureshi HN, Masood U, Riaz MS, Ali K, et al. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. Inform Med Unlocked 2020;20:100378. [doi: 10.1016/j.imu.2020.100378] [Medline: 32839734]
15. Brown C, Chauhan J, Grammenos A, Han J, Hasthanasombat A, Spathis D, et al. Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data. arXiv Preprint posted online June 10, 2020. [doi: 10.48550/arXiv.2006.05919]
16. Pinkas G, Karny Y, Malachi A, Barkai G, Bachar G, Aharonson V. SARS-CoV-2 detection from voice. IEEE Open J Eng Med Biol 2020;1:268-274. [doi: 10.1109/ojemb.2020.3026468]
17. Han J, Brown C, Chauhan J, Grammenos A, Hasthanasombat A, Spathis D, et al. Exploring automatic COVID-19 diagnosis via voice and symptoms from crowdsourced Data. 2021 Presented at: 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); June 6-11, 2021; Toronto, Ontario, Canada p. 8328-8332. [doi: 10.1109/ICASSP39728.2021.9414576]
18. Coppock H, Gaskell A, Tzirakis P, Baird A, Jones L, Schuller B. End-to-end convolutional neural network enables COVID-19 detection from breath and cough audio: a pilot study. BMJ Innov 2021 Apr 16;7(2):356-362 [FREE Full text] [doi: 10.1136/bmjinnov-2021-000668] [Medline: 34192022]
19. Andreu-Perez J, Perez-Espinosa H, Timonet E, Kiani M, Giron-Perez M, Benitez-Trinidad AB, et al. A generic deep learning based cough analysis system from clinically validated samples for point-of-need Covid-19 test and severity levels. IEEE Trans Serv Comput 2021:1-1. [doi: 10.1109/TSC.2021.3061402]
20. Yu F, Yan L, Wang N, Yang S, Wang L, Tang Y, et al. Quantitative detection and viral load analysis of SARS-CoV-2 in infected patients. Clin Infect Dis 2020 Jul 28;71(15):793-798 [FREE Full text] [doi: 10.1093/cid/ciaa345] [Medline: 32221523]
21. Wu J, Li W, Shi X, Chen Z, Jiang B, Liu J, et al. Early antiviral treatment contributes to alleviate the severity and improve the prognosis of patients with novel coronavirus disease (COVID-19). J Intern Med 2020 Jul 20;288(1):128-138. [doi: 10.1111/joim.13063] [Medline: 32220033]
22. Voinsky I, Baristaite G, Gurwitz D. Effects of age and sex on recovery from COVID-19: analysis of 5769 Israeli patients. J Infect 2020 Aug;81(2):e102-e103 [FREE Full text] [doi: 10.1016/j.jinf.2020.05.026] [Medline: 32425274]
23. Lechien JR, Chiesa-Estomba CM, Place S, Van Laethem Y, Cabaraux P, Mat Q, COVID-19 Task Force of YO-IFOS. Clinical and epidemiological characteristics of 1420 European patients with mild-to-moderate coronavirus disease 2019. J Intern Med 2020 Sep;288(3):335-344 [FREE Full text] [doi: 10.1111/joim.13089] [Medline: 32352202]
24. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. Lancet Infect Dis 2020 Aug;20(8):911-919. [doi: 10.1016/s1473-3099(20)30287-5]
25. Siddiqi HK, Mehra MR. COVID-19 illness in native and immunosuppressed states: a clinical-therapeutic staging proposal. J Heart Lung Transplant 2020 May;39(5):405-407 [FREE Full text] [doi: 10.1016/j.healun.2020.03.012] [Medline: 32362390]
26. Chen J, Qi T, Liu L, Ling Y, Qian Z, Li T, et al. Clinical progression of patients with COVID-19 in Shanghai, China. J Infect 2020 May;80(5):e1-e6 [FREE Full text] [doi: 10.1016/j.jinf.2020.03.004] [Medline: 32171869]
27. University of Cambridge. COVID-19 Sounds App. URL: https://www.covid-19-sounds.org/en/ [accessed 2022-06-06]
28. Xia T, Spathis D, Brown C, Ch J, Grammenos A, Han J, et al. COVID-19 sounds: a large-scale audio dataset for digital respiratory screening. 2021 Presented at: Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2); December 6-14, 2021; Virtual-only.
29. Hershey S, Chaudhuri S, Ellis DPW, Gemmeke JF, Jansen A, Moore RC, et al. CNN architectures for large-scale audio classification. 2017 Presented at: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); June 19, 2017; New Orleans, LA p. 131-135. [doi: 10.1109/icassp.2017.7952132]

XSL•FO
RenderX

Abbreviations

AUROC: area under the receiver operating characteristic curve
CT: computed tomography
DTW: dynamic time warping
GRU: gated recurrent unit
RT-PCR: reverse transcription polymerase chain reaction

Edited by C Basch; submitted 04.02.22; peer-reviewed by R Eikelboom; comments to author 31.03.22; revised version received 14.04.22; accepted 18.04.22; published 21.06.22.

Please cite as:
Dang T, Han J, Xia T, Spathis D, Bondareva E, Siegele-Brown C, Chauhan J, Grammenos A, Hasthanasombat A, Floto RA, Cicuta P, Mascolo C
Exploring Longitudinal Cough, Breath, and Voice Data for COVID-19 Progression Prediction via Sequential Deep Learning: Model Development and Validation
J Med Internet Res 2022;24(6):e37004
URL: https://www.jmir.org/2022/6/e37004
doi:10.2196/37004
PMID:35653606

©Ting Dang, Jing Han, Tong Xia, Dimitris Spathis, Erika Bondareva, Chloë Siegele-Brown, Jagmohan Chauhan, Andreas Grammenos, Apinan Hasthanasombat, R Andres Floto, Pietro Cicuta, Cecilia Mascolo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.06.2022. 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 Fear of Infection and Sufficient Vaccine Reservation Information on Rapid COVID-19 Vaccination in Japan: Evidence From a Retrospective Twitter Analysis

Qian Niu[1*], MS; Junyu Liu[2*], ME; Masaya Kato[1], MS; Momoko Nagai-Tanima[1], PhD; Tomoki Aoyama[1], MD, PhD

[1]Department of Human Health Sciences, Graduate School of Medicine, Kyoto University, Kyoto, Japan

[2]Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University, Kyoto, Japan

[*]these authors contributed equally

Corresponding Author:
Momoko Nagai-Tanima, PhD
Department of Human Health Sciences
Graduate School of Medicine
Kyoto University
53, Kawahara-Cho, Shogoin
Sakyo-Ku
Kyoto, 606-8507
Japan
Phone: 81 075 751 3964
Email: tanima.momoko.8s@kyoto-u.ac.jp

Abstract

Background: The global public health and socioeconomic impacts of the COVID-19 pandemic have been substantial, rendering herd immunity by COVID-19 vaccination an important factor for protecting people and retrieving the economy. Among all the countries, Japan became one of the countries with the highest COVID-19 vaccination rates in several months, although vaccine confidence in Japan is the lowest worldwide.

Objective: We attempted to find the reasons for rapid COVID-19 vaccination in Japan given its lowest vaccine confidence levels worldwide, through Twitter analysis.

Methods: We downloaded COVID-19–related Japanese tweets from a large-scale public COVID-19 Twitter chatter data set within the timeline of February 1 and September 30, 2021. The daily number of vaccination cases was collected from the official website of the Prime Minister's Office of Japan. After preprocessing, we applied unigram and bigram token analysis and then calculated the cross-correlation and Pearson correlation coefficient (r) between the term frequency and daily vaccination cases. We then identified vaccine sentiments and emotions of tweets and used the topic modeling to look deeper into the dominant emotions.

Results: We selected 190,697 vaccine-related tweets after filtering. Through n-gram token analysis, we discovered the top unigrams and bigrams over the whole period. In all the combinations of the top 6 unigrams, tweets with both keywords "reserve" and "venue" showed the largest correlation with daily vaccination cases (r=0.912; P<.001). On sentiment analysis, negative sentiment overwhelmed positive sentiment, and fear was the dominant emotion across the period. For the latent Dirichlet allocation model on tweets with fear emotion, the two topics were identified as "infect" and "vaccine confidence." The expectation of the number of tweets generated from topic "infect" was larger than that generated from topic "vaccine confidence."

Conclusions: Our work indicates that awareness of the danger of COVID-19 might increase the willingness to get vaccinated. With a sufficient vaccine supply, effective delivery of vaccine reservation information may be an important factor for people to get vaccinated. We did not find evidence for increased vaccine confidence in Japan during the period of our study. We recommend policy makers to share accurate and prompt information about the infectious diseases and vaccination and to make efforts on smoother delivery of vaccine reservation information.

(J Med Internet Res 2022;24(6):e37466) doi:10.2196/37466

XSL•FO
RenderX

KEYWORDS

COVID-19; vaccine hesitancy; Japan; social media; text mining

Introduction

COVID-19 has spread worldwide since its first case in December 2019 and has become a public health emergency of international concern [1]. Until September 30, 2021, Japan experienced 5 waves of the COVID-19 pandemic [2]. The surge of COVID-19 in Japan occurred during the Tokyo Olympics, bringing the cumulative number of COVID-19 cases to 1,556,998 when the Games finished. However, with the lifting of the fourth national state of emergency on September 30, 2021, the pandemic was effectively contained nationwide, and the number of new confirmed cases abruptly decreased. The high vaccination rate in Japan was considered to have caused a decline in the community infections during the fifth wave [3].

A high vaccination rate is thought to be promoted by high vaccine confidence [4]. According to the US Centers of Disease Control and Prevention, "Vaccine confidence is the belief that vaccines work, are safe, and are part of a trustworthy medical system" [5]. A global survey that did not include Japan showed that the potential acceptance of a COVID-19 vaccine largely varied among countries [6]. Japan ranks among the countries with the lowest vaccine confidence worldwide according to a survey in 2020 [7]. Another survey conducted before large-scale vaccination in Japan indicated that Japan ranked last with regard to confidence in COVID-19 vaccines among 15 countries [8]. Gordon and Reich [9] explained the historical reasons for low vaccine confidence in Japan. Kunitoki et al [10] proposed that barriers to vaccine access and use mainly result from effective public communication and called for rebuilding vaccine confidence in Japan.

However, Japan's speed of vaccination has been impressive since large-scale vaccination was opened up (May 24, 2021). Japan's first-dose vaccination rate was approximately 6.8% by June 1, 2021, and over 70% of the population accepted at least one dose until September 30, 2021 [11]. Notably, vaccination was not mandatory and was administered only with the recipient's consent. A survey of multiple countries reported the coexistence of a high level of uncertainty about the safety of COVID-19 vaccines and a high willingness to get vaccinated [12], which indicates that Japan may not be a special case. The reason for the contradiction between the rapid growth in the COVID-19 vaccination rate and low vaccine confidence in Japan is worth studying and maybe instructive for propelling worldwide vaccination against infectious diseases.

Twitter is a widespread social media platform that has attracted the increasing attention of public health researchers because of its advantages of large amounts, real-time availability, and ease of public searching and access [13]. With a large amount of real-time COVID-19–related posts, Twitter has been widely used for public opinion mining toward COVID-19 during the pandemic, providing policy makers with substantiated evidence [12,14,15]. Lyu et al [14] reported the trend of topics and sentiments of English tweets for approximately 11 months since the World Health Organization declared COVID-19 a pandemic. Yousefinaghani et al [12] reported the dominance of positive sentiments and more vaccine objection and hesitancy than vaccine interest. Huangfu et al [15] reported the results of topic modeling and sentiment analysis of tweets between December 8, 2020, and April 8, 2021. Eibensteiner et al [16] reported willingness to vaccinate despite the safety concerns of vaccines, according to a survey on a Twitter poll. Besides, Twitter is the most popular social media platform in Japan [17], owning 58.2 million users as of October 2021 [18], making Twitter analysis more powerful for COVID-19 research in Japan. A Twitter analysis by Niu et al [19] reported that the Japanese public's negative sentiment overwhelmed the positive sentiment toward the COVID-19 vaccine before and at the beginning of the large-scale vaccination campaign.

This retrospective study aimed to identify public sentiments and concerns associated with rapid COVID-19 vaccination in Japan. We hypothesized that the increase in vaccination rates might be due to subjective factors including increased public confidence in vaccines (S1) and fear of infection (S2), and objective factors including adequate vaccine supply (O1) and effective delivery reservation–related vaccine information (O2). To test these hypotheses, we collected vaccine-related tweets posted between February 1 and September 30, 2021. Then, we preprocessed the collected tweets and conducted a unigram token analysis, sentiment analysis, and topic modeling.

Methods

Overview

In previous works of large-scale Twitter analyses, after preprocessing, there are mainly 4 types of natural language processing (NLP) methods: n-gram token analysis [12,15,20,21], sentiment analysis [12,14,15,20-25], topic modeling [12,14,15,20,22-25], and geographical analysis [22,24]. The geographical analysis is less important in our work because the range of our research is a whole country instead of subareas. In this work, we followed previous works in applying n-gram token analysis, sentiment analysis, and topic modeling. Code in this work will be shared on the web [26].

Data Collection and Preprocessing

The data used in this study were obtained from a large-scale public COVID-19 Twitter chatter data set [27] updated by the Georgia State University's Panacea Lab. The data set provided the IDs, posting time, and the languages of all the tweets were provided in the data set. We downloaded COVID-19–related Japanese tweets between February 1, 2021, the month the first person was vaccinated, and September 30, 2021, when the first-dose vaccination rate exceeded 70%. In addition, data on the number of vaccination cases were collected from the official website of the Prime Minister's Office of Japan (PMOJ) [28].

The downloaded tweets were then cleaned and processed. Retweets were filtered using the Python package tweepy. Tweets that included no keywords related to vaccines were deleted. The keywords used in the filtering are listed in Multimedia Appendix

XSL•FO
RenderX

1. It is worth noting that the three vaccine brands (Pfizer, Moderna, and AstraZeneca) that were approved by the Japanese government were included in the keywords. Other vaccine brands were excluded because we attempted to focus more on the brands adopted in the vaccination process. Frequent misspellings (eg, "Modelna") was also included in the keywords. Weblinks, special characters, emojis, and "amp" (ampersands) were removed, and all full-width English characters were converted to half-width, lowercase characters.

For convenience, all Japanese words in our results were directly presented in English translations. The English-Japanese translation table is provided in Multimedia Appendix 1. In order to minimize the influence of difference between languages, all the translations in our results were carried out as the last step by directly replacing the Japanese words in the graphs with corresponding English words; therefore, they would not influence the statistical results.

Unigram and Bigram Token Analysis

Tokenization is necessary before many other NLP tasks, especially for many non-Latin languages, such as Japanese. We removed the predefined English and Japanese stop words in the Python packages NLTK [29] and SpaCy [30] and tokenized all collected vaccine-related tweets using the Python package SpaCy into unigrams or bigrams for statistical analysis, as reported by Kwok et al [27]. We sorted the unigram tokens or bigram tokens in descending order of term frequency over the entire period. Similar to Liu et al [24], we used the pruned exact linear time (PELT) algorithm [31] to find the first change point of the term frequency. Unigrams before the first change point were regarded as top unigrams, and the term frequencies of the unigrams after the change point were significantly lower than those of the top unigrams. Similar processes were carried out for bigrams. To eliminate the difference in the number of days between months, the monthly term frequency was defined by dividing the total term frequency by the number of days each month for each top unigram or bigram.

Correlation coefficients were widely used in social media analysis. In Google Trends analysis, correlations were calculated between reported cases of infectious disease and the trends of search for relevant keywords. In Twitter analyses, correlations between the daily cases of infection or death and the number of related tweets or sentiment scores, were also investigated [24,25]. In this work, correlation analyses were adopted to find out the factors from the top unigrams that are most related to the COVID-19 vaccination campaign. We first calculated the cross-correlations between the number of tweets containing the top unigrams or bigrams and the vaccination cases and then observed the time lags when maximum cross-correlation appeared for each unigram and bigram. Pearson correlation coefficients (r) between top unigrams or bigrams and the vaccination cases were also calculated.

Sentiment Analysis

After n-gram analysis, sentiment analyses were often used to explore the real-time public attitudes in social media analysis related to COVID-19 vaccination, which may reflect the acceptance of COVID-19 vaccines and related policies [12,14,20,22,24]. The trend of negative sentiments may provide potential evidence for vaccine hesitancy [23]. In this work, sentiment analysis was applied to all vaccine-related tweets. Cloud services were used in this study because there were no reliable public models for sentiment analysis in the Japanese language. We selected Amazon Web Services (AWS) for consistency with previous work [29]. The tweets were divided into positive, negative, neutral, or a mixture of positive and negative tweets using the AWS. Fine-grained emotions were also explored using the Japanese version of the NRC Emotion Lexicon [32]. The NRC Emotion Lexicon is a dictionary of words and their associated scores for eight emotions: anticipation, trust, joy, surprise, anger, disgust, fear, and sadness. The positive and negative tweets were tokenized, and the degree of valence (DOV) for the eight emotions was calculated by adding up the scores for the unigrams that appeared in the NRC Emotion Lexicon. Finally, we calculated the daily average DOV by dividing the number of positive and negative tweets on that day to show the trend of each emotion.

Topic Modeling

Topic modeling were applied to identify fine-grained information from tweets of different sentiments [12,15,24]. Based on the sentiment analysis results, we summarized the topics to look deeper into the dominant emotion in the tweets. Latent Dirichlet allocation (LDA) is often used in tweet topic modeling studies [14,15,20,22,23]. In this study, LDA regards tweets as being generated from different topics, and each topic generates tweets with a Dirichlet distribution. A Python package scikit-learn was used to determine the best number of topics. Log likelihood was adopted as the metric for selection, and 5-fold cross correlation was applied to avoid overfitting. As shown in Multimedia Appendix 1, we chose 2 as the number of topics for LDA modeling, which showed the highest log likelihood score. We used scikit-learn for LDA topic modeling and displayed the top 10 keywords and their weights related to each topic. The weights were the pseudocounts of the keywords in a topic. The themes of topics were summarized from the top 10 keywords by 3 volunteers. The volunteers were first asked to work out the themes of the topics independently, and then they had a meeting to finally reach an agreement on the themes.

We then checked the trends of tweets related to different topics. Defining the i-th tweet in all collected tweets as d_i, and the j-th topic of the LDA model as t_j, the probability of a tweet d_i coming from t_j was calculated using the fitted LDA model as p_{ij}. For tweets posted each day, the expectation of the number of tweets generated from topic j was calculated by summing p_{ij} on that day. The ratio between the expected number of tweets generated from each topic was also plotted to show the trend of public attention under dominant emotion.

Ethics Approval

This study used publicly available and accessible tweets collected by Georgia State University's Panacea Lab, allowing free download. We assert that our analysis is compliant with Twitter's usage policy in aggregate form without identifying specific individuals who published the Twitter posts. Furthermore, the number of vaccination cases downloaded from

XSL•FO
RenderX

the PMOJ are open government data. Therefore, the activities described do not meet the requirements of human subject research and did not require review by an institutional review board.

Results

Data Summary

We downloaded 979,636 Japanese tweets posted between February 1 and September 30, 2021, according to the ID and region information in the data set. After filtering, 190,697 vaccine-related tweets were selected. As a result, the total number of vaccine-related tweets increased from 14,758 tweets in February to 34,692 in August and then decreased to 27,824 in September.

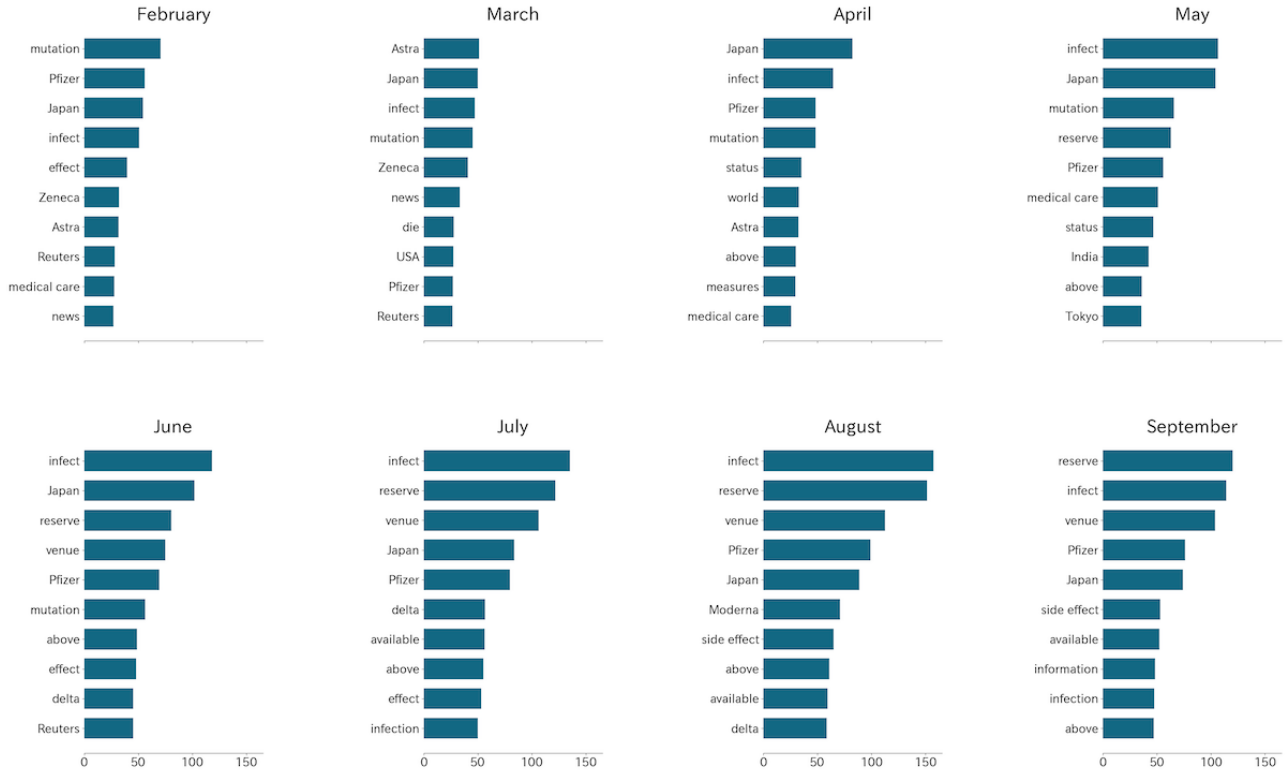
Unigram and Bigram Token Analysis

The change point of unigram term frequencies detected by the PELT algorithm was 6, and the top 6 unigrams were Japanese words for “infection,” “Japan,” “reserve,” “Pfizer,” “venue,” and “mutation.” The unigram “side effects,” related to the safety of vaccines, ranked eighth overall. The unigrams “infect,” “reserve,” and “venue” gradually ranked in the top 3 from February to September, as shown in Figure 1.

The change point of bigram term frequencies detected by the PELT algorithm was 5, and the top 5 were Japanese bigrams for “Astra + Zeneca,” “reserve + available,” “article + Reuters,” “venue + reserve,” and “medical-care + workers.” The bigrams “reserve + available” and “venue + reserve” ranked in the top-2 from June to September, and the ranking of “Astra + Zeneca” decreased since May, as shown in Figure 2.

Regarding correlation analysis of unigrams, the time lags for “reserve” and “venue” were 0, and the vaccination cases led the number of tweets containing “infection” for 5 days. After calculating r between the daily number of tweets containing each top unigram and vaccination cases, significant r values (P<.001) were obtained for all unigrams except “mutation.” The largest r value for the daily vaccination cases was from unigrams “infection” (r=0.746), “reserve” (r=0.829), and “venue” (r=0.908). We then checked the daily number of tweets containing all the combinations of the 3 unigrams showing a strong correlation and found the highest r value (r=0.912; P<.001) for tweets containing both “reserve” and “venue.” By randomly selecting 5 days and checking the source of all the tweets on those days, we found that the 95% CI of tweets containing both “reserve” and “venue” posted by official accounts or mainstream media was 96.0%-100%. The trend of tweets containing both unigrams “reserve” and “venue” compared with the daily vaccination cases is shown in Figure 3.

Figure 1. Translation of the top 10 unigrams of each month. The lengths of the bars represent the monthly term frequencies in tweets of each month.



XSL•FO
RenderX

Figure 2. Translation of the top 10 bigrams of each month. The lengths of the bars represent the monthly term frequencies in tweets of each month.

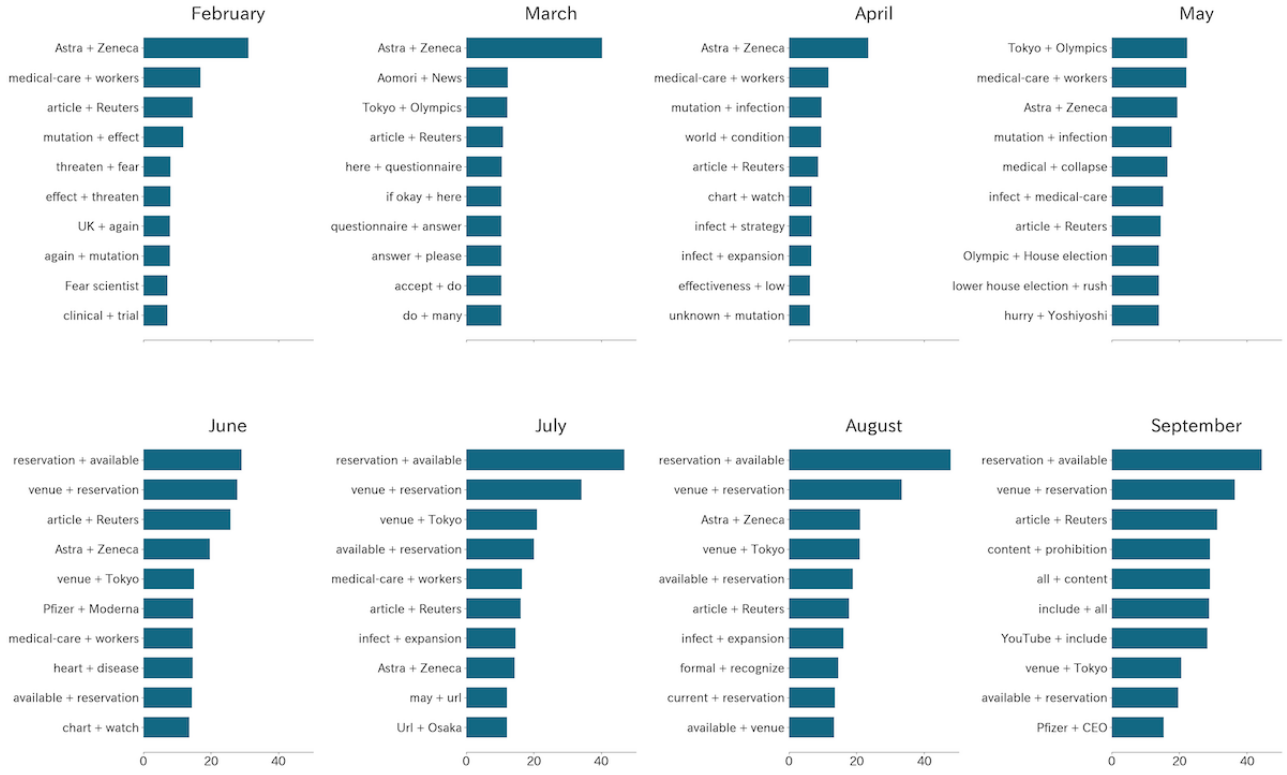
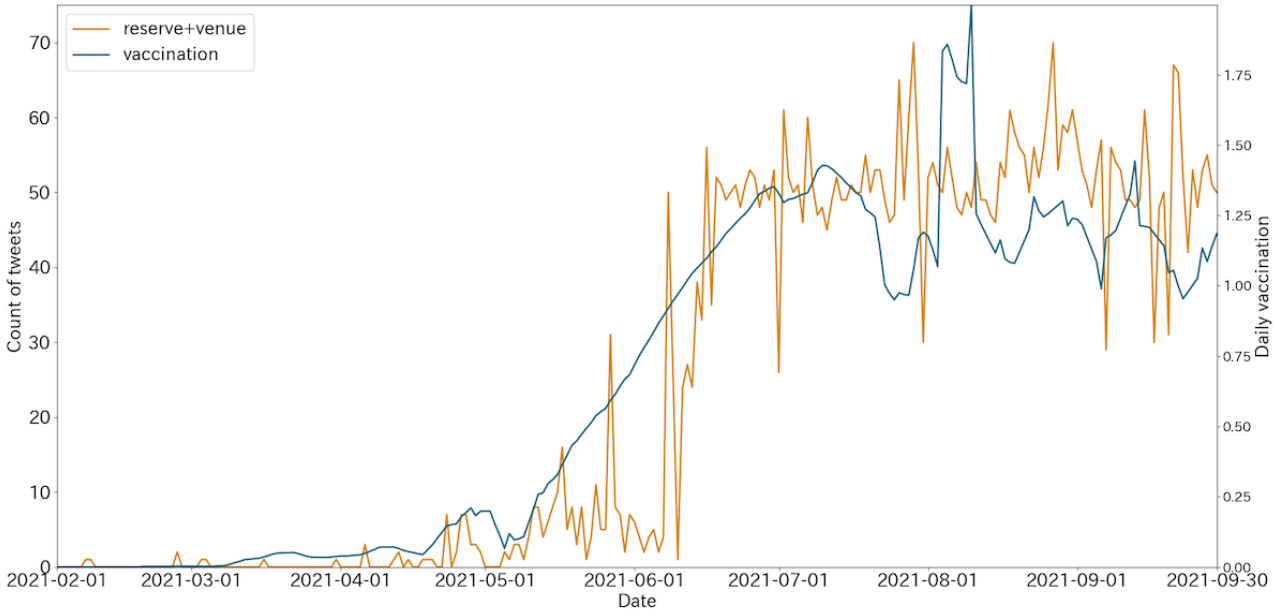


Figure 3. Trend of tweets containing both unigrams “reserve” and “venue,” and the curve for daily first-dose vaccination cases.



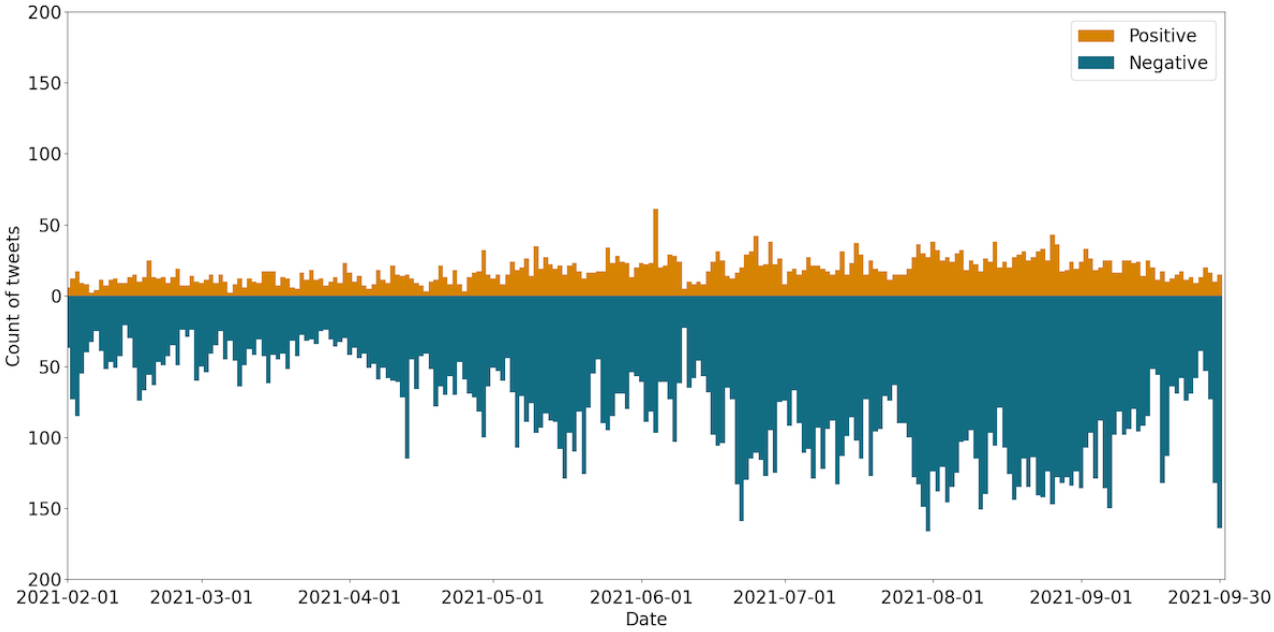
As for bigrams, the bigram “venue + reserve” overlapped with the unigram analysis and was excluded from this part. The time lags for bigrams “reserve + available” and “article + Reuters” were 0, and vaccination cases led “Astra + Zeneca” and “medical-care + workers” for 116 and 63 days, respectively. The bigrams “reserve + available” and “article + Reuters” had the highest cross-correlations than the others. The bigrams “Astra + Zeneca” (r=–0.331), “reserve + available” (r=0.908), and “article + Reuters” (r=0.229) showed significant correlations (P<.001) except for “medical-care + workers” (r=–0.055). On manual evaluation by 3 volunteers, we found that 95.4% of the tweets that contain the bigrams “reserve + available” were the same as those of containing the combination of unigrams “venue” and “reserve.”

Sentiment Analysis

For all tweets, 4453 (2.3%) were positive, 19,340 (10.1%) were negative, 164,687 (86.4%) were neutral, and 2217 (1.2%) were mixed positive and negative sentiments. A comparison between the daily numbers of tweets marked as positive and negative is shown in Figure 4. Negative sentiments overwhelmed positive sentiments for all days.

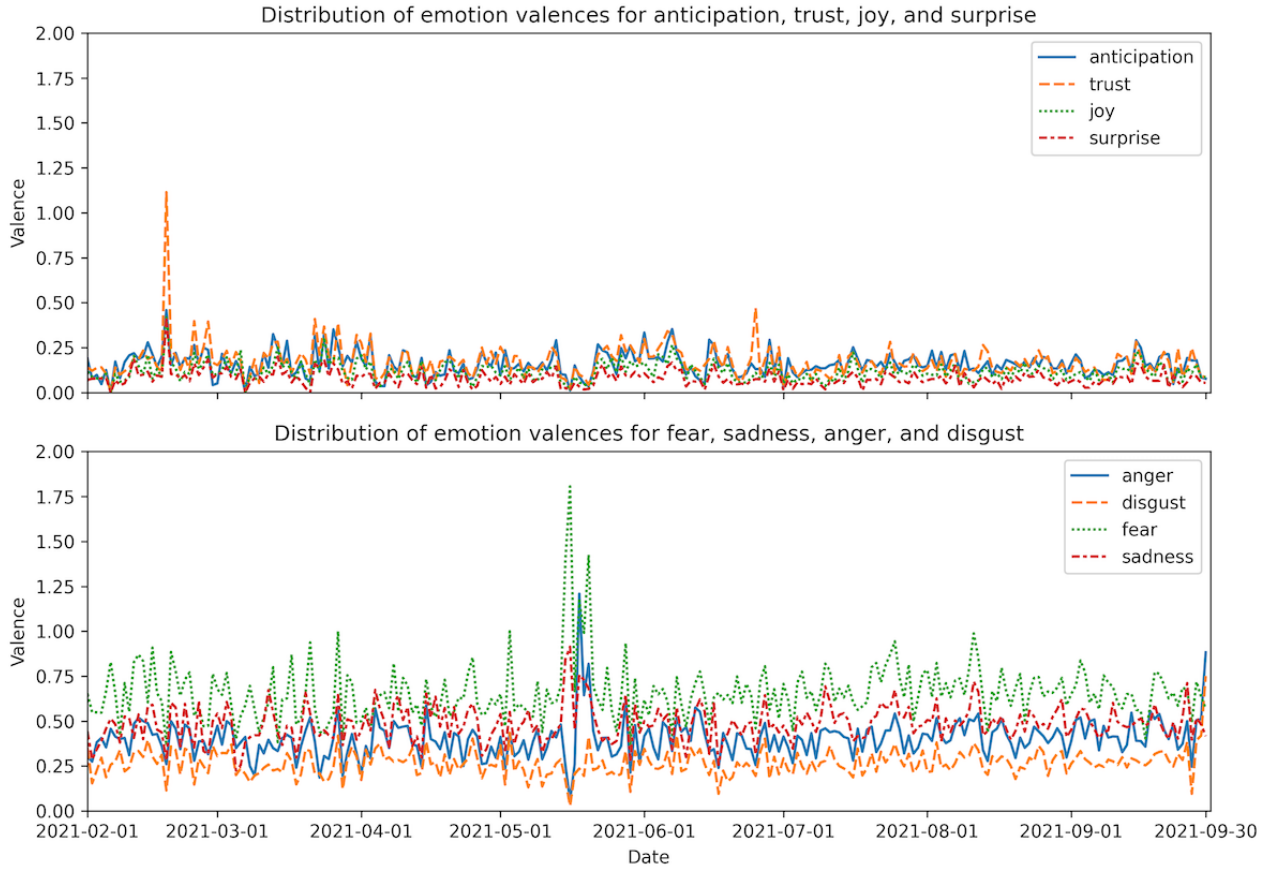
XSL•FO
RenderX

Figure 4. Comparison between the daily number of tweets marked positive (orange) and negative (green).



The DOVs for the 8 emotions are shown in Figure 5. The daily average DOV of anger (0.404), disgust (0.268), fear (0.659), sadness (0.486), overwhelmed anticipation (0.163), trust (0.173), joy (0.118), and surprise (0.081). Fear was the dominant emotion during this period. Here, we defined the peaks of emotion as larger than 3 times the daily average DOV for that emotion. Trust peaked (1.114) on February 18, 2021. From May 13 to 18, 2021, there were several peaks of fear.

Figure 5. Daily average degree of valence of 8 emotions in the vaccine-related tweets.



XSL•FO
RenderX

Topic Modeling

The top 10 keywords for each LDA topic are shown in Figure 6. The theme of topic 1 is "infect," and that of topic 2 is "vaccine confidence." It is also noticeable that the weight of "infect" (14,895) in topic-1 was over 3 times that of the second keyword "Japan" (4359), but the weight of "Pfizer" (4348) in topic 2 was only 15.5% larger than the second keyword "die" (3763).

The ratio between the expectation of the number of "infect"-related tweets and "vaccine confidence"–related tweets is shown in Figure 7. The total expectation of the number of tweets generated from topic 1 ("infect," n=30,288) is larger than that generated from topic 2 ("vaccine confidence," n=27,572), and the mean ratio between the expectation of the daily number of tweets generated from topics 1 and 2 is significantly larger than 1 (P<.01). On 68.2% of days, the expectation of the number of tweets generated from "infect" was larger than that generated from "vaccine confidence."

Figure 6. Top 10 keywords of 2 topics by latent Dirichlet allocation modeling. The bars represent the weights, which can be regarded as the pseudocounts of the keywords in each topic.

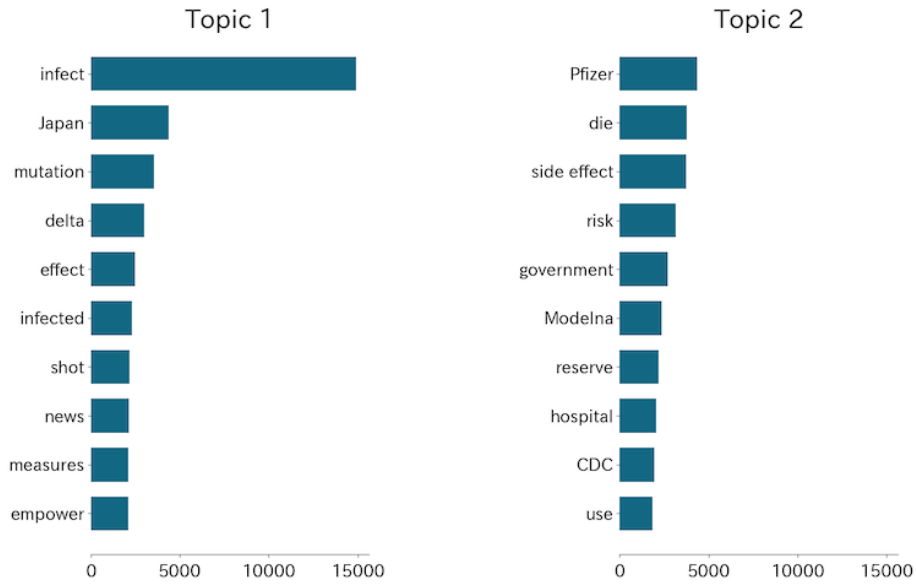
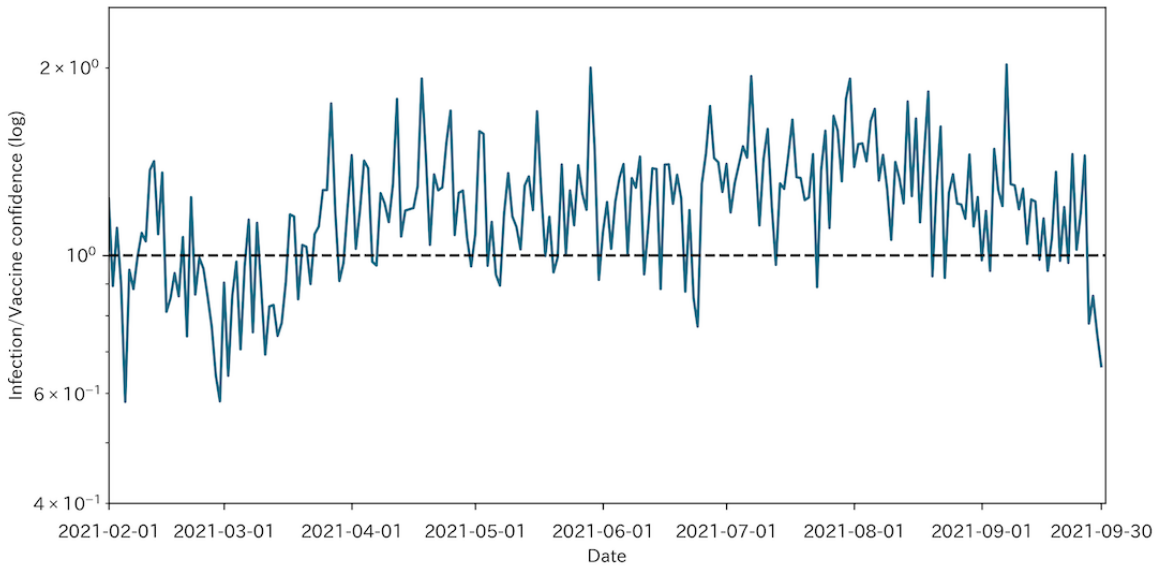


Figure 7. Ratio between the expectation of the number of "infection"-related tweets and "vaccine confidence"–related tweets.



Discussion

Principal Findings

A high vaccination rate is thought to be promoted by high vaccine confidence [16,33-36], but Japan achieved a high vaccination rate in several months, with the lowest vaccine confidence in the world. This retrospective study aimed to determine the reasons for the fast vaccination process in Japan, which may be instructive for propelling worldwide vaccination for infectious diseases. Based on previous studies [16,34-37], we hypothesized that subjective factors, including increased

XSL•FO
RenderX

vaccine confidence (S1) and fear of infection (S2), and objective factors including adequate vaccine supply (O1) and effective delivery of reservation-related vaccine information (O2). Our results indicate that hypotheses S2 and O2 might have driven the public to be vaccinated. No evidence supporting hypothesis S1 was found in our results. Evidence for hypothesis O1 can be found in the history of vaccine supply on the official website of the PMOJ (Prime Minister of Japan and his Cabinet) and is not discussed in this paper.

Several results support hypothesis S2. In the unigram token analysis shown in Figure 1, the keyword "infect" ranked among the top 3, except in February, and ranked first from May to August during Japan's fourth and fifth wave infections. The keywords "venue" and "reserve" also ranked up from May. No keywords related to increased vaccine confidence were found. The sentiment analysis shown in Figure 4 showed that negative sentiment overwhelmed positive sentiments, consistent with the results of Chen et al [35] that Japan showed dissatisfaction compared with neighboring countries. Combined with our result that "infect" was the top keyword and "side effect" ranked eighth in the unigram token analysis, our results support that the Japanese public was more concerned about infection than the side effects of COVID-19 vaccines.

More evidence for hypothesis S2 was obtained from the topic modeling results. From the keywords of topic 1 ("infect"), we can see that the public was concerned about the infection and death rate. The mutated virus and empowered cases also led to fear. Willis et al [37] found that less fear of infection may lead to a lower willingness to be vaccinated, which is complementary to our results. From the keywords of topic 2 ("vaccine confidence"), we can see that the side effects of the vaccines were the most concerning, but the following keywords were related to the effectiveness of vaccines on the mutated virus, reservation of vaccines, and medical care conditions. Previous surveys in different countries have indicated that fear of vaccine safety is the key factor for low vaccine acceptance [38,39]. Furthermore, the "side effect" weight in topic 2 was much less than that of "infect" in topic 1. The top keywords in the two topics indicated that people were more concerned about COVID-19 rather than the side effects of vaccines. Bendau et al [40] reported a significant positive correlation between fears of infection and vaccine acceptance and a significant negative correlation between fear of vaccine safety and vaccine acceptance. Therefore, it is important to distinguish the mainstream fear emotion to determine the reason for the high vaccination rate. Figure 7 provides details about the ratio between the expected number of tweets related to "infect" and "vaccine confidence." In most cases, the ratio was larger than 1, indicating that the public was more concerned about infection rather than the safety and effectiveness of vaccines. Higher ratios were observed in April and from July to end-September, which were periods of Japan's fourth and fifth waves of infection. There was also a relatively long period of less than one ratio from mid-February to mid-March, which was the period when the vaccines were less effective against the mutated virus (February 10), severe side effects of the AstraZeneca vaccine were observed (March 12), and several side effects were observed in Japan (February 21, March 7, and March 10). However, the ratio soon increased because of the fourth wave of infections. This example also proved that fear of infection overcame the vaccine safety concern.

We also provide evidence of a strong relationship between vaccination and hypothesis O2. Bigram analysis in Figure 2 showed that "reservation + available" ranked first since June, shortly after large-scale vaccination started, which might reflect the strong concerns about vaccine reservation by the public. Unigram token analysis in Figure 3 showed that tweets including the keywords "reserve" and "venue" were significantly highly correlated (r>0.9; P<.01) with the daily number of vaccination cases in Japan, and most of them were from government official accounts. The bigram "reservation + available" also showed a high correlation (r>0.9; P<.01) with the daily vaccination cases. Because reservation information should always lead to the actual vaccination, this result indicated that in addition to sufficient vaccine supply, reservation information delivery might also be important in large-scale vaccination. Furthermore, the time lag for the maximum cross-correlation was 0, which may indicate the efficiency of the reservation information posted on Twitter. Our results were consistent with Fu's [41] finding that inflexible information systems for vaccine reservation can impair immunization services in the community.

We did not find any evidence for hypothesis S1. Macaraan reported a shift from hesitancy to confidence toward the COVID-19 vaccination program among Filipinos [36]. Okubo [42] reported a shift from hesitancy to confidence but also admitted that the shift might come from the differences in the survey metrics in previous studies [43]. Following these studies, we looked for a similar shift in sentiment or emotions from negative to positive, but negative sentiments overwhelmed positive sentiments as shown in Figure 4, and fear dominates all the emotions in Figure 5. The positive emotions "anticipation," "trust," and "joy" did not increase during the entire period. These two results made it difficult to conclude increased vaccine confidence.

Our results were partially related to the 5 C model (confidence, competence, convenience, calculation, and collective responsibility) measuring vaccine hesitancy [36,44]. Confidence and complacency are two subjective measures that are directly related to individuals. In our work, the LDA theme "vaccine confidence" belonged to "confidence," and "fear of infection" belonged to "complacency." In Japan, fear of infection may drive a high vaccination rate. The delivery of reservation information may be an extension to "convenience," which was previously defined as "physical availability, affordability and willingness-to-pay, geographical accessibility, ability to understand (language and health literacy), and appeal of immunization service affect uptake" [45]. Our work indicates that information about vaccination reservations should also be considered for the convenience of vaccination.

Limitations

We admit that our research might have some potential limitations: (1) the imbalance of the demographics of Twitter users in Japan [46] may cause bias in the results; (2) the status of the user on a certain day (at home or not, other events on that day, etc) may also bias the data set [47]; (3) owing to the lack

XSL•FO
RenderX

of a reliable public model for sentiment analysis in the Japanese language, the cloud service AWS was used for sentiment analysis; (4) filtering keywords may include irrelevant or missing related tweets; (5) antivaccine tweets, especially rumors, were not distinguished or analyzed separately in this study. However, feature works can be combined with classical surveys to train the sentiment analysis model and model to distinguish rumors from tweets to overcome these limitations.

Conclusions

This retrospective study aimed to determine the reasons for the fast vaccination process in Japan, which might be instructive for propelling worldwide vaccination toward infectious diseases. In conclusion, our work indicated that awareness of the danger of COVID-19 increased the willingness to be vaccinated; with a sufficient supply of vaccines, effective reservation information delivery might provide more opportunities for people to be vaccinated. Models measuring vaccine hesitancy might also need to add efficiency in delivering reservation information as a metric. Based on our findings, we recommend public health policy makers and the government to share accurate and prompt information about the infectious diseases and vaccination. Furthermore, efforts on tied cooperation among multilevel relevant organizations and new media operations may help achieve smoother delivery of vaccine reservation information.

Acknowledgments

This work was supported by the JST SPRING (grant number JPMJSP2110).

Authors' Contributions

QN and JL performed analyses and drafted the manuscript. All authors conceived the study, interpreted the results, and revised the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The keywords used for the selection of vaccine-related tweets and the corresponding translation; English translations used in our paper and the corresponding original Japanese words; Mean Log likelihood scores for different LDA topic numbers using five-fold cross-validation.
[DOCX File , 29 KB - jmir_v24i6e37466_app1.docx]

References

1. Romer D, Jamieson KH. Conspiracy theories as barriers to controlling the spread of COVID-19 in the U.S. Soc Sci Med 2020 Oct;263:113356 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113356] [Medline: 32967786]
2. Experts warn of 6th wave as COVID cases decrease in Japan but rise overseas. The Mainichi. 2021 Oct 30. URL: https://mainichi.jp/english/articles/20211030/p2a/00m/0na/020000c [accessed 2022-02-22]
3. Fifth wave had fewer big clusters possibly due to vaccinations. The Asahi Shimbun. 2021 Oct 15. URL: https://www.asahi.com/ajw/articles/14461519 [accessed 2022-02-22]
4. Larson H. Vaccine confidence and public trust as drivers of vaccine failure. Int J Infect Dis 2014 Apr;21:50. [doi: 10.1016/j.ijid.2014.03.522]
5. What Is Vaccine Confidence? Centers for Disease Control and Prevention. URL: https://www.cdc.gov/vaccines/covid-19/vaccinate-with-confidence/building-trust.html [accessed 2022-02-22]
6. Lazarus JV, Ratzan SC, Palayew A, Gostin LO, Larson HJ, Rabin K, et al. A global survey of potential acceptance of a COVID-19 vaccine. Nat Med 2021 Feb 20;27(2):225-228 [FREE Full text] [doi: 10.1038/s41591-020-1124-9] [Medline: 33082575]
7. de Figueiredo A, Simas C, Karafillakis E, Paterson P, Larson HJ. Mapping global trends in vaccine confidence and investigating barriers to vaccine uptake: a large-scale retrospective temporal modelling study. The Lancet 2020 Sep;396(10255):898-908. [doi: 10.1016/s0140-6736(20)31558-0]
8. Mahase E. Covid-19: UK has highest vaccine confidence and Japan and South Korea the lowest, survey finds. BMJ 2021 Jun 04;373:n1439. [doi: 10.1136/bmj.n1439] [Medline: 34088703]
9. Gordon A, Reich MR. The Puzzle of Vaccine Hesitancy in Japan. J Jpn Stud 2021;47(2):411-436. [doi: 10.1353/jjs.2021.0047]
10. Kunitoki K, Funato M, Mitsunami M, Kinoshita T, Reich MR. Access to HPV vaccination in Japan: Increasing social trust to regain vaccine confidence. Vaccine 2021 Oct 01;39(41):6104-6110 [FREE Full text] [doi: 10.1016/j.vaccine.2021.08.085] [Medline: 34507858]
11. Ritchie H, Mathieu E, Rodés-Guirao L, Appel C, Giattino C, Ortiz-Ospina E. Coronavirus Pandemic (COVID-19). Our World in Data. URL: https://ourworldindata.org/covid-vaccinations [accessed 2022-01-22]

XSL•FO
RenderX

12. Yousefinaghani S, Dara R, Mubareka S, Papadopoulos A, Sharif S. An analysis of COVID-19 vaccine sentiments and opinions on Twitter. Int J Infect Dis 2021 Jul;108:256-262 [FREE Full text] [doi: 10.1016/j.ijid.2021.05.059] [Medline: 34052407]
13. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a Tool for Health Research: A Systematic Review. Am J Public Health 2017 Jan;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
14. Lyu JC, Han EL, Luli GK. COVID-19 Vaccine-Related Discussion on Twitter: Topic Modeling and Sentiment Analysis. J Med Internet Res 2021 Jun 29;23(6):e24435 [FREE Full text] [doi: 10.2196/24435] [Medline: 34115608]
15. Huangfu L, Mo Y, Zhang P, Zeng DD, He S. COVID-19 Vaccine Tweets After Vaccine Rollout: Sentiment-Based Topic Modeling. J Med Internet Res 2022 Feb 08;24(2):e31726 [FREE Full text] [doi: 10.2196/31726] [Medline: 34783665]
16. Eibensteiner F, Ritschl V, Nawaz FA, Fazel SS, Tsagkaris C, Kulnik ST, et al. People's Willingness to Vaccinate Against COVID-19 Despite Their Safety Concerns: Twitter Poll Analysis. J Med Internet Res 2021 Apr 29;23(4):e28973 [FREE Full text] [doi: 10.2196/28973] [Medline: 33872185]
17. Social Media Stats Japan. StatCounter. URL: https://gs.statcounter.com/social-media-stats/all/japan [accessed 2022-01-22]
18. Leading countries based on number of Twitter users as of January 2022. Statista. URL: https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/ [accessed 2022-01-22]
19. Niu Q, Liu J, Kato M, Shinohara Y, Matsumura N, Aoyama T, et al. Public Opinion and Sentiment Before and at the Beginning of COVID-19 Vaccinations in Japan: Twitter Analysis. JMIR Infodemiology 2022;2(1):e32335 [FREE Full text] [doi: 10.2196/32335] [Medline: 35578643]
20. 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]
21. Yang X, Sornlertlamvanich V. Public Perception of COVID-19 Vaccine by Tweet Sentiment Analysis. 2021 Presented at: 2021 International Electronics Symposium (IES); September 29-30, 2021; Surabaya. [doi: 10.1109/ies53407.2021.9594036]
22. Xie Z, Wang X, Jiang Y, Chen Y, Huang S, Ma H, et al. Public Perception of COVID-19 Vaccines on Twitter in the United States. medRxiv Preprint posted online October 18, 2021. [FREE Full text] [doi: 10.1101/2021.10.16.21265097] [Medline: 34704100]
23. Monselise M, Chang C, Ferreira G, Yang R, Yang CC. Topics and Sentiments of Public Concerns Regarding COVID-19 Vaccines: Social Media Trend Analysis. J Med Internet Res 2021 Oct 21;23(10):e30765 [FREE Full text] [doi: 10.2196/30765] [Medline: 34581682]
24. Liu S, Liu J. Public attitudes toward COVID-19 vaccines on English-language Twitter: A sentiment analysis. Vaccine 2021 Sep 15;39(39):5499-5505 [FREE Full text] [doi: 10.1016/j.vaccine.2021.08.058] [Medline: 34452774]
25. Shim J, Ryu K, Lee SH, Cho E, Lee YJ, Ahn JH. Text Mining Approaches to Analyze Public Sentiment Changes Regarding COVID-19 Vaccines on Social Media in Korea. Int J Environ Res Public Health 2021 Jun 18;18(12):6549 [FREE Full text] [doi: 10.3390/ijerph18126549] [Medline: 34207016]
26. COVID_fear. GitHub. URL: https://github.com/juniorliu95/COVID_fear [accessed 2022-06-06]
27. Banda JM, Tekumalla R, Wang G, Yu J, Liu T, Ding Y, et al. A large-scale COVID-19 Twitter chatter dataset for open scientific research -- an international collaboration. arXiv Preprint posted online April 7, 2020. [Medline: 32550247]
28. Prime Minister of Japan and His Cabinet. URL: https://japan.kantei.go.jp/ [accessed 2022-02-22]
29. Bird S, Klein E, Loper E. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. Newton, MA: O'Reilly Media, Inc; 2009.
30. Engebreth G. Bug Fixes. In: PHP 8 Revealed. Berkeley, CA: Apress; 2021:87-99.
31. Dorcas Wambui G. The Power of the Pruned Exact Linear Time(PELT) Test in Multiple Changepoint Detection. AJTAS 2015;4(6):581. [doi: 10.11648/j.ajtas.20150406.30]
32. Mohammad S, Turney P. Crowdsourcing a word-emotion association lexicon. Comput Intell 2013;29:465. [doi: 10.1111/j.1467-8640.2012.00460.x]
33. Deveaud R, SanJuan E, Bellot P. Accurate and effective latent concept modeling for ad hoc information retrieval. Document numérique 2014 Apr 30;17(1):61-84. [doi: 10.3166/dn.17.1.61-84]
34. Mori H, Naito T. A rapid increase in the COVID-19 vaccination rate during the Olympic and Paralympic Games 2021 in Japan. Hum Vaccin Immunother 2022 Dec 31;18(1):2010440-2010442. [doi: 10.1080/21645515.2021.2010440] [Medline: 34893009]
35. Chen CW, Lee S, Dong MC, Taniguchi M. What factors drive the satisfaction of citizens with governments' responses to COVID-19? Int J Infect Dis 2021 Jan;102:327-331 [FREE Full text] [doi: 10.1016/j.ijid.2020.10.050] [Medline: 33115678]
36. Betsch C, Schmid P, Heinemeier D, Korn L, Holtmann C, Böhm R. Beyond confidence: Development of a measure assessing the 5C psychological antecedents of vaccination. PLoS One 2018 Dec 7;13(12):e0208601 [FREE Full text] [doi: 10.1371/journal.pone.0208601] [Medline: 30532274]
37. Willis DE, Andersen JA, Bryant-Moore K, Selig JP, Long CR, Felix HC, et al. COVID-19 vaccine hesitancy: Race/ethnicity, trust, and fear. Clin Transl Sci 2021 Nov;14(6):2200-2207 [FREE Full text] [doi: 10.1111/cts.13077] [Medline: 34213073]

XSL•FO
RenderX

38. Pugliese-Garcia M, Heyerdahl LW, Mwamba C, Nkwemu S, Chilengi R, Demolis R, et al. Factors influencing vaccine acceptance and hesitancy in three informal settlements in Lusaka, Zambia. Vaccine 2018 Sep 05;36(37):5617-5624 [FREE Full text] [doi: 10.1016/j.vaccine.2018.07.042] [Medline: 30087047]
39. Robertson E, Reeve KS, Niedzwiedz CL, Moore J, Blake M, Green M, et al. Predictors of COVID-19 vaccine hesitancy in the UK household longitudinal study. Brain Behav Immun 2021 May;94:41-50 [FREE Full text] [doi: 10.1016/j.bbi.2021.03.008] [Medline: 33713824]
40. Bendau A, Plag J, Petzold MB, Ströhle A. COVID-19 vaccine hesitancy and related fears and anxiety. Int Immunopharmacol 2021 Aug;97:107724 [FREE Full text] [doi: 10.1016/j.intimp.2021.107724] [Medline: 33951558]
41. Fu C. Milestone and challenges: lessons from defective vaccine incidents in China. Hum Vaccin Immunother 2020 Sep 06;16(1):80-80 [FREE Full text] [doi: 10.1080/21645515.2019.1646579] [Medline: 31339795]
42. Okubo R, Yoshioka T, Ohfuji S, Matsuo T, Tabuchi T. COVID-19 Vaccine Hesitancy and Its Associated Factors in Japan. Vaccines (Basel) 2021 Jun 17;9(6):662 [FREE Full text] [doi: 10.3390/vaccines9060662] [Medline: 34204465]
43. Machida M, Nakamura I, Kojima T, Saito R, Nakaya T, Hanibuchi T, et al. Acceptance of a COVID-19 Vaccine in Japan during the COVID-19 Pandemic. Vaccines (Basel) 2021 Mar 03;9(3) [FREE Full text] [doi: 10.3390/vaccines9030210] [Medline: 33802285]
44. Wiysonge CS, Ndwandwe D, Ryan J, Jaca A, Batouré O, Anya BM, et al. Vaccine hesitancy in the era of COVID-19: could lessons from the past help in divining the future? Hum Vaccin Immunother 2022 Dec 31;18(1):1-3 [FREE Full text] [doi: 10.1080/21645515.2021.1893062] [Medline: 33684019]
45. MacDonald NE, SAGE Working Group on Vaccine Hesitancy. Vaccine hesitancy: Definition, scope and determinants. Vaccine 2015 Aug 14;33(34):4161-4164 [FREE Full text] [doi: 10.1016/j.vaccine.2015.04.036] [Medline: 25896383]
46. Statista. URL: https://www.statista.com/markets/424/topic/540/social-media-user-generated-content/ [accessed 2022-04-30]
47. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018 Jun 14;13(6):e0198857 [FREE Full text] [doi: 10.1371/journal.pone.0198857] [Medline: 29902270]

Abbreviations

amp: ampersands
AWS: Amazon Web Services
DOV: degree of valence
LDA: latent Dirichlet allocation
NLP: natural language processing
PELT: pruned exact linear time
PMOJ: Prime Minister's Office of Japan

Edited by M Gisondi, J Faust; submitted 22.02.22; peer-reviewed by C Tsagkaris, R Gore, V Ritschl; comments to author 19.04.22; revised version received 09.05.22; accepted 30.05.22; published 09.06.22.

Please cite as:
Niu Q, Liu J, Kato M, Nagai-Tanima M, Aoyama T
The Effect of Fear of Infection and Sufficient Vaccine Reservation Information on Rapid COVID-19 Vaccination in Japan: Evidence From a Retrospective Twitter Analysis
J Med Internet Res 2022;24(6):e37466
URL: https://www.jmir.org/2022/6/e37466
doi:10.2196/37466
PMID:35649182

©Qian Niu, Junyu Liu, Masaya Kato, Momoko Nagai-Tanima, Tomoki Aoyama. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.06.2022. 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

Navigating the Credibility of Web-Based Information During the COVID-19 Pandemic: Using Mnemonics to Empower the Public to Spot Red Flags in Health Information on the Internet

Jessica Stokes-Parish[1], BN, MNurs, PhD

Faculty of Health Sciences and Medicine, Bond University, Robina, Australia

Corresponding Author:
Jessica Stokes-Parish, BN, MNurs, PhD
Faculty of Health Sciences and Medicine
Bond University
14 University Drive
Robina, 4227
Australia
Phone: 61 755951468
Email: jstokesp@bond.edu.au

Abstract

Misinformation creates challenges for the general public in differentiating truth from fiction in web-based content. During the COVID-19 pandemic, this issue has been amplified due to high volumes of news and changing information. Evidence on misinformation largely focuses on understanding the psychology of misinformation and debunking strategies but neglects to explore critical thinking education for the general public. This viewpoint outlines the science of misinformation and the current resources available to the public. This paper describes the development and theoretical underpinnings of a mnemonic (Conflict of Interest, References, Author, Buzzwords, Scope of Practice [CRABS]) for identifying misinformation in web-based health content. Leveraging evidence-based educational strategies may be a promising approach for empowering the public with the confidence needed to differentiate truth from fiction in an infodemic.

(J Med Internet Res 2022;24(6):e38269) doi:10.2196/38269

KEYWORDS

science communication; critical appraisal; social media; health literacy; digital literacy; misinformation; COVID-19; online health; infodemic; infodemiology

Introduction

Overview

Recognizing misinformation in web-based content is becoming increasingly difficult. The general public struggles with differentiating credible health information from fiction, but we do not know how best to equip them to do so. In a world where information is at our fingertips, differentiating fact from fiction is a priority. This paper explores the science of misinformation and proposes an accessible framework for identifying misinformation in health content on the internet.

Background

The COVID-19 pandemic triggered an overabundance of information, including false and misleading information that has contributed to confusion, mistrust, and risk-taking behaviors. This kind of information excess is defined as an *infodemic* [1,2]. From daily press conferences to viral videos, health professionals and the general public alike have struggled to keep up with the overload of health information. The inundation of misinformation, disinformation, and contradictory information has obscured access to credible information.

Misinformation in science communications is not a new thing. *Misinformation* is defined as the inadvertent sharing of false or misleading information, whereas *disinformation* is the deliberate sharing of false or misleading information with the intent to harm [3]. Both topics are of great interest to psychologists and researchers. Prolific misinformation researchers Lewandowsky and colleagues [4] suggest that misinformation may arise when the situation is evolving or when the information is piecemeal. This is certainly the case with the pandemic, during which we have seen changes in information that was correct at a certain time, such as the use of masks to prevent SARS-CoV-2 transmission. Other sources of misinformation include rumors, politicians and governments, vested interests, and the media [4,5].

XSL•FO
RenderX

The Landscape of Misinformation

Misinformation is shared on a variety of platforms—Twitter, Reddit, WhatsApp, and Facebook to name a few [6]. However, misinformation is not limited to social media; it is also present in traditional media platforms, such as articles in magazines, on websites, and on the news. For example, in an analysis of health information on the internet, researchers found that of 1300 websites on safe infant sleep, only 43.5% provided correct information [7]. In another study on conception information, only 1 in 2 websites contained accurate information on conception [8]. The examples go on and on, particularly in the case of COVID-19, with multiple accounts of misinformation regarding COVID-19 treatments [9-11].

Topics of misinformation occur in a wide variety of fields, such as health and climate sciences [3]. Although it is difficult to quantify which topics have the most focus, we can get an indication by looking to research. Most research related to health misinformation focuses on vaccines, communicable (eg, HIV or COVID-19) and noncommunicable diseases (eg, cancer and diabetes), drugs (eg, tobacco), treatments, autism, and eating disorders [3,6,12].

Although misinformation pertains to the inadvertent sharing of false or partially false information, there is a more sinister kind of misinformation—disinformation. *Disinformation* describes sharing false or partially false information with the intent to harm or profit [13] (the term *fake news* is not used in this summary, as it is not supported by literature surrounding false information). Disinformation is a type of warfare strategy that has been linked to creating confusion regarding vaccination and disrupting election campaigns, as well as issues such as climate change [13,14].

How Does Misinformation Spread?

There are several decades of research dedicated to this issue, so this viewpoint will not attempt to cover the breadth of research on this complex issue. Instead, this paper will briefly outline why misinformation might spread. The author of this paper considers the following two broad categories of reasons: external and internal reasons. Externally, social media platforms amplify misinformation and disinformation due to their reach and the complex algorithms at play [4,5]. Internally, misinformation and disinformation disrupt our cognitive processes, fragmenting our ability to think logically. The little we do know about how and why misinformation spreads is that it is most often spread by individuals who hold positions of influence (eg, social media influencers or politicians) and share messages with personal opinions and strong negative tones [15]. In addition, a person's relationship with, or their view of, an individual sharing a piece of information influences perceived credibility; that is, if a person likes the individual and knows them well, the person is more likely to believe the information shared and is less likely to do a credibility check [4,15]. Misinformation is amplified by the impact of confirmation bias; people are more prone to misinformation that supports their worldview or ideology [16,17].

Health and Digital Literacy in an Infodemic

Although technology platforms such as Facebook and Twitter have a role in curbing the proliferation of misinformation and disinformation, digital literacy and health literacy are key factors in slowing the spread of misinformation and disinformation. *Health literacy* can be defined as the "ability of an individual to obtain and translate knowledge and information in order to maintain and improve health in a way that is appropriate to the individual and system contexts" [18]. Coldwell-Neilson [19] defines *digital literacy* as "the ability to identify and use technology confidently, creatively and critically to meet the demands and challenges of living, learning and working in a digital society." People with lower health literacy seek out health information less often and have a lesser ability to interpret health messages [20]. We also know that those with lower digital literacy are less able to identify reliable news sources or manipulated images [16], and those with less digital and health literacy are more likely to share false information [21].

What Is the Solution?

As the infodemic is unlikely to disappear anytime soon, we must consider ways to approach information on the internet. We are quick to defer to experts or exclaim "trust the science" as a sort of mantra for ordinary people. This does not engender trust or transparency in science but rather undermines attempts to engage in conversation about science, reinforces harmful hierarchies, and even leads to people falling for misinformation [22]. This mantra ignores the complexities and nuances of trust and engagement with scientific evidence, such as the influence of political persuasion, worldviews, and personal experiences [23]. Instead of restricting autonomy to that of scientists, it is the suggestion of this author that we consider ways to improve digital and health literacy to empower the general public to make informed decisions about the information they read [24].

What Exists?

Several resources on digital and health literacy exist. A quick keyword search of *health literacy course* and *health literacy training* on Google highlights the variety of resources from universities and not-for-profit organizations. For example, ScienceUpFirst—an initiative borne out of the COVID-19 pandemic—focuses on credible pandemic information [25]. Although they have a page on credible sources, this page focuses on who ScienceUpFirst considers credible as opposed to identifying components of credibility [25]. In a 2020 systematic review, researchers found that very little research focuses on critical thinking; even then, the limited research focused on student populations as opposed to the general public [26]. In addition, many courses on digital literacy, health literacy, or critical appraisal are recommended to health professionals, such as the Centre for Culture, Ethnicity and Health's courses [27] and Cochrane Training [28]. Research on misinformation extensively explores debunking, fact-checking, and prebunking (ie, preparing a viewer for incoming misinformation) [4,5]. To improve the health literacy of the general public, we should provide accessible appraisal resources, thereby allowing individuals to feel empowered when it comes to health information. In keeping with the constructivist philosophy, the

XSL•FO
RenderX

framework presented herein proposes that the general public should become collaborators in critical appraisal.

Methods and Theoretical Framework

Overview of Mnemonic Development

Drawing from the constructivist lens (ie, knowledge is subjective and informed by experiences), this paper considered the literature on credibility and critical appraisal and drew from this author's expertise as an educator to develop a mnemonic [15]. A mnemonic is a specific strategy for enhancing memory with the aim of improving the recall of information [29]. The purpose of the mnemonic in this instance was purely to create a memorable word (and visual) and a mental model for assessing health information on the internet [30,31].

The Framework

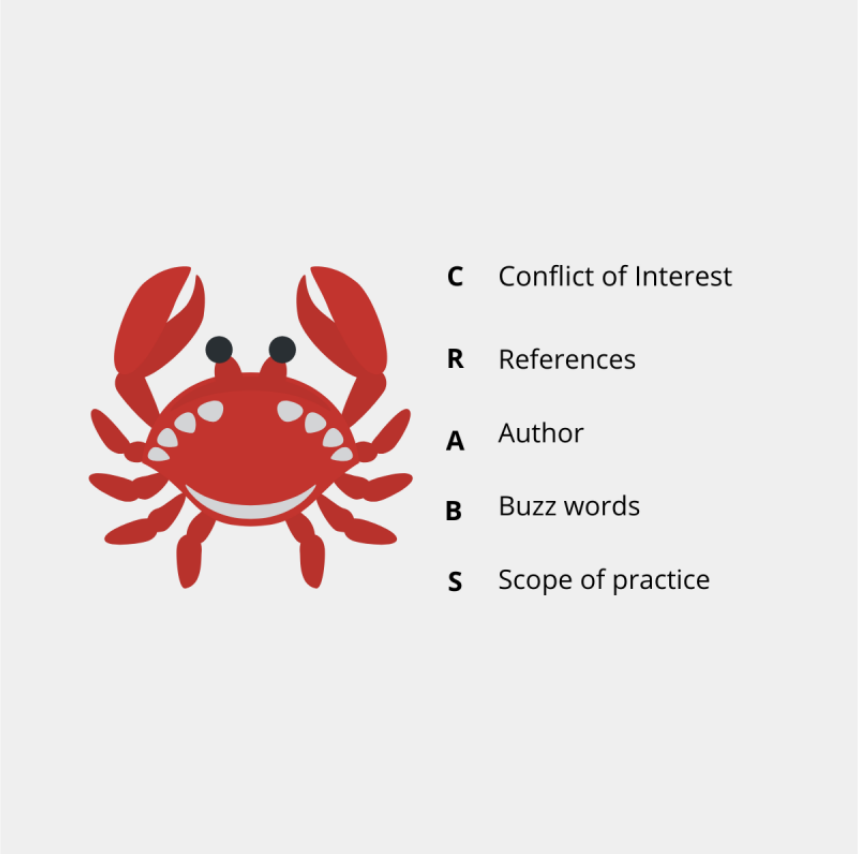
The mnemonic was developed by using an iterative process. This included an unstructured review of teaching materials for undergraduate and postgraduate health professions education (ie, materials that were used for teaching at the time of writing this paper), the use of library guides, and subsequent crowdsourcing on social media platforms [32-34]. Questions such as "how do you flag questionable content online" and "how would you review content online for accuracy" were used to engage readers. This process resulted in the development of the mnemonic *CRABS* (Conflict of Interest, References, Author, Buzzwords, Scope of Practice; Figure 1).

This paper's author presented the framework development work at professional development events and published it on several social media platforms. This was presented to registered nurses in Australia for a professional development activity on exploring credible content in the media. The feedback was overwhelmingly positive regarding the mnemonic, with 70% of participants identifying the mnemonic as their key takeaway from the activity. At the time of writing this paper, the framework has had significant reach on this author's social media platforms (Table 1). In addition to this, the work has been amplified on other social media influencers' posts, culminating in 68,000 unique views, and translated in other languages [35-37].

In addition to this, the work has been published on various media platforms, such as the Australian Broadcasting Corporation, lifestyle magazines, and high school education resources [38-40]. In health care, the framework has been shared in professional development resources, so that health educators can use the framework for their programs [41,42].

Now, we move on to the framework and underpinning rationale.

Figure 1. Illustration of the CRABS framework for credibility. CRABS: Conflict of Interest, References, Author, Buzzwords, Scope of Practice.



XSL•FO
RenderX

Table 1. Summary of impact.

Platform	Likes, n	Shares, n	Saves, n	Total number of unique views
Instagram	3783	1280	1691	54,492
Twitter	27	14	N/A[a]	8330

[a]N/A: not applicable.

C—Conflict of Interest

Conflicts of interest occur when an individual stands to benefit from a certain message or decision, making the information less reliable. Conflicts can be overt or subtle [43]. One example is an individual who owns a nutrition supplement company. This is an overt conflict of interest, as they are likely to prefer their product over others because it benefits them financially. In a more subtle example, a physician may have a family member who owns a company that manufactures wound care products; while there may not be any formal agreement, this relationship may influence the decisions that the physician makes about wound care [44]. In research, conflicts of interest may undermine the validity of results and undercut integrity. There have been many reports of trial sponsors inducing favorable results in research [45]. Conflicts of interest are not limited to finances; they can also include conflicts related to politics, policies, and employment [44]. Conflicts of interest should be considered when judging health information on the internet.

R—References

References are a useful gauge of content on the internet, as they indicate several things—supporting data, the body of evidence, the quality of evidence, and plagiarism [46]. Supporting data are obvious to scientists; one cannot make a claim without evidence. However, in health information on the internet, particularly on social media, the use of references is less common. Reminding viewers to consider references may assist them in considering the weight to give a claim. Additionally, references can be a good indication of whether the content author has training and an understanding of the body of evidence related to the topic. For instance, authors not citing sentinel work in their blogs can be a red flag for incomplete information. In addition to these checks, references should be checked for recency (science changes fast) and the quality of scientific sources. The issue of predatory publishing is not a small one. Predatory publishers accept manuscripts for a fee without performing any quality checks, thereby allowing poor-quality research or misinformation to proliferate [47]. Finally, plagiarism is not limited to scientific mediums; social media is rife with instances of content thieving and misattribution [48,49].

A—Author

Anyone can write about anything. The internet provides opportunities for everyone to have a voice and has fewer gatekeepers than traditional media [50]. A person's expertise and qualifications (or lack thereof) in relation to a topic is important to consider when determining how much weight to give to content. Social media verification, whereby an account is given a blue tick to verify that they are indeed a real person, is not an indicator of credibility in the traditional sense. Credibility literature states that there are 5 things to consider with regard to credibility—accuracy, authority, objectivity, currency, and scope [50]. The *Author* item of the CRABS framework encourages readers to check an individual's training, qualifications, and credentials. In 5 to 10 minutes, a reader can verify qualifications, explore the level of training that an individual has, and learn about the views of an author's peers (ie, their views on the author). For instance, if an author claims that they conduct research in a given field, the number of publications on the topic that are under their name should be considered.

B—Buzzwords

Buzzwords are words or phrases that have become fashionable by being used often, but they sometimes have little meaning. Buzzwords, or overly jargon-filled words, are not always designed to deceive people, but they can be used to mislead people. For example, when food packaging includes buzzwords such as *organic*, consumers are more likely to believe that a product is healthy—the health halo effect [51]. Linguistics research argues that clues are in the language used; emotional language is an indication that information is not credible [24]. News that is inaccurate or fake is more likely to use adverbs and verbs and present information with more certainty. This makes it challenging for credible science information, which frequently hedges certainty and does not overstate claims, to compete against noncredible information [52]. Other work suggests that framing the information in a certain way is a key for identifying misinformation. For example, topics of personal concern (eg, health information), emotive topics (eg, one's children), and the use of personalization pronouns (eg, *you*) can influence readers [17,53]. Overall, the trigger word *buzzwords* may help an individual to scan for jargon, marketing strategies, and emotional language that might frame their perception.

S—Scope of Practice

The scope of practice describes the practice of a profession that combines an individual's qualifications and expertise, the setting of practice, and the needs of clients [54]. In a health care setting, it is difficult to overreach the scope of practice due to highly regulated workforces. However, on social media, the scope is mostly unmonitored (but not necessarily unregulated). Most do not set out to overreach their scope of practice; however, it is a slippery slope. A nurse providing specific nutrition advice for newborns may be inappropriate if the nurse has not undergone additional training, depending on the situational context. In addition, it is easy to overstate expertise or specialty due to the halo of authority portrayed on social media. For example, a junior physician can inadvertently portray themselves as an expert in hormones while not having completed their endocrinology training.

XSL•FO
RenderX

Framework Application and Implications

The CRABS framework is intended to be applied as an overarching concept at a first glance of web-based content. It is not intended to be a full critical appraisal and may inadvertently exclude key qualities of appraisal that would be otherwise identified. One limitation may be that the framework could inadvertently exclude information that is credible due to the piecemeal nature of social media.

There are opportunities for expansion. For example, to verify the quality of the work, further research should be undertaken to determine content validity. Content validity analyses could include expert consensus methods, such as the Delphi method. Following this, consumer representatives (eg, members of the general public from a wide variety of demographic populations) could be engaged to rate the usability of the framework in assessing web-based health information. The framework should be further assessed for reliability and construct testing to ensure that the framework can indeed be used to identify accurate and false information and that it works for various users. Once the work has been validated, it could be used in critical appraisal guides, misinformation resources, and educational campaigns.

Although anecdotal, feedback has suggested that the content of the framework is representative of the issue and that it is usable among various users; however, more work is required to fully develop the framework. In its present state, the work could be presented as a conceptual model for assessing web-based health information, serving as a trigger for critical appraisal. Despite its origins in the COVID-19 pandemic, the work has scope for application beyond this. Areas rife with misinformation (eg, infant sleeping information, as identified earlier in this viewpoint) could be relevant areas.

Conclusions

In this era of infodemia, the general public requires accessible tools to navigate health information on the internet. Drawing from misinformation and educational research can provide us with tools to navigate this complex issue and develop resources. Using mnemonics is a practical strategy for encoding memory and developing mental models for critical appraisal. The CRABS model may provide a useful strategy for achieving this. More research is needed to explore the validity and usability of such a model for the general public.

Conflicts of Interest

None declared.

References

1. Infodemic. World Health Organization. URL: https://www.who.int/health-topics/infodemic#tab=tab_1 [accessed 2021-09-07]
2. Gisondi MA, Barber R, Faust JS, Raja A, Strehlow MC, Westafer LM, et al. A deadly infodemic: Social media and the power of COVID-19 misinformation. J Med Internet Res 2022 Feb 01;24(2):e35552 [FREE Full text] [doi: 10.2196/35552] [Medline: 35007204]
3. 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]
4. Lewandowsky S, Ecker UKH, Seifert CM, Schwarz N, Cook J. Misinformation and its correction: Continued influence and successful debiasing. Psychol Sci Public Interest 2012 Dec;13(3):106-131. [doi: 10.1177/1529100612451018] [Medline: 26173286]
5. Cook J, Ecker U, Lewandowsky S. Misinformation and how to correct it. In: Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource. Hoboken, New Jersey: John Wiley & Sons; May 15, 2015.
6. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of health misinformation on social media: Systematic review. J Med Internet Res 2021 Jan 20;23(1):e17187. [doi: 10.2196/17187] [Medline: 33470931]
7. Chung M, Oden RP, Joyner BL, Sims A, Moon RY. Safe infant sleep recommendations on the Internet: let's Google it. J Pediatr 2012 Dec;161(6):1080-1084 [FREE Full text] [doi: 10.1016/j.jpeds.2012.06.004] [Medline: 22863258]
8. Kedzior SGE, Bianco-Miotto T, Breen J, Diener KR, Donnelley M, Dunning KR, et al. It takes a community to conceive: an analysis of the scope, nature and accuracy of online sources of health information for couples trying to conceive. Reprod Biomed Soc Online 2019 Dec;9:48-63 [FREE Full text] [doi: 10.1016/j.rbms.2019.08.004] [Medline: 32021914]
9. Rachul C, Marcon AR, Collins B, Caulfield T. COVID-19 and 'immune boosting' on the internet: a content analysis of Google search results. BMJ Open 2020 Oct 26;10(10):e040989 [FREE Full text] [doi: 10.1136/bmjopen-2020-040989] [Medline: 33109677]
10. Wagner DN, Marcon AR, Caulfield T. "Immune Boosting" in the time of COVID: selling immunity on Instagram. Allergy Asthma Clin Immunol 2020 Sep 03;16:76 [FREE Full text] [doi: 10.1186/s13223-020-00474-6] [Medline: 32905318]
11. Cuan-Baltazar JY, Muñoz-Perez MJ, Robledo-Vega C, Pérez-Zepeda MF, Soto-Vega E. Misinformation of COVID-19 on the internet: Infodemiology study. JMIR Public Health Surveill 2020 Apr 09;6(2):e18444 [FREE Full text] [doi: 10.2196/18444] [Medline: 32250960]

XSL•FO
RenderX

12. Naeem SB, Bhatti R, Khan A. An exploration of how fake news is taking over social media and putting public health at risk. Health Info Libr J 2021 Jun;38(2):143-149 [FREE Full text] [doi: 10.1111/hir.12320] [Medline: 32657000]
13. Kapantai E, Christopoulou A, Berberidis C, Peristeras V. A systematic literature review on disinformation: Toward a unified taxonomical framework. New Media Soc 2020 Sep 20;23(5):1301-1326. [doi: 10.1177/1461444820959296]
14. Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, et al. Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. Am J Public Health 2018 Oct;108(10):1378-1384. [doi: 10.2105/AJPH.2018.304567] [Medline: 30138075]
15. Sbaffi L, Rowley J. Trust and credibility in web-based health information: A review and agenda for future research. J Med Internet Res 2017 Jun 19;19(6):e218 [FREE Full text] [doi: 10.2196/jmir.7579] [Medline: 28630033]
16. 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]
17. Zhao Y, Da J, Yan J. Detecting health misinformation in online health communities: Incorporating behavioral features into machine learning based approaches. Inf Process Manag 2021 Jan;58(1):102390. [doi: 10.1016/j.ipm.2020.102390]
18. Liu C, Wang D, Liu C, Jiang J, Wang X, Chen H, et al. What is the meaning of health literacy? A systematic review and qualitative synthesis. Fam Med Community Health 2020 May;8(2):e000351 [FREE Full text] [doi: 10.1136/fmch-2020-000351] [Medline: 32414834]
19. Decoding digital literacy. Decoding Digital Literacy. URL: https://decodingdigitalliteracy.org/ [accessed 2022-06-06]
20. 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]
21. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015 Mar 17;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
22. O'Brien TC, Palmer R, Albarracin D. Misplaced trust: When trust in science fosters belief in pseudoscience and the benefits of critical evaluation. J Exp Soc Psychol 2021 Sep;96:104184. [doi: 10.1016/j.jesp.2021.104184]
23. Rowland J, Estevens J, Krzewińska A, Warwas I, Delicado A. Trust and mistrust in sources of scientific information on climate change and vaccines: Insights from Portugal and Poland. Sci Educ (Dordr) 2022 Jan 09:1-26 [FREE Full text] [doi: 10.1007/s11191-021-00304-0] [Medline: 35035097]
24. Asr FT, Taboada M. Big data and quality data for fake news and misinformation detection. Big Data Soc 2019 May 23;6(1):205395171984331 [FREE Full text] [doi: 10.1177/2053951719843310]
25. Credible sources to share. ScienceUpFirst. URL: https://www.scienceupfirst.com/learn-more/ [accessed 2022-06-06]
26. Machete P, Turpin M. The use of critical thinking to identify fake news: A systematic literature review. 2020 Apr 01 Presented at: 19th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society; April 6-8, 2020; Skukuza, South Africa p. 235-246 URL: https://link.springer.com/chapter/10.1007/978-3-030-45002-1_20 [doi: 10.1007/978-3-030-45002-1_20]
27. Health literacy training courses. Centre for Culture, Ethnicity & Health. URL: https://www.ceh.org.au/health-literacy-training-courses/ [accessed 2022-06-06]
28. Online learning. Cochrane Training. URL: https://training.cochrane.org/online-learning [accessed 2022-06-06]
29. Bellezza FS. Mnemonic devices: Classification, characteristics, and criteria. Rev Educ Res 1981 Jun 01;51(2):247-275. [doi: 10.3102/00346543051002247]
30. Putnam AL. Mnemonics in education: Current research and applications. Transl Issues Psychol Sci 2015;1(2):130-139. [doi: 10.1037/tps0000023]
31. Manalo E. Uses of mnemonics in educational settings: A brief review of selected research. Psychologia 2002;45(2):69-79. [doi: 10.2117/psysoc.2002.69]
32. Evidence-based practice: Critical appraisal tools. Charles Sturt University. URL: https://libguides.csu.edu.au/ebp/critical_appraisal [accessed 2022-06-06]
33. Medicine for Students: Critical appraisal. The University of Sydney. URL: https://libguides.library.usyd.edu.au/c.php?g=508081&p=6644861 [accessed 2022-06-06]
34. Medicine and biomedical science: Evidence-based practice. University of Newcastle Library Guides. URL: https://libguides.newcastle.edu.au/medicine-biomedical-science/ebp [accessed 2022-06-06]
35. Blunden P. Paediatric nurse + Mum on Instagram. Instagram. 2022 May 22. URL: https://www.instagram.com/reel/CdnIYexuTkY/?utm_source=ig_web_copy_link [accessed 2022-05-29]
36. Novakovich J. Jen @theecowell #BeautyScience on Instagram. Instagram. 2021 Dec 03. URL: https://www.instagram.com/p/CXBaACtAx6N/?utm_source=ig_web_copy_link [accessed 2022-05-29]
37. Danielle J. Juf Danielle - Uitlegvideo's on Instagram. Instagram. 2021 Aug 02. URL: https://www.instagram.com/p/CSEPQxLtfUP/?utm_source=ig_web_copy_link [accessed 2022-05-29]
38. Judd B. Norman Swan, Sophie Scott and Kylie Quinn answer your questions about the COVID-19 vaccines. ABC News. 2021 Jan 26. URL: https://www.abc.net.au/news/2021-01-27/covid-19-vaccines-your-questions-answered/13094164 [accessed 2022-06-06]

XSL•FO
RenderX

39. Stokes-Parish J. Helping teens tackle social media with confidence. Lights Out Magazine. 2022. URL: https://www.boarding.org.au/our-community/lights-out-journal-1 [accessed 2022-04-11]
40. Mills H. Demystifying science with Dr Jess. In: Swell Magazine. Newcastle, New South Wales, Australia: Swell Magazine; Mar 01, 2022:6.
41. Stokes-Parish J. CRABS - The credibility framework. Jessica Stokes-Parish Education. URL: https://www.jessicastokesparish.com/blog/crabs-the-credibility-framework [accessed 2022-06-06]
42. Szabo R. 'Dr Google': using the internet for good. O&G Magazine (forthcoming). 2022. URL: https://www.ogmagazine.org.au/category/24/2-24/ [accessed 2022-05-29]
43. Muth CC. Conflict of interest in medicine. JAMA 2017 May 02;317(17):1812. [doi: 10.1001/jama.2017.4044] [Medline: 28464142]
44. Rahman-Shepherd A, Balasubramaniam P, Gautham M, Hutchinson E, Kitutu FE, Marten R, et al. Conflicts of interest: an invisible force shaping health systems and policies. Lancet Glob Health 2021 Aug;9(8):e1055-e1056 [FREE Full text] [doi: 10.1016/S2214-109X(21)00202-3] [Medline: 34297953]
45. Romain PL. Conflicts of interest in research: looking out for number one means keeping the primary interest front and center. Curr Rev Musculoskelet Med 2015 Jun;8(2):122-127 [FREE Full text] [doi: 10.1007/s12178-015-9270-2] [Medline: 25851417]
46. Santini A. The importance of referencing. J Crit Care Med (Targu Mures) 2018 Feb 08;4(1):3-4. [doi: 10.2478/jccm-2018-0002]
47. Grudniewicz A, Moher D, Cobey KD, Bryson GL, Cukier S, Allen K, et al. Predatory journals: no definition, no defence. Nature 2019 Dec;576(7786):210-212. [doi: 10.1038/d41586-019-03759-y] [Medline: 31827288]
48. Bailey M. On misogynoir: citation, erasure, and plagiarism. Fem Media Stud 2018 Mar 13;18(4):762-768. [doi: 10.1080/14680777.2018.1447395]
49. Bryson D. Using research papers: citations, referencing and plagiarism. J Vis Commun Med 2012 Jun;35(2):82-84. [doi: 10.3109/17453054.2012.690128] [Medline: 22747269]
50. Metzger MJ. Making sense of credibility on the web: Models for evaluating online information and recommendations for future research. J Am Soc Inf Sci Technol 2007 Sep 21;58(13):2078-2091. [doi: 10.1002/asi.20672]
51. Northup T. Truth, lies, and packaging: How food marketing creates a false sense of health. Food Stud 2014 Mar 29;3(1):9-18. [doi: 10.18848/2160-1933/cgp/v03i01/40545]
52. Pérez-Rosas V, Kleinberg B, Lefevre A, Mihalcea R. Automatic detection of fake news. arXiv Preprint posted online on August 23, 2017. [FREE Full text]
53. Singh VK, Ghosh I, Sonagara D. Detecting fake news stories via multimodal analysis. J Assoc Inf Sci Technol 2020 May 04;72(1):3-17. [doi: 10.1002/asi.24359]
54. Full scope of practice. Queensland Health. URL: https://www.health.qld.gov.au/ahwac/html/full-scope#:~:text=The%20full%20scope%20of%20practice,competent%20and%20authorised%20to%20perform [accessed 2022-06-06]

Abbreviations

CRABS: Conflict of Interest, References, Author, Buzzwords, Scope of Practice

Edited by M Gisondi, M Strehlow; submitted 30.03.22; peer-reviewed by SF Tsao, S Gordon; comments to author 04.05.22; revised version received 29.05.22; accepted 30.05.22; published 17.06.22.

Please cite as:
Stokes-Parish J
Navigating the Credibility of Web-Based Information During the COVID-19 Pandemic: Using Mnemonics to Empower the Public to Spot Red Flags in Health Information on the Internet
J Med Internet Res 2022;24(6):e38269
URL: https://www.jmir.org/2022/6/e38269
doi:10.2196/38269
PMID:35649183

©Jessica Stokes-Parish. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 17.06.2022. 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

XSL•FO
RenderX

information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Understanding How and by Whom COVID-19 Misinformation is Spread on Social Media: Coding and Network Analyses

Yuehua Zhao[1,2], PhD; Sicheng Zhu[1], BA; Qiang Wan[1], BA; Tianyi Li[1], BA; Chun Zou[1], BA; Hao Wang[1,2], PhD; Sanhong Deng[1,2], PhD

[1]School of Information Management, Nanjing University, Nanjing, China

[2]Jiangsu Key Laboratory of Data Engineering and Knowledge Service, Nanjing University, Nanjing, China

Corresponding Author:
Sanhong Deng, PhD
School of Information Management
Nanjing University
163 Xianlin Road
Qixia District
Nanjing, 210023
China
Phone: 86 2589685996
Email: sanhong@nju.edu.cn

Abstract

Background: During global health crises such as the COVID-19 pandemic, rapid spread of misinformation on social media has occurred. The misinformation associated with COVID-19 has been analyzed, but little attention has been paid to developing a comprehensive analytical framework to study its spread on social media.

Objective: We propose an elaboration likelihood model–based theoretical model to understand the persuasion process of COVID-19–related misinformation on social media.

Methods: The proposed model incorporates the central route feature (content feature) and peripheral features (including creator authority, social proof, and emotion). The central-level COVID-19–related misinformation feature includes five topics: medical information, social issues and people's livelihoods, government response, epidemic spread, and international issues. First, we created a data set of COVID-19 pandemic–related misinformation based on fact-checking sources and a data set of posts that contained this misinformation on real-world social media. Based on the collected posts, we analyzed the dissemination patterns.

Results: Our data set included 11,450 misinformation posts, with medical misinformation as the largest category (n=5359, 46.80%). Moreover, the results suggest that both the least (4660/11,301, 41.24%) and most (2320/11,301, 20.53%) active users are prone to sharing misinformation. Further, posts related to international topics that have the greatest chance of producing a profound and lasting impact on social media exhibited the highest distribution depth (maximum depth=14) and width (maximum width=2355). Additionally, 97.00% (2364/2437) of the spread was characterized by radiation dissemination.

Conclusions: Our proposed model and findings could help to combat the spread of misinformation by detecting suspicious users and identifying propagation characteristics.

(J Med Internet Res 2022;24(6):e37623) doi:10.2196/37623

KEYWORDS

health misinformation; COVID-19; social media; misinformation spread; infodemiology; global health crisis; misinformation; theoretical model; medical information; epidemic; pandemic

Introduction

Background

As early as February 15, 2020, the General Director of the World Health Organization stated at the Munich Security Conference, "We are not only just fighting an epidemic; but also an infodemic" [1]. Owing to quarantine restrictions imposed during pandemics, information access is limited to the internet and social media, which facilitates misinformation spread. According to one survey, 87% of internet users were exposed to pandemic-related misinformation [2].

XSL•FO
RenderX

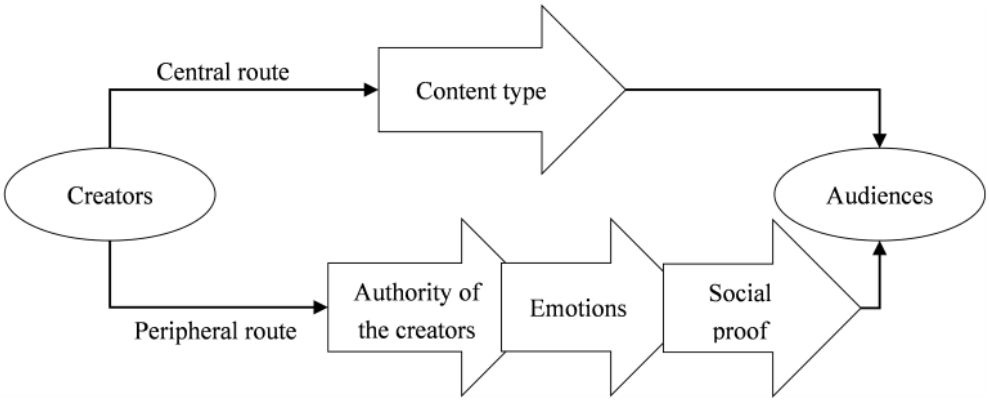
The spread of misinformation on social media can be amplified by information silos and echo chambers with personally tailored content. Kouzy et al [3] reported that 153 out of 673 tweets (24.8%) they examined contained COVID-19 pandemic misinformation, and 107 out of 673 tweets (17.4%) contained unverifiable information. Misinformation about the origin of the virus that originated from social media accounts has attracted more than 20 million hits [4].

Theoretical Context

On social media, misinformation can be defined as messages that aim to persuade other users. Persuasion theories state that the disseminator, message content, and recipient all have an impact on communication. Apart from studying the posts themselves, it is also necessary to examine the users who spread misinformation on social media. To uncover the characteristics of the spreaders of misinformation, we relied on persuasion theories that can help understand how misinformation is spread on social media. According to the elaboration likelihood model (ELM), a widely used persuasion model, users form their attitudes toward a message using either the central or peripheral path [5]. In the central path, users evaluate the quality and strength of the information, whereas in the peripheral route, they focus more on superficial factors such as the source's reputation, visual appeal, and presentation [6]. Therefore, the characteristics of online health misinformation can be divided into two levels: central and peripheral [7]. The characteristics of the central-level features of online coronavirus misinformation have been documented in the context of various countries [8,9] with misinformation spreading rapidly worldwide. However, little is known about the peripheral characteristics of COVID-19–related misinformation. Therefore, the first research question of this study was as follows: What are the central- and peripheral-level features of disseminated misinformation related to COVID-19? To answer this question, we propose a theoretical model of the persuasion process of COVID-19–related misinformation on social media based on the ELM. As shown in Figure 1, the central route feature (content type) and peripheral features (including the creator's authority, social proof, and emotion) are emphasized.

The dissemination of misleading information leads to increased public uncertainty, lack of belief in trustworthy sources, and, as a result, increased spread and ineffective containment of the virus [4]. To date, studies have largely focused on the extent of misinformation associated with coronavirus disease [3]. In contrast, a comprehensive understanding of the virality of COVID-19–related misinformation, particularly in the context of social media, is lacking. Accordingly, the second research question that guided this study was the following: How does COVID-19–related misinformation spread on social media? To answer this question, we measured the virality of COVID-19–related misinformation on social media from multiple perspectives, including the number of reposts, depth of reposts, width of reposts, and repost speed.

Figure 1. Theoretical model of the spread of COVID-19–related misinformation on social media.



Related Work

Persuasion Theory and Misinformation

Persuasion can be defined as "human communication that is designed to influence others by modifying their beliefs, values, or attitudes" [10]. The ELM views persuasion primarily as a cognitive event, meaning that the recipients of persuasive messages use mental processes of motivation and reasoning (or lack thereof) to accept or reject the messages [5].

In the peripheral route, messages rely on the emotional involvement of the recipient, and the recipient is persuaded by more superficial means. Cialdini [11,12] identified seven common cues that signal the use of a peripheral message: authority, commitment, contrast, liking, reciprocity, scarcity, and social proof. As a peripheral cue, authority can be used to convince the audience to accept the presented beliefs or behaviors. Previous work in psychology has shown that participants tend to believe information from people they consider credible [13]. From the social media perspective, the number of followers and followings (friends) of users may represent their social capital [14], which may indicate their authority. Posts by users with numerous followers (eg, opinion leaders) are perceived as trustworthy [15]. Moreover, Suh et al [16] found that the number of followers/followings of social media users positively influences the retweet probability of their posts.

Peripheral evidence of social proof is based on the age-old concept of peer pressure [11,12]. Empirical evidence suggests that engagement metrics provided by social media platforms, such as the number of likes and posts, also increase belief in social media content, especially in the case of misinformation [17]. Finally, rather than focusing on facts, the peripheral route

XSL•FO
RenderX

depends on associations with positive attributes such as positive emotions [18].

Spread of Misinformation on Social Media

Bode and Vraga [19] defined misinformation as "the factually incorrect information that is not backed up with evidence." Zhou et al [20] showed that misinformation with emotional and comparative terms is more likely to spread than correct information.

Moreover, researchers examined user-based characteristics to further understand the types of individuals who post or spread misinformation on social media [21]. Verification status, often assigned to official accounts and public figures to inform people that the account is authentic, is often used to measure the credibility of social media content [22]. Because crises are defined as emotional situations, the function of emotions in studies of misinformation connected to public health emergencies requires further investigation [23].

Existing research on the propagation characteristics of misinformation spread has focused on temporal factors [24,25] rather than propagation structure [26]. To capture the high-order propagation patterns of misinformation spread on social media, Ma et al [26] constructed a propagation network of misinformation on Twitter and identified that misinformation is typically first posted by a low-profile user, followed by some popular users who help spread it further, whereas genuine information is first posted by a prominent user and then directly shared by many general users.

Characteristics of COVID-19–Related Misinformation

Table 1 summarizes previous research on COVID-19–related misinformation on social media. Song et al [9] examined the types of misinformation disseminated during the COVID-19 pandemic in South Korea by analyzing fact-checking posts. Ceron et al [8] collected data from two Twitter accounts of Brazilian fact-checking projects and presented the refuted themes during the pandemic. Chen and Tang [27] reported that the spread of misinformation in public health emergencies had obvious characteristics of localization and high reproducibility.

In addition to content-based characteristics, the studies show that tweets from unverified accounts contain more misinformation compared to those from verified accounts (31% for unverified accounts, 12.6% for verified accounts; $P<.001$) [3]. Twitter accounts with more followers have fewer tweets with false information (20.1%, $P<.001$). Cinelli et al [28] found that while the number of posts from suspicious sources accounts for 70% of the volume of posts from trusted sources, the volume of responses to the former is three times higher than that to the latter.

In summary, there are two potential gaps in the existing literature that we address in this study. Previous studies have examined the characteristics of misinformation about the COVID-19 pandemic from several perspectives [3,8,9,27,29]. Psychological research has demonstrated the effect of news source credibility on persuasion [13], especially in the case of misinformation [30]. However, there is limited empirical research on the *source* of misinformation related to the COVID-19 pandemic (ie, the users who post the misinformation) and the misinformation dissemination patterns on social media platforms. In addition, based on the ELM, we proposed that the peripheral-level features of the spread of misinformation include the authority of the creators, emotion, and social proof, and then investigated the characteristics of the misinformation related to COVID-19 on social media. To the best of our knowledge, no study has considered all of the aforementioned features of COVID-19–related misinformation.

Table 1. Previous research on COVID-19–related misinformation on social media.

Study	Title	Method	Data	Source
Song et al [9]	The South Korean government's response to combat COVID-19 misinformation: analysis of "Fact and Issue Check" on the Korea Centers for Disease Control and Prevention website	Content analysis	90 posts	Korea Centers for Disease Control and Prevention (KCDC) website
Kouzy et al [3]	Coronavirus goes viral: quantifying the COVID19 misinformation epidemic on Twitter	Statistical analysis	673 tweets	Twitter
Ceron et al [8]	Fake news agenda in the era of COVID-19: identifying trends through fact-checking content	Topic analysis	5115 tweets	Twitter
Qin [29]	Analysis of the characteristics of health rumors in public health emergencies: Taking the "Shuanghuanglian" incident during the COVID-19 as an example	Case analysis	134 headings	COVID-19–related rumor list announced by Dingxiangyuan.com
Chen and Tang [27]	Analysis of circulating characteristics of rumors on Weibo in public emergencies: a case study of COVID-19 epidemic	Coding and visual analysis	968 posts	Weibo Rumor Refuting

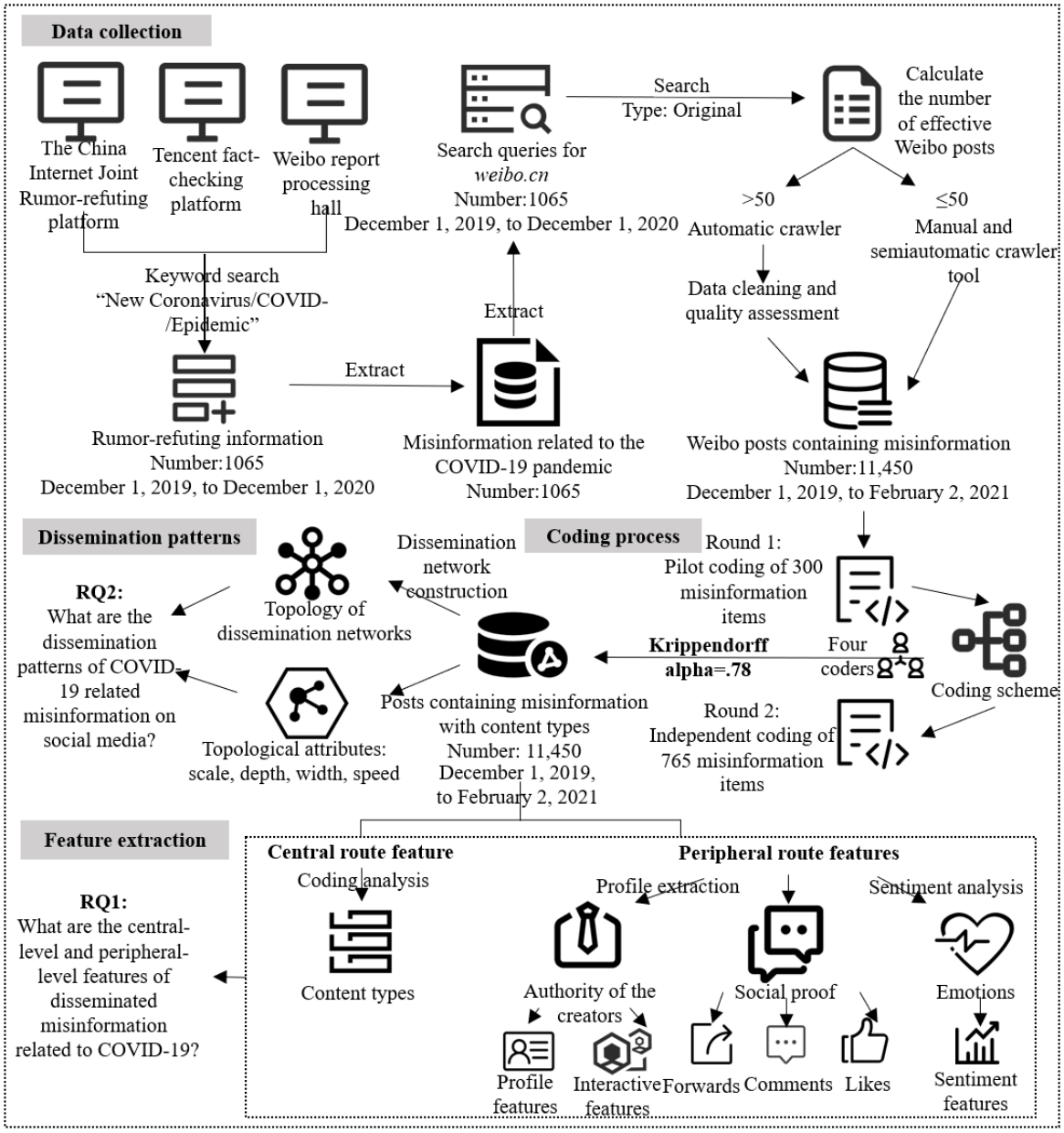
Methods

Process

First, we created a data set of COVID-19 pandemic-related misinformation based on fact-checking sources and then created another data set of circulated posts that contained this misinformation from a real-world social media platform. Based on the collected posts, we further analyzed the dissemination patterns and proposed peripheral-level characteristics of the coronavirus misinformation circulated on social media. The detailed data collection and analysis procedures are described in Figure 2.

XSL•FO
RenderX

Figure 2. Data collection and data analysis process.



Data Collection

Definition of Misinformation

Accurate identification of unknown misinformation is difficult for the general public because it requires multidisciplinary expertise. Reliable access to misinformation can be achieved by processing authoritative disconfirming information. For example, from the rebuttal information "Smoking can prevent coronavirus infection. This is false," we can extract the misinformation "smoking can prevent coronavirus infection."

Obtaining Misinformation Related to the COVID-19 Pandemic

As sources of authoritative misinformation, we selected three authoritative online platforms: China Internet Joint Rumor-Refuting Platform [31] (operated by the Office of the Central Cyberspace Affairs Commission), Tencent Fact Checking Platform [32] (operated by a large-scale internet integrated service provider), and Weibo Report Processing Hall [33] (operated by a large-scale public social media platform). We developed a web crawler to automatically collect the rumor-refuting information posted on each platform under the keywords "Novel Coronavirus (新冠病毒)/COVID/Epidemic (疫情)" from December 1, 2019, to December 1, 2020. After manual cleaning of irrelevant and repetitive information, we obtained 1065 COVID-19 pandemic–related misinformation posts.

Collecting Posts Containing COVID-19 Pandemic–Related Misinformation

To collect the circulated posts containing COVID-19 pandemic–related misinformation, we extracted keywords from all the collected misinformation, and then created corresponding queries for an advanced search on the Weibo.cn website. Considering the possibility of delayed and long-term dissemination of misinformation, the query search was limited

XSL•FO
RenderX

to original posts between December 1, 2019, and February 2, 2021.

To ensure the accuracy of the collected posts, the first round of collection was performed by manual query using a semiautomated collection tool. If there were more than 50 valid retrieved posts, the second round of collection was performed using an automated web crawler followed by data cleaning. After two collection rounds, Weibo posts containing misinformation were matched with the corresponding misinformation and 11,450 posts were finally identified.

Coding Process

A summary of previous research [34-36] on the classification of pandemic-related misinformation on social media can be found in Table S1 of Multimedia Appendix 1. Inspired by previous research, to characterize the content of COVID-19 misinformation, we manually coded the 1065 misinformation posts according to the six steps proposed by Richards and Hemphill [37]: (1) preliminary organization and planning, (2) open and axial coding, (3) development of a preliminary codebook, (4) pilot testing of the codebook, (5) final coding process, and (6) review of the codebook and finalization of the themes.

The coding scheme developed in this study is shown in Table 2. The COVID-19–related misinformation was classified into five content types: government response (Chinese-related), spread of epidemic (Chinese-related), medical information (Chinese-related), social issues and livelihood of people (Chinese-related), and international issues. Following Krippendorff's [38] method, four coders were recruited to conduct two rounds of manual labeling. In the first round, a pilot labeling of 300 misinformation items was performed. After further discussion, the remaining 765 misinformation items were independently labeled by the four coders in the second round, with an α value of .78. This meant that the four coders achieved substantial agreement in the topic assignment [39]. In the coding process, for misinformation items involving two or more topics, all four coders discussed and classified the items into the most relevant categories. Subsequently, 11,450 posts were labeled according to the labeling of the corresponding misinformation.

Table 2. COVID-19 pandemic–related misinformation topics.

Topic	Illustration	Example
Government response (Chinese-related)	Information related to traffic control, resumption of work and school, suspension of work and school, epidemic prevention measures, and others	It is said that after the disinfectant powder is sprayed over Wuhan today, patients with fever will be transported to designated hospitals.
Spread of the epidemic (Chinese-related)	Information related to the spread of the pandemic	The son-in-law of the Guanghan family came back from Wuhan for a few days. The family concealed their working address and went to play cards every day. He became ill today. The neighbors were very angry and went to smash his house.
Medical information (Chinese-related)	Information related to the virus itself, infection, prevention, treatment, disinfection, and other medical information	A doctor friend sent it. In response to this new type of coronavirus, the content of vitamin C (to fight the virus) and echinacea (to enhance immunity) can be used to prevent it.
Social issues and livelihood of people (Chinese-related)	Information related to celebrities, donation assistance, social aspects, and people's livelihood	National level response! All rented houses, apartments, shops and factories will be rent-free for one month in February, and rent-free for half a month in March and April! I hope that all "landlords" will respond positively! Overcome the difficulties together
International issues	Information related to other countries' response, online political rumors	Japan sent a 1,000-member medical team to Wuhan without masks and slogans.

Extraction of Post Features

Through the weibo.cn/repost/ website using a Weibo ID, we obtained the specific forwarding, liking, and commenting information of each post. Based on the forwarding relationships, we then created the forwarding network of the collected posts. Following Avram et al [17], we used engagement metrics, including the numbers of likes, comments, and forwards of each post, to represent the social proof features of posts containing misinformation (as summarized in Table 3).

XSL•FO
RenderX

Table 3. Features of posts containing COVID-19–related misinformation and users who have posted them.

Category	Description	Data type
Social proof features		
Forwards	Frequency of forwarding	Integer
Comments	Frequency of commenting	Integer
Likes	Frequency of liking	Integer
Profile features		
Verification status	Verified or not	Verified/Not verified
Verification type	Verification type	Category
Mrank	Weibo membership level	Integer (0-7)
Urank	User level	Integer (0-48)
Interactive features		
Posts_count	Number of posts	Integer
Followers_count	Number of followers	Integer
Following_count	Number of followings	Integer

Extraction of User Features

Apart from users who could not be captured because they were, for example, blocked, a total of 11,301 users who had published posts containing misinformation about COVID-19 were collected on Weibo.

Weibo's user authentication mechanism provides a channel for different types of users to prove their identity. The type of verification includes verified personal users, government users, media users, and businesses. The user level, as the basic characteristic of Weibo users, can largely represent the activity level of accounts. The higher the user level, the more active the user is. Membership level reflects users' habits in using Weibo. Users with a high membership level can be considered loyal users.

In addition to profile features, the interactive characteristics (ie, numbers of followers, followings, and posts) can also characterize users' authority on social media. The number of posts reflects the user's engagement on the social media platform. Users with a considerable number of followers can share their opinions with a large group of people [16], whereas users with many followings have a broad range of information sources [14]. Therefore, both profile features and interactive features characterize the authority of users who have published posts containing misinformation. Detailed descriptions of these features can be found in Table 3.

Sentiment Analysis

Sentiment characteristics have been recognized as effective features for distinguishing online rumors and fake reviews [40]. In this study, we performed sentiment analysis by applying a pretrained convolutional neural network model to the collected data set on the Baidu Senta system [41]. Senta incorporates sentiment knowledge into pretrained models and produces new state-of-the-art results on most of the test data sets [42]. Senta is exposed to a vast corpus from multiple settings as an open sentiment analysis platform, which considerably increases the validity of its analyses [42,43]. Senta uses an unsupervised method to automatically mine emotional knowledge, and comprehensively surpassed other approaches in 14 typical tasks of Chinese emotional analysis [41]. Each post that is input into Senta returns an outcome of positive sentiment likelihood ranging from 0 to 1, which can be utilized as the sentiment feature of the post.

Construction of the Dissemination Network

To describe the prevalence of coronavirus-related misinformation on social media, in addition to the number of forwards of each post, we crawled the detailed forwarding information for each post in the data set created in the previous step of the research and collected a list of forwards of the original misinformation posts. The forwarding information for each post included the users who forwarded the original post, the content of the reposts, and the forwards and likes that the reposts received. The Weibo platform uses the "//" symbol to divide the forwarded content into different forwarding levels. Therefore, the forwarding level of each post can be extracted based on the forwarded content. In addition, the dissemination network of each post can be constructed using a series of reposts based on the corresponding forwarding relationships. Thus, apart from posts that could not be captured because they were, for example, blocked or deleted, we constructed a dissemination network for a total of 2437 posts that contained information about COVID-19. In these networks, each node represents an individual post, whereas a directed link represents a forwarding relationship from the source node to the repost node. For example, if post A forwards the original post B, then an edge is drawn from nodes B to A.

Extraction of Dissemination Scale, Depth, Width, and Speed

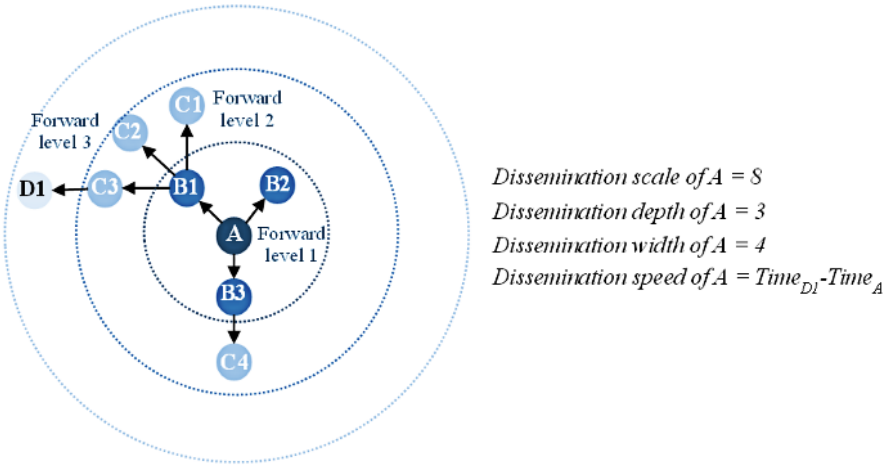
In the constructed dissemination networks, each node represents a single post that was involved in the spread of misinformation related to COVID-19. Based on the network for each original post, the dissemination scale refers to the number of nodes in

XSL•FO
RenderX

the network, corresponding to the number of forwards for the original post. The dissemination depth indicates the highest level of repost in the network of the original post, whereas the dissemination width is equal to the number of nodes at the level with the largest number of nodes in the network.

Figure 3 shows an example of a dissemination network for a post. Node A denotes an original post containing COVID-19–related misinformation. Posts B1, B2, and B3 forward the original post. Subsequently, post B1 is forwarded by posts C1, C2, and C3, while post B3 is forwarded by post C4; post C3 is also forwarded by D1. In this case, the original post A has been forwarded eight times and thus has a dissemination scale of 8. The second level of forwarding involves most posts, which means that the propagation width of the original post A is equal to 4. The highest level of forwarding signifies that the dissemination depth is 3.

Figure 3. Illustration of the dissemination scale, depth, width, and speed of a sample post. Each node represents a single post that was involved in the spread of misinformation related to COVID-19.



Ethical Considerations

No ethics approval was required as this study was based on publicly available data and involved no personally identifiable data.

Results

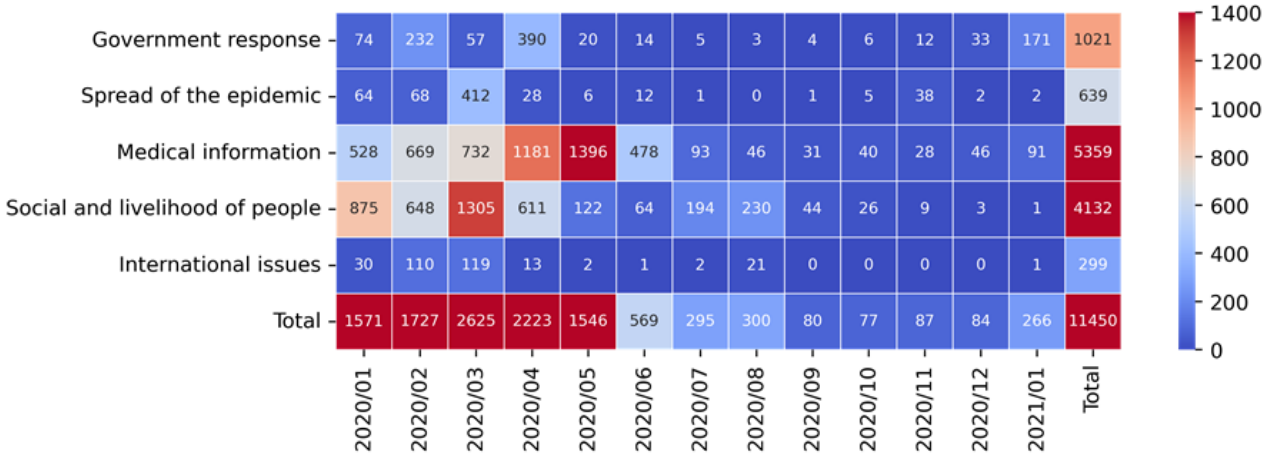
Central Route Feature of Posts

To answer the first research question, coding analysis was performed to identify the content types/topics of posts containing COVID-19–related misinformation. A total of 11,450 such posts were categorized under five topics: government response (n=1021), spread of the epidemic (n=639), medical information (n=5359), social issues and livelihood of people (n=4132), and international issues (n=299). The most common theme was medical misinformation (5359/11,450, 46.80%), including misinformation about the virus, infection, prevention, treatment, and disinfection. The second most popular topic was social issues and livelihood of people (4132/11,450, 36.09%), especially related to fake statements about celebrities. This category also included posts referring to donations that were refuted.

To distinguish different topics of posts, the number of posts and corresponding dates are plotted in Figure 4, where the warmer color represents a more popular misinformation topic in a given period. Social and livelihood misinformation emerged most prominently in mid-March when several widely circulated misinformation topics appeared, including statements about a Malaysian shaman who can cast a spell to cure the coronavirus. Posts with misinformation about medical information appeared most frequently from February to May, which coincided with the most severe period of the epidemic in China.

XSL•FO
RenderX

Figure 4. Changes in the number of posts containing misinformation over time.



Peripheral Route Features of Posts

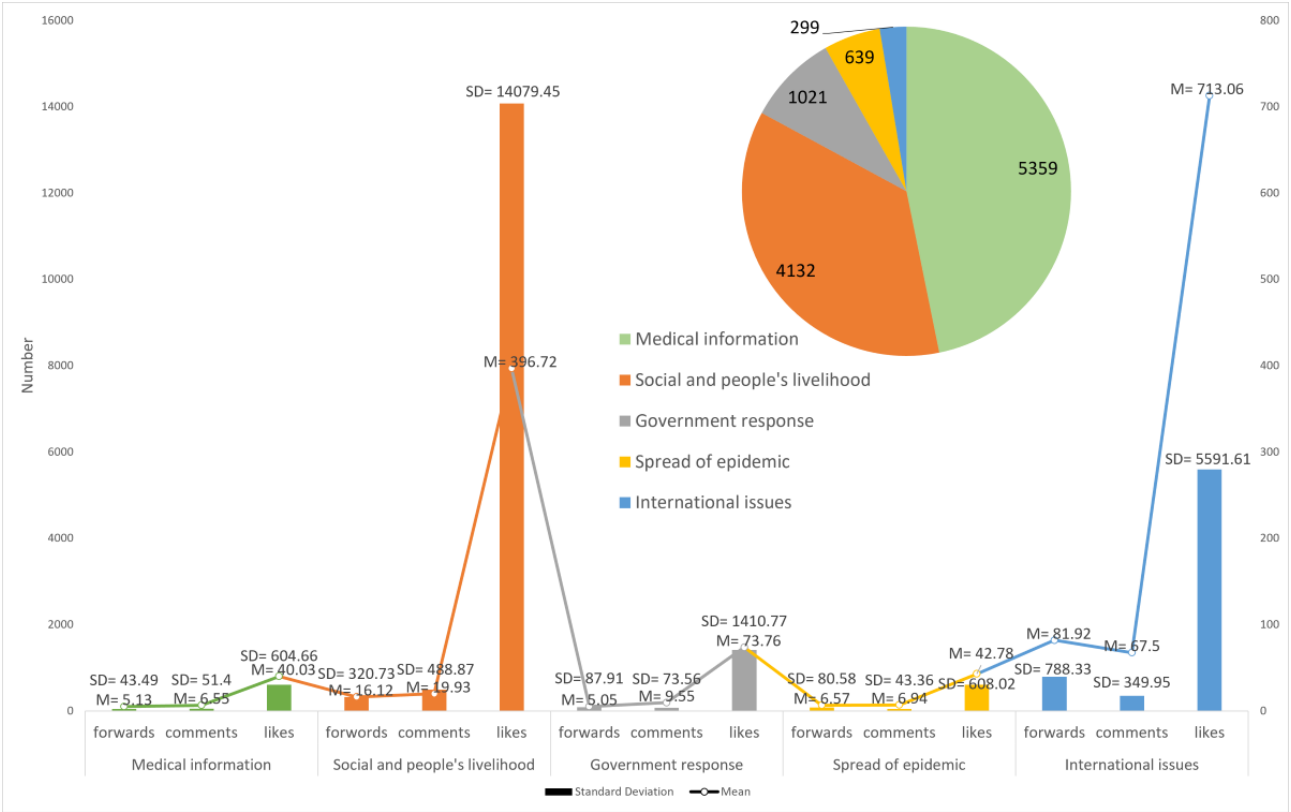
Overview

To answer the first research question, we also examined the social proof features of the collected posts, the sentiment features of the posts, and the authority features of the users who posted them.

Social Proof Features of the Posts

The collected posts with COVID-19–related misinformation received 11 forwards, 13 comments, and 189 likes, on average. The pie chart in Figure 5 shows the number of posts containing misinformation on different topics, and the bars and the line show the standard deviations (on the left Y-axis) and average numbers (on the right Y-axis) of likes, comments, and forwards that the posts on each topic received. Considering the topics of misinformation, posts with the most likes (mean 713), comments (mean 67), and forwards (mean 82) contained misinformation on international issues, whereas posts with medical information received the least attention, with an average of 40 likes, 7 comments, and 5 forwards (see Figure 5). Among all five topics, posts containing misinformation on international issues accounted for only 2.61% (299/11,450) of all posts, but achieved much higher attention. This may suggest that misinformation involving international issues, while less frequent, is much more viral across all three types of social proof features, indicating the potentially serious consequences of such misinformation.

Figure 5. Social proof features of posts related to various misinformation topics.



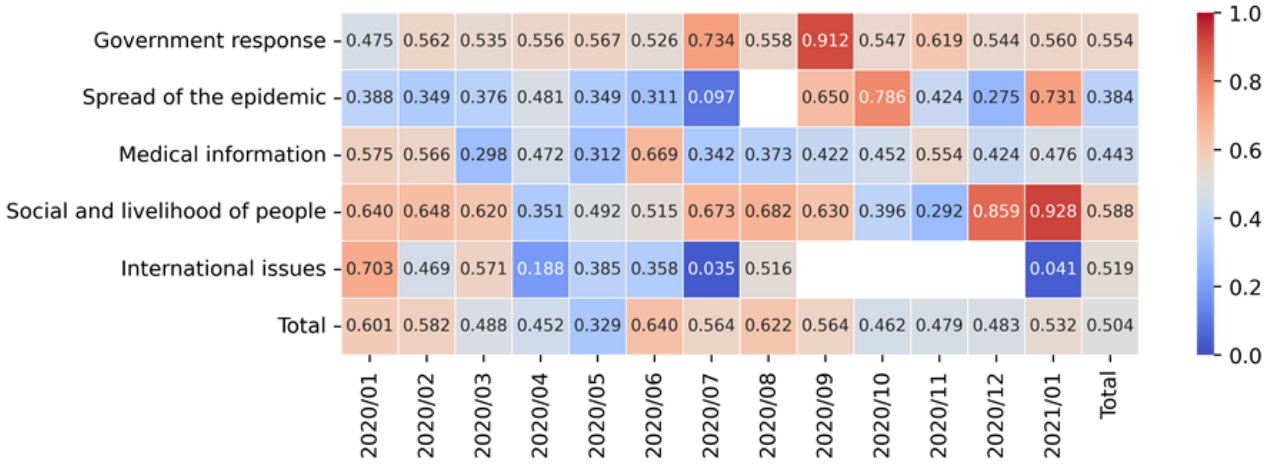
XSL•FO
RenderX

Sentiment Features of the Posts

Considering the misinformation topics, Figure 6 describes the variance of the mean value of the generated sentiment intensity value of the COVID-19-related misinformation posts. The warmer the color, the more positive the emotion, and conversely, the colder the color, the more negative the emotion. The average sentiment value of the posts remained neutral to positive over time. Some of the posts that contained misinformation about unproven preventive measures and treatments appeared to be relatively positive. For example, "A hot bath in the home is the easiest, most effective, and least costly way to protect susceptible people!" conveyed extremely optimistic emotions. In contrast to the general negative sentiment that prevailed among the public during the pandemic [44], posts containing misinformation tended to express positive emotions to attract public attention.

In contrast to other topics, posts spreading misinformation related to spread of the epidemic tended to be consistently negative. In particular, misinformation related to the lifting of the lockdown and traffic restrictions expressed very negative emotions, such as "Harbin is closed! Urgent city closure. No chance for any travel."

Figure 6. Sentiment features of posts related to various misinformation topics.



Authority Features of Users

Among the users who posted messages containing misinformation related to COVID-19, certified users accounted for 46.60% (5266/11,301). Among them, verified personal users were the most prominent sources (2475/5266, 47.00%), followed by media users (1159/5266, 22.01%) and government accounts (1013/5266, 19.24%). The number of nonverified users was only 6.8% more than that of verified users, representing 53.40% (6035/11,301) of the messages. This suggests that when detecting misinformation, whether it was published by a verified account cannot be a criterion for determining the authority of the information.

We found obvious differences between the authority characteristics of users who posted messages containing misinformation on different topics. For international issues, certified users (160/292, 54.8%) outnumbered uncertified users (132/292, 45.2%), and certified users (3160/5310, 59.51%) also outnumbered uncertified users for medical information. By contrast, misinformation about social issues and people's livelihoods was more likely to be posted by uncertified users (2586/4093, 63.18%) than by certified users (1507/4093, 36.82%), and misinformation about the spread of the epidemic was also more likely to be posted by uncertified users (409/630, 64.9%).

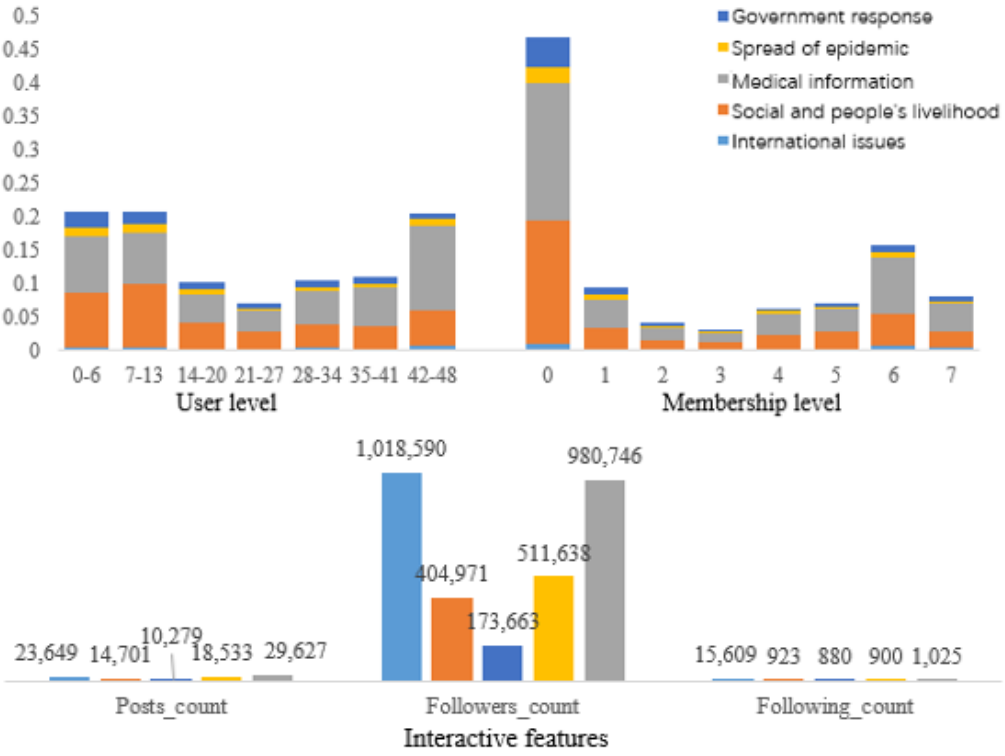
The user-level and membership-level distributions of users posting misinformation on various topics are shown in the upper part of Figure 7. Considering the user and membership levels, the users with the lowest levels (in the range of 0-13 for user level and 0 for membership level) were the most responsible for publishing misinformation, accounting for 41.24% (4660/11,301) and 46.77% (5285/11,301), respectively. Surprisingly, the second most responsible user groups were those with the highest user (in the range of 42-48; 2320/11,301, 20.53%) and membership (in the range of 6-7; 2680/11,301, 23.71%) levels. This may indicate that misinformation is posted by both the least and most active users. For international misinformation and medical misinformation, users with the highest user levels (in the range of 42-48) accounted for the most (1516/5602, 27.06%), while misinformation on the other three topics tended to be posted more by users with lower user levels (in the range of 0-13; 2758/5699, 48.39%).

The lower part of Figure 7 shows the average interactive features of users who posted messages containing misinformation on different topics. The number of followers of misinformation publishers averaged over 100,000, demonstrating the great social capital they have on social media.

In comparison, users who posted misinformation related to international issues and medical information tended to have higher authority than those who posted about the government response, social issues and livelihood of people, and spread of the epidemic. The numbers of posts, followers, and followings of users who posted misinformation related to the government response were the lowest among the five topics, representing users who had less authority.

XSL•FO
RenderX

Figure 7. User-level and membership-level distributions and average interactive features of users posting misinformation about various topics.



Dissemination Patterns of the Posts

Based on the constructed dissemination network for each of the 2437 posts containing COVID-19–related misinformation, we extracted the dissemination scale, depth, maximum width, average width, and speed for each post. In this study, maximum width measured the number of nodes involved in the widest level and average width measured the average number at all levels. Table 4 summarizes the descriptive statistics for the dissemination patterns of all posts. In our data set, the most widely disseminated post comprised 7604 users in the dissemination network. The posts that spread the deepest were forwarded by 14 levels of users, whereas the posts that spread the most widely were forwarded by 2355 users in a given dissemination level. On average, the dissemination scale of all posts was 19.7, with an average depth of 1.5 and an average maximum width of 20.5.

Table 4. Descriptive statistics of the dissemination patterns.

Dissemination measures	Mean (SD)	Maximum
Scale	19.7 (236.03)	7604
Depth	1.5 (0.99)	14
Maximum width	20.5 (87.82)	2355
Average width	15.9 (23.74)	688
Speed	2.4 (8.20)	96.9

Figure 8 shows the 95% CI plots for the means of the dissemination scale, depth, maximum width, average width, and speed for the posts related to each topic. Posts containing misinformation related to social issues and livelihood of people clearly engaged many more users in the dissemination network than posts on other topics. Compared to the other topics, misinformation related to the government response (average dissemination scale 28.2) and social issues and livelihood of people (average dissemination scale 27.4) engaged more users in the discussions and propagation. In terms of dissemination depth, maximum width, and average width, posts on international issues reached a larger audience at each dissemination level and resulted in more in-depth propagation. Unlike other topics, most international misinformation posts attracted some level of public attention (with a forwarding rate of 168/299, 56.2%), suggesting that this type of misinformation is more likely to be subject to widespread, large-scale, and heated mainstream discussions.

Based on the structure, we divided the dissemination network of the misinformation posts into three main types: (1) radiation dissemination network, where the first-level dissemination is wider than all other levels; (2) sector dissemination network, where the width of the other levels in the dissemination network is wider than that of the first level, and the node with the highest forwarding volume reaps more forwards than likes; and (3) viral dissemination network, where the width of other levels in the dissemination network is larger than that of the first level, and

XSL•FO
RenderX

the node with the highest like volume reaps more likes than forwards. Figure 9 shows representative examples of all three networks.

Examination of the posts revealed that 97.00% (2364/2437) were disseminated through the radiation dissemination network, only 0.98% (24/2437) belonged to the sector dissemination network, and 2.01% (49/2437) belonged to the viral dissemination network. As shown in Figure 9a, the width of the first-level forwarding radiated from the original node is much larger than that of the subsequent layers, which is reflected in the much higher density of nodes around the root node than that of nodes on the other levels. In this type of dissemination network, the user who created the original post usually has higher authority (eg, a large number of followers); such users include public organizations and news media. In addition, the content of the posts is likely political. The node density tends to decrease as the level increases, indicating that the potential impact of these posts declined after the first-level forwarding. For the dissemination networks shown in Figure 9b and c, although the width of the first level is relatively large, the nodes at the other levels are fan-shaped, with the nodes radiating from multiple sublevel nodes. In the spread of posts pertaining to the sector and viral dissemination networks, there are multiple nodes with a high degree of propagation ability. In contrast to the posts present in the radiation dissemination network, some opinion leaders, who possessed a large number of followers and were not verified as authoritative institutions or public organizations, also played a role in the spreading of the posts in the sector and viral dissemination networks.

Figure 8. Confidence interval plots for the dissemination patterns.

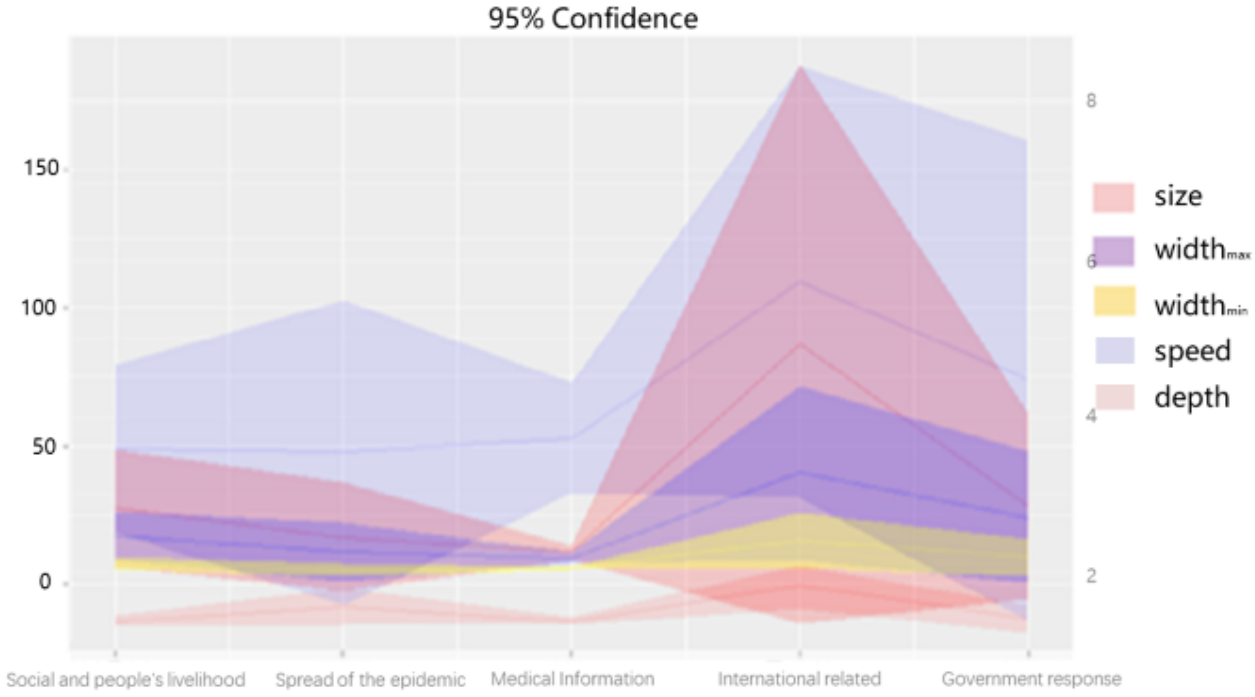
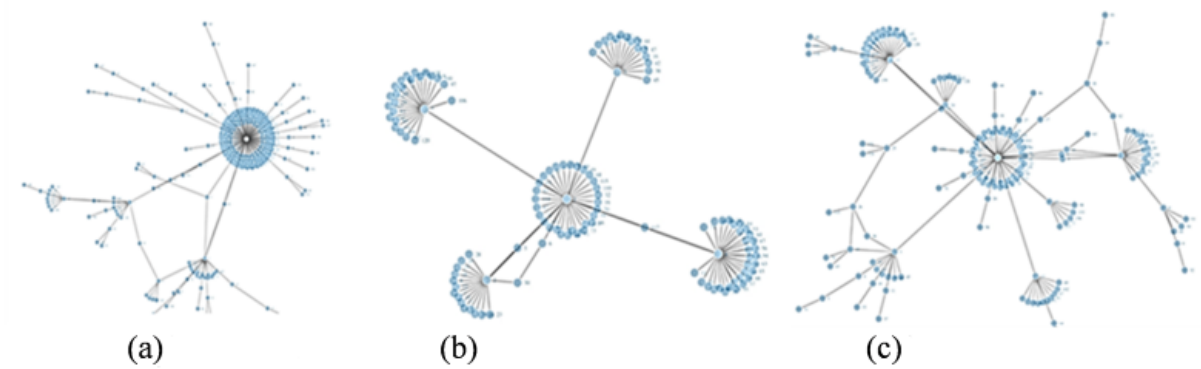


Figure 9. Examples of each dissemination network type. (a) Radiation dissemination network. (b) Sector dissemination network. (c) Viral dissemination network.



Clustering Analysis of Misinformation Disseminators

To further characterize the users who posted misinformation about coronavirus on social media, we leveraged the k-means clustering algorithm to classify users based on user authority features (including user level, membership level, posts count, follower count, and following count). To ensure the quality of the clustering, the Nbclust function in R was used to test

XSL•FO
RenderX

different values of k. Based on the elbow method, 5 was selected as the optimal number of clusters. Figure 10 shows the users in the form of a scatterplot matrix created using the ggpairs function from the GGally R package [45]. Each point is colored by a cluster identified using the k-means clustering algorithm.

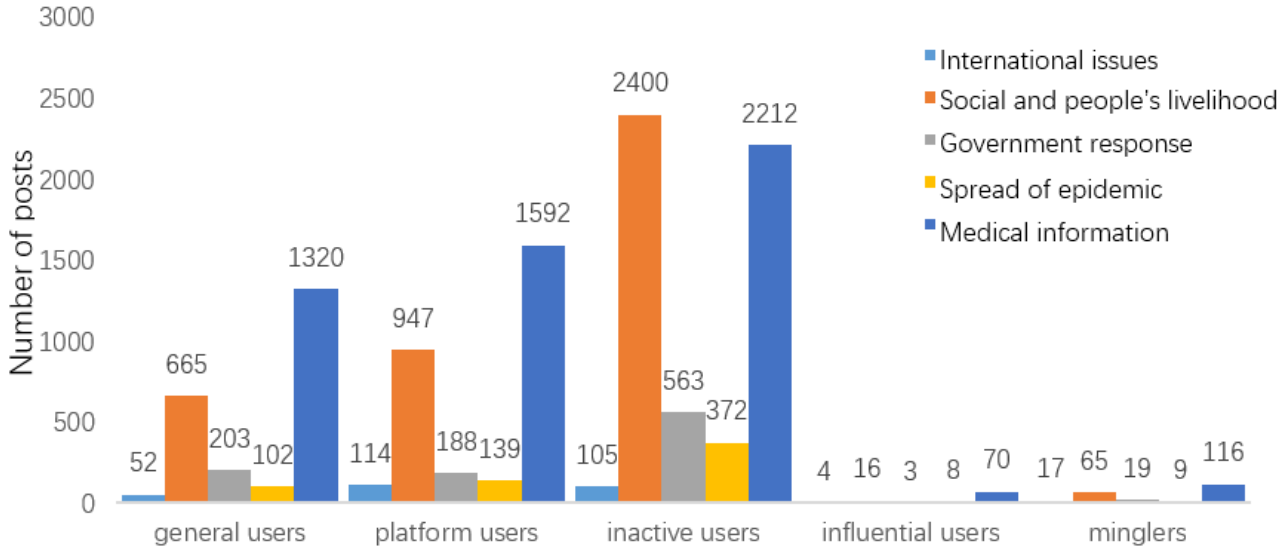
As determined by the k-means clustering algorithm, users who posted misinformation were classified into five groups: general users, platform users, inactive users, influential users, and minglers. A total of 2342 users were classified as general users, who participate in social media but are less willing to pay for membership. They tended to have a low membership level but a high user level, and their performance in terms of the number of posts and followers was relatively normal. The behavioral patterns of platform users (comprising 2980 users) were similar to those of general users, except for their membership level. They tended to have significantly higher membership levels, indicating that they were both actively participating in social interactions and purchasing memberships to enjoy the privileges. The largest group, inactive users, comprised 5652 users who appeared at lower frequencies for all five features. In contrast, influential users, the smallest group of users (comprising 101 users), posted more frequently than others and also reaped a large number of followers. Users in this group tended to remain in the highest position with respect to both user and membership levels. Finally, users in the mingler group had a higher number of followers than the other user groups, but they made fewer posts. The characteristics of the 226 users in the mingler group were consistent with those identified by Kozinets [46] as users who maintain strong social ties on social media while being marginally interested in activities.

The distribution of different types of users posting misinformation on different topics is shown in Figure 11. Inactive users were more likely to post misinformation about social issues and livelihood of people, which significantly differed from the other four user types. Influential users contributed the most to medical misinformation dissemination, but were less likely to post misinformation related to social issues and people's livelihoods.

Figure 10. Scatterplot and correlation matrix of user authority features. a: The correlation is significant at a significance level of .001 (two-sided); b: The correlation is significant at a significance level of .01 (two-sided); c: The correlation is significant at a significance level of .05 (two-sided).

XSL•FO
RenderX

Figure 11. Distribution of various types of users posting misinformation related to various topics.



Correlations Between Topological Attributes and User Authority Features

We also performed a correlation analysis to test whether the dissemination network features were significantly related to the authority features of the users who created the posts. The Spearman rank correlation coefficient was used to measure the correlations between the authority features of the creators and the features of the resulting dissemination network. Table 5 shows that the number of followers and membership level were significantly correlated with all five characteristics of the dissemination network (ie, dissemination scale, depth, average width, maximum width, and speed) under two-sided test conditions (P<.01).

From a network perspective, messages posted by users with numerous followers tend to receive more attention on social media. Users with high membership levels are more likely to engage in social media interactions. Similar to the context of disaster-related information [14], the number of followers was found to have a significant positive effect on the dissemination scale, depth, and speed of the post. This suggests that misinformation from users with high authority levels have a high intensity in wide broadcasts, and thus can have serious consequences.

Table 5. Spearman correlation (ρ) analysis between topological attributes and user authority features.

Dissemination variables	Posts count	Followers count	Following count	Membership level	User level
Scale					
ρ	0.114	0.344	0.009	0.171	0.107
P value	<.001	<.001	.77	<.001	.001
Maximum width					
ρ	0.103	0.349	0.008	0.17	0.1
P value	.002	<.001	.80	<.001	.002
Average width					
ρ	0.081	0.345	–0.007	.171	0.096
P value	.01	<.001	.83	<.001	.003
Depth					
ρ	0.174	0.197	0.106	0.08	0.118
P value	<.001	<.001	.001	.02	<.001
Speed					
ρ	0.023	0.174	–0.003	0.105	0.047
P value	.48	<.001	.93	.001	.16

XSL•FO
RenderX

Discussion

Principal Findings

Understanding the underlying psychology of why people fall for misinformation is key to developing effective interventions against it [30]. This study of posts containing COVID-19–related misinformation reveals important insights into its dissemination on social media. To analyze the central-level feature, the COVID-19–related misinformation content was classified into five types: medical information (5359/11,450, 46.80%), social and livelihood of people (4132/11,450, 36.09%), government response (1021/11,450, 8.92%), spread of the epidemic (639/11,450, 5.58%), and international issues (299/11,450, 2.61%). The classification of the content is consistent with previous research on recent epidemics [34]. Consistent with previous studies on COVID-19 in which most of the misinformation was medical-related [1,3], we found that the predominant themes of misinformation were related to medical information, as well as to social and livelihood issues. In particular, posts with medical misinformation appeared most frequently during the most severe phase of the pandemic.

Misinformation that attracts attention can trigger intense discussions, thus promoting the spread of information. In addition to the central-level feature, social proofs of posts with COVID-19–related misinformation showed that such misinformation was actively responded to (with an average of 11 forwards, 13 comments, and 189 likes). Interestingly, misinformation related to international issues accounted for 2.61% (299/11,450) of all posts but achieved alarmingly higher attention (with an average of 82 forwards, 67 comments, and 713 likes), suggesting that misinformation involving international issues tends to go viral on social media and thus can have serious consequences. This is consistent with empirical findings on Twitter, where COVID-19–related conspiracy misinformation is most likely to spread [47].

In contrast to the negative sentiment that emerged among the public during the pandemic [48], positive to neutral sentiment in the misinformation posts appeared during the pandemic. In particular, some misinformation about the untested prevention measures and treatment seemed extremely positive and captured the public's attention. Our results support the idea that misinformation topics should be considered when designing misinformation interventions during the pandemic [47].

Analysis of user profile characteristics revealed that users with the lowest and highest levels of user and membership levels were the most responsible for publishing misinformation. Our results suggest that both the least and most active users are prone to sharing misinformation. In contrast to the empirical results of Kouzy et al [3], where unverified Twitter accounts published significantly more misinformation than verified accounts, verified Weibo users accounted for nearly half (5266/11,301, 46.60%) of all messages containing COVID-19–related misinformation. This suggests that user verification status is not applicable to coronavirus-related misinformation detection on Weibo.

The average number of followers of the misinformation publishers was extremely high (>100,000), indicating the credibility and social influence they possess on social media. Some marketing-oriented accounts changed the main part of genuine news to attract users. As for the medical misinformation, some corporate accounts fabricated misinformation (eg, "Natto can inactivate the virus") to promote their product.

The average dissemination scale of misinformation posts was 19.7, with an average depth of 1.5 and an average maximum width of 20.5. Li et al [14] reported average dissemination scale and depth for disaster-related posts of 68.64 and 1.113, respectively. This suggests that COVID-19–related misinformation posts spread deeper than disaster-related posts, while the dissemination scale was lower than that of disaster-related posts. Moreover, posts about international issues were the most likely to have profound and lasting effects on social media, with the highest numbers in terms of dissemination depth, maximum width, and average width. This highlights the importance of dealing with COVID-19–related misinformation, especially that related to international issues.

In capturing the topological attributes of the dissemination network, three main types of networks can be distinguished in the spread of misinformation posts: radiation, sector, and viral. Unlike rumor-spreading on Twitter, in which the news is usually first posted by a low-impact user and then shared by some popular users [26], the majority of COVID-19 misinformation on Weibo was represented by the radiation dissemination network, in which the messages were first posted by a prominent user and then directly shared by many general users. In addition, the original user tended to have higher authority (public organizations and news media), suggesting the crucial role of influential users in the spread of COVID-19 misinformation; this is consistent with the results of Wang et al [47], who found that Donald Trump's tweets potentially influenced people's information-sharing behavior.

Limitations

This study has several limitations. First, we only examined misinformation about COVID-19 circulating on Weibo. In addition, we selected "Novel Coronavirus (新冠病毒)/COVID/Epidemic (疫情)" as COVID-19–related keywords. However, due to the potential early inconsistency in disease terminology, users may have used other keywords that were not collected by this study, such as 武汉肺炎 (Wuhan pneumonia) and 不明原因肺炎 (unknown-cause pneumonia), to describe COVID-19–related conversations or topics. Therefore, the characteristics identified in our study may not represent all COVID-19–related misinformation. Future studies should consider misinformation on other social media platforms to ascertain the stability of these findings. Second, we focused on Chinese-language misinformation. Misinformation in other languages about the pandemic could lead to different results, which should also be explored in future work.

Conclusions

In the COVID-19 pandemic, we witnessed a massive infodemic in which fake news and conspiracy theories were spread, especially on social media. This study provides a comprehensive

XSL•FO
RenderX

examination of the COVID-19 misinformation spread on a social media platform.

The theoretical contributions of this study lie in the following two aspects. Although efforts have been made to analyze the COVID-19–related misinformation on social media platforms, no comprehensive analytical framework guided by psychological theory exists to study such misinformation, particularly related to COVID-19. Based on the ELM, this work provides a first step toward understanding the underlying persuasion process of COVID-19–related misinformation. By developing a theoretical model of the persuasion process, this study includes a comprehensive set of features to understand the spread of COVID-19–related misinformation on social media. Moreover, whereas previous studies have generally considered the detection of pandemic misinformation as a binary classification problem, our results show that misinformation on different topics appears to have different characteristics in terms of emotion, social engagement metrics, and publisher authority characteristics. Therefore, this study suggests that the development of misinformation detection algorithms and prevention mechanisms should consider the specific topics of misinformation. It is necessary to develop targeted strategies based on the characteristics of misinformation on different topics.

The practical contributions of this study are two-fold. First, although COVID-19–related misinformation has been widely studied, to our knowledge, no research has attempted to uncover the comprehensive characteristics of users who post misinformation about the novel coronavirus on social media. Therefore, this study examined both the profile features and the interactive characteristics of misinformation authors. By revealing the characteristics of misinformation publishers, our results not only extend the research on analyzing COVID-19–related misinformation but also provide a possible solution to the issue of detecting suspicious users who may be prone to posting misinformation. Moreover, the significant positive correlations among the authority features of the users and the topological attributes of the dissemination network indicate the possible influence of authority features on the spread of misinformation. To combat misinformation, our results suggest that it is important for influential users, public organizations, and news media to be aware of their responsibility to provide verified information, especially during a public health crisis.

Acknowledgments

This study was supported in part by the National Natural Science Foundation of China (72004091, 72174083) and the Humanity and Social Science Foundation of Ministry of Education of China (20YJC870014).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Table S1. Summary of previous research on the classification of coronavirus-related misinformation on social media.
[DOCX File , 17 KB - jmir_v24i6e37623_app1.docx]

References

1. Cuan-Baltazar JY, Muñoz-Perez MJ, Robledo-Vega C, Pérez-Zepeda MF, Soto-Vega E. Misinformation of COVID-19 on the internet: infodemiology study. JMIR Public Health Surveill 2020 Apr 09;6(2):e18444 [FREE Full text] [doi: 10.2196/18444] [Medline: 32250960]
2. Jiang S. Infodemic study on the spread of and response to rumors about COVID-19. Studies Science Popularization 2020;15(1):70-78. [doi: 10.19293/j.cnki.1673-8357.2020.01.011]
3. Kouzy R, Abi Jaoude J, Kraitem A, El Alam MB, Karam B, Adib E, et al. Coronavirus goes viral: quantifying the COVID-19 misinformation epidemic on Twitter. Cureus 2020 Mar 13;12(3):e7255. [doi: 10.7759/cureus.7255] [Medline: 32292669]
4. Mian A, Khan S. Coronavirus: the spread of misinformation. BMC Med 2020 Mar 18;18(1):89 [FREE Full text] [doi: 10.1186/s12916-020-01556-3] [Medline: 32188445]
5. Petty R, Cacioppo J. The elaboration likelihood model of persuasion. Adv Exp Soc Psychol 1986;19:123-205. [doi: 10.1007/978-1-4612-4964-1_1]
6. Ebnali M, Kian C. Nudge users to healthier decisions: A design approach to encounter misinformation in health forums. 2019 Presented at: AHFE 2019 International Conference on Applied Human Factors in Communication of Design; July 24-28, 2019; Washington, DC. [doi: 10.1007/978-3-030-20500-3_1]
7. Zhao Y, Da J, Yan J. Detecting health misinformation in online health communities: incorporating behavioral features into machine learning based approaches. Inf Process Manage 2021 Jan;58(1):102390. [doi: 10.1016/j.ipm.2020.102390]
8. Ceron W, de-Lima-Santos M, Quiles MG. Fake news agenda in the era of COVID-19: identifying trends through fact-checking content. Online Soc Netw Media 2021 Jan;21:100116. [doi: 10.1016/j.osnem.2020.100116]
9. Song Y, Ko L, Jang SH. The South Korean government's response to combat COVID-19 misinformation: analysis of "Fact and Issue Check" on the Korea Centers for Disease Control and Prevention website. Asia Pac J Public Health 2021 Jul 06;33(5):620-622. [doi: 10.1177/10105395211014705] [Medline: 33955279]

XSL•FO
RenderX

10. Simons H. Persuasion: understanding, practice, and analysis. Melbourne, Australia: Addison-Wesley Publishing Company; 1976.
11. Cialdini R. Interpersonal influence. In: Shavitt S, Brock TC, editors. Persuasion: psychological insights and perspectives. Needham Heights, MA: Allyn & Bacon; 1994:195-217.
12. Cialdini RB. Influence: the psychology of persuasion. New York, NY: Harper Business; 2006.
13. Pornpitakpan C. The persuasiveness of source credibility: a critical review of five decades' evidence. J Appl Social Pyschol 2004 Feb;34(2):243-281. [doi: 10.1111/j.1559-1816.2004.tb02547.x]
14. Li L, Tian J, Zhang Q, Zhou J. Influence of content and creator characteristics on sharing disaster-related information on social media. Inf Manage 2021 Jul;58(5):103489. [doi: 10.1016/j.im.2021.103489]
15. Turcotte J, York C, Irving J, Scholl RM, Pingree RJ. News recommendations from social media opinion leaders: effects on media trust and information seeking. J Comput-Mediat Comm 2015 Jun 01;20(5):520-535. [doi: 10.1111/jcc4.12127]
16. Suh B, Hong L, Pirolli P, Chi E. Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. 2010 Presented at: SOCIALCOM '10: 2010 IEEE Second International Conference on Social Computing; August 20-22, 2010; Washington, DC. [doi: 10.1109/socialcom.2010.33]
17. Avram M, Micallef N, Patil S, Menczer F. Exposure to social engagement metrics increases vulnerability to misinformation. HKS Misinfo Review 2020 Jul 25;1(5):1-11. [doi: 10.37016/mr-2020-033]
18. Dainton M, Zelley ED. Applying communication theory for professional life: a practical introduction. Thousand Oaks, CA: SAGE Publications, Inc; 2010.
19. Bode L, Vraga EK. In related news, that was wrong: the correction of misinformation through related stories functionality in social media. J Commun 2015 Jun 23;65(4):619-638. [doi: 10.1111/jcom.12166]
20. Zhou C, Li K, Lu Y. Linguistic characteristics and the dissemination of misinformation in social media: The moderating effect of information richness. Inf Process Manage 2021 Nov;58(6):102679. [doi: 10.1016/j.ipm.2021.102679]
21. Chen S, Xiao L, Mao J. Persuasion strategies of misinformation-containing posts in the social media. Inf Process Manage 2021 Sep;58(5):102665. [doi: 10.1016/j.ipm.2021.102665]
22. Castillo C, Mendoza M, Poblete B. Information credibility on Twitter. 2011 Presented at: 20th International Conference on World Wide Web; March 28-April 1, 2011; Hyderabad, India. [doi: 10.1145/1963405.1963500]
23. van der Meer TGLA, Jin Y. Seeking formula for misinformation treatment in public health crises: the effects of corrective information type and source. Health Commun 2020 May 14;35(5):560-575. [doi: 10.1080/10410236.2019.1573295] [Medline: 30761917]
24. Kwon S, Cha M, Jung K. Rumor detection over varying time windows. PLoS One 2017;12(1):e0168344 [FREE Full text] [doi: 10.1371/journal.pone.0168344] [Medline: 28081135]
25. Kwon S, Cha M, Jung K, Chen W, Wang Y. Prominent features of rumor propagation in online social media. 2013 Presented at: 2013 IEEE 13th International Conference on Data Mining; December 7-10, 2013; Dallas, TX. [doi: 10.1109/icdm.2013.61]
26. Ma J, Gao W, Wong K. Detect rumors in microblog posts using propagation structure via kernel learning. 2017 Presented at: 55th Annual Meeting of the Association for Computational Linguistics; July 2017; Vancouver, BC. [doi: 10.18653/v1/p17-1066]
27. Chen D, Tang S. Analysis of circulating characteristics of rumors on Weibo in public health emergencies: A case study of COVID-19 epidemic. Studies on Science Popularization 2021;16(01):98. [doi: 10.19293/j.cnki.1673-8357.2021.01.005]
28. 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]
29. Qin Y. Analysis of the characteristics of health rumors in public health emergencies: Taking the Shuanghuanglian incident during the COVID-19 as an example. J News Res 2020;11(20):104-115.
30. Pennycook G, Rand DG. The psychology of fake news. Trends Cogn Sci 2021 May;25(5):388-402 [FREE Full text] [doi: 10.1016/j.tics.2021.02.007] [Medline: 33736957]
31. China Internet Joint Rumor-refuting platform. URL: https://www.piyao.org.cn/2020yqpy/ [accessed 2020-12-01]
32. Tencent Fact Checking Platform. URL: https://vp.fact.qq.com/home [accessed 2020-12-01]
33. Weibo Report Processing Hall. URL: https://service.account.weibo.com/ [accessed 2020-12-01]
34. Hu W, Chen H, Wang Q. Psychological motivations and countermeasures of rumors in COVID-19 epidemic. J Dali Univ 2021;6(1):111-116. [doi: 10.3969/j.issn.2096-2266.2021.01.018]
35. Wang K. Content analysis of rumors related to the Novel Coronavirus epidemic and reflections on governance--Analysis of 368 samples based on Nvivo 11. J Northeast Agricult Univ (Social Science Edition) 2020;18(05):27-35.
36. Yao A, Ma J, Lin Y, Liu Q, Zhang Z. The rumor evolution and governance strategy of major public health emergencies. Inf Sci 2020;38(07):22-29. [doi: 10.13833/j.issn.1007-7634.2020.07.004]
37. Richards KAR, Hemphill MA. A practical guide to collaborative qualitative data analysis. J Teach Phys Educ 2018 Apr;37(2):225-231. [doi: 10.1123/jtpe.2017-0084]
38. Krippendorff K. Content analysis: an introduction to its methodology. Newbury Park, CA: SAGE Publishing; 2012.
39. Zapf A, Castell S, Morawietz L, Karch A. Measuring inter-rater reliability for nominal data - which coefficients and confidence intervals are appropriate? BMC Med Res Methodol 2016 Aug 05;16(1):93 [FREE Full text] [doi: 10.1186/s12874-016-0200-9] [Medline: 27495131]

XSL•FO
RenderX

40. Rout JK, Singh S, Jena SK, Bakshi S. Deceptive review detection using labeled and unlabeled data. Multimed Tools Appl 2016 Aug 19;76(3):3187-3211. [doi: 10.1007/s11042-016-3819-y]
41. Tian H, Gao C, Xiao X, Liu H, He B, Wu H, et al. SKEPntiment knowledge enhanced pre-training for sentiment analysis. arXivcs. URL: http://arxiv.org/abs/2005.05635 [accessed 2022-05-09]
42. Zhu Y, Cao L, Xie J, Yu Y, Chen A, Huang F. Using social media data to assess the impact of COVID-19 on mental health in China. Psychol Med 2021 Apr 20:1-8. [doi: 10.1017/S0033291721001598] [Medline: 33875036]
43. Chen Y, Wang T, Zhu S, Lian P. Will you miss me if I am leaving? Unexpected market withdrawal of Norwegian Joy and customer satisfaction. Tourism Management 2020 Feb;76:103951. [doi: 10.1016/j.tourman.2019.103951]
44. Xu Q, Shen Z, Shah N, Cuomo R, Cai M, Brown M, et al. Characterizing Weibo social media posts From Wuhan, China during the early stages of the COVID-19 pandemic: qualitative content analysis. JMIR Public Health Surveill 2020 Dec 07;6(4):e24125 [FREE Full text] [doi: 10.2196/24125] [Medline: 33175693]
45. Barter RL, Yu B. Superheat: An R package for creating beautiful and extendable heatmaps for visualizing complex data. J Comput Graph Stat 2018;27(4):910-922. [doi: 10.1080/10618600.2018.1473780] [Medline: 30911216]
46. Kozinets RV. E-tribalized marketing?: the strategic implications of virtual communities of consumption. Eur Manag J 1999 Jun;17(3):252-264. [doi: 10.1016/s0263-2373(99)00004-3]
47. Wang X, Zhang M, Fan W, Zhao K. Understanding the spread of COVID-19 misinformation on social media: The effects of topics and a political leader's nudge. J Assoc Inf Sci Technol 2021 Sep 27;73(5):726-737. [doi: 10.1002/asi.24576] [Medline: 34901312]
48. Usher K, Durkin J, Martin S, Vanderslott S, Vindrola-Padros C, Usher L, et al. Public sentiment and discourse on domestic violence during the COVID-19 pandemic in Australia: analysis of social media posts. J Med Internet Res 2021 Oct 01;23(10):e29025 [FREE Full text] [doi: 10.2196/29025] [Medline: 34519659]

Abbreviations

ELM: elaboration likelihood model

Edited by M Gisondi, M Gottlieb; submitted 28.02.22; peer-reviewed by Q Xu, W Ceron, N Chalghaf, T Lin; comments to author 19.04.22; revised version received 10.05.22; accepted 07.06.22; published 20.06.22.

Please cite as:
Zhao Y, Zhu S, Wan Q, Li T, Zou C, Wang H, Deng S
Understanding How and by Whom COVID-19 Misinformation is Spread on Social Media: Coding and Network Analyses
J Med Internet Res 2022;24(6):e37623
URL: https://www.jmir.org/2022/6/e37623
doi:10.2196/37623
PMID:35671411

©Yuehua Zhao, Sicheng Zhu, Qiang Wan, Tianyi Li, Chun Zou, Hao Wang, Sanhong Deng. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.06.2022. 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

COVID-19 Vaccine Fact-Checking Posts on Facebook: Observational Study

Haoning Xue[1], MA; Xuanjun Gong[1], MA, MSc; Hannah Stevens[1], MA

Department of Communication, University of California, Davis, Davis, CA, United States

Corresponding Author:
Haoning Xue, MA
Department of Communication
University of California, Davis
One Shields Avenue
Davis, CA, 95616
United States
Phone: 1 5303048532
Email: hnxue@ucdavis.edu

Abstract

Background: Effective interventions aimed at correcting COVID-19 vaccine misinformation, known as fact-checking messages, are needed to combat the mounting antivaccine infodemic and alleviate vaccine hesitancy.

Objective: This work investigates (1) the changes in the public's attitude toward COVID-19 vaccines over time, (2) the effectiveness of COVID-19 vaccine fact-checking information on social media engagement and attitude change, and (3) the emotional and linguistic features of the COVID-19 vaccine fact-checking information ecosystem.

Methods: We collected a data set of 12,553 COVID-19 vaccine fact-checking Facebook posts and their associated comments (N=122,362) from January 2020 to March 2022 and conducted a series of natural language processing and statistical analyses to investigate trends in public attitude toward the vaccine in COVID-19 vaccine fact-checking posts and comments, and emotional and linguistic features of the COVID-19 fact-checking information ecosystem.

Results: The percentage of fact-checking posts relative to all COVID-19 vaccine posts peaked in May 2020 and then steadily decreased as the pandemic progressed (r=–0.92, df=21, t=–10.94, 95% CI –0.97 to –0.82, P<.001). The salience of COVID-19 vaccine entities was significantly lower in comments (mean 0.03, SD 0.03, t=39.28, P<.001) than in posts (mean 0.09, SD 0.11). Third-party fact checkers have been playing a more important role in more fact-checking over time (r=0.63, df=25, t=4.06, 95% CI 0.33-0.82, P<.001). COVID-19 vaccine fact-checking posts continued to be more analytical (r=0.81, df=25, t=6.88, 95% CI 0.62-0.91, P<.001) and more confident (r=0.59, df=25, t=3.68, 95% CI 0.27-0.79, P=.001) over time. Although comments did not exhibit a significant increase in confidence over time, tentativeness in comments significantly decreased (r=–0.62, df=25, t=–3.94, 95% CI –0.81 to –0.31, P=.001). In addition, although hospitals receive less engagement than other information sources, the comments expressed more positive attitudinal valence in comments compared to other information sources (b=0.06, 95% CI 0.00-0.12, t=2.03, P=.04).

Conclusions: The percentage of fact-checking posts relative to all posts about the vaccine steadily decreased after May 2020. As the pandemic progressed, third-party fact checkers played a larger role in posting fact-checking COVID-19 vaccine posts. COVID-19 vaccine fact-checking posts continued to be more analytical and more confident over time, reflecting increased confidence in posts. Similarly, tentativeness in comments decreased; this likewise suggests that public uncertainty diminished over time. COVID-19 fact-checking vaccine posts from hospitals yielded more positive attitudes toward vaccination than other information sources. At the same time, hospitals received less engagement than other information sources. This suggests that hospitals should invest more in generating engaging public health campaigns on social media.

(J Med Internet Res 2022;24(6):e38423) doi:10.2196/38423

KEYWORDS

COVID-19 vaccine; fact checking; misinformation correction; sentiment analysis; social media; COVID-19; vaccination; misinformation; health information; online information; infodemic; public sentiment

XSL•FO
RenderX

Introduction

Background

As of May 4, 2022, the novel COVID-19 outbreak had caused 994,551 deaths and 81,574,159 cases in the United States [1]. Compared to COVID-19 deaths per capita in developed countries with similarly aged populations (ie, the United Kingdom, France, Spain, Canada), the United States has the highest number of deaths per 100,000 people [1]. This may be because, despite widespread vaccine availability, the United States has the lowest rate of individuals who are fully vaccinated and boosted (30%) compared to developed countries with similarly aged populations (eg, 52% in Canada). Indeed, those who haven't received all 3 doses of the vaccine account for the majority of deaths and severe cases in the United States [2-4]. Furthermore, vaccine hesitancy has constrained public health officials' efforts to mitigate the pandemic through herd immunity [5,6].

Thus, effectively communicating the necessity of getting the COVID-19 vaccine is essential to mitigating the COVID-19 pandemic. Although some public officials have endorsed the COVID-19 vaccine, others have fostered vaccine hesitancy by broadcasting misinformation (ie, inaccurate health information), which is often disseminated widely via social media [7,8]. The prevalence of US adults who indicate they primarily get news information from social media (ie, 68%) [9] has given rise to an *infodemic*, wherein public confidence in the COVID-19 vaccine is shaken by the overload of COVID-19 misinformation on social media [10-12]. Indeed, Loomba et al found that exposure to COVID-19 vaccine misinformation significantly decreases the intention to receive the COVID-19 vaccine [13]. Thus, effective interventions aimed at correcting COVID-19 vaccine misinformation, known as fact-checking messages, are needed to combat the mounting antivaccine infodemic and alleviate vaccine hesitancy [14,15].

Research efforts have therefore focused on experimentally testing the efficacy of COVID-19 vaccine misinformation fact-checking messages, finding that accurate misinformation correction messages can effectively mitigate health misinformation in certain contexts, namely when the message is from a credible information source (ie, health institutions, research institutions, and news media) [16-19]. Thus, although we know fact checking from credible sources can be effective, the extent to which credible information sources share fact-checking information and the ways in which the public engages with COVID-19 vaccine fact checks in naturalistic social media environments remain neglected in the literature. This work aims to fill this gap by investigating (1) the changes in the public's attitude toward COVID-19 vaccines over time, (2) the effectiveness of COVID-19 vaccine fact-checking information on social media engagement and attitude change, and (3) the emotional and linguistic features of the COVID-19 vaccine fact-checking information ecosystem. This study expands our knowledge of the COVID-19 vaccine information environment on social media and contributes to our understanding of the effectiveness of COVID-19 vaccine fact-checking messages on the public's attitudes toward the COVID-19 vaccine. A novel contribution of this study lies in the usage of entity-targeted sentiment powered by Google Cloud Natural Language AI (where AI is artificial intelligence) that enables us to capture the exact attitude toward the COVID-19 vaccine among the public [20].

The Public's Attitude Toward COVID-19 Vaccines

Vaccine attitude determines people's intention to vaccinate and the consequential vaccine uptake behaviors [21,22]. However, exposure to misinformation can cause a decline in people's vaccination intention to receive COVID-19 vaccines in both the United Kingdom and the United States [13]. Thus, to assess how effective vaccination campaigns are against COVID-19 misinformation, such as fact-checking messages on social media, it is important to investigate the actual impact on the public's attitude toward the COVID-19 vaccine. Previous experimental studies have shown that fact-checking interventions can promote people's positive attitude toward vaccines and increase the accuracy of beliefs about vaccination [12,23]. However, empirical observational evidence for fact-checking messages' effects on the public's attitude toward the COVID-19 vaccine in the real world is still lacking. In this study, we explore this question by examining how the public's attitude toward the COVID-19 vaccine, as reflected by attitudinal linguistic markers related to COVID-19 vaccine–related entities in comments attached to fact-checking posts, changes as a response to fact-checking posts over time. Accordingly, we ask research question (RQ)1a: *How does the attitude toward the COVID-19 vaccine in fact-checking comments change over time?*

In addition, it is also important to investigate the attitude toward COVID-19 vaccine fact-checking messages itself. Emotionally charged messages are found to influence vaccination intent more than facts and statistics [24], and discrete emotions impact vaccine behaviors differently [25]. A 2020 study found that emotionally positive COVID-19 health messages predict compliance with COVID-19 public health guidelines when the messages evoke highly positive responses [26]. In contrast, desensitization to emotionally charged, negatively valenced COVID-19 messages can prompt folks to become disengaged, unmotivated to take protective action, and susceptible to health misinformation [27-29]. Thus, the extent to which credible health sources broadcast positively valenced, emotionally charged messages is consequential; accordingly, we pose the following RQ1b: *How does the attitude toward the COVID-19 vaccine in fact-checking posts change over time?*

Effects of Fact-Checking Information Sources

We know politicians are a prevalent source of vaccine misinformation on social media (Featherstone et al. [17-19]), and the credibility of the source of vaccine information can determine fact-checking message efficacy. Indeed, health information from sources with authority (eg, health institutions) are perceived as more credible [30]. However, the extent to which credible information sources share misinformation corrections and the ways in which the public engages with COVID-19 vaccine fact-checking messages in naturalistic social media environments remain unclear. Thus, we pose the following RQ2: *How do different information sources of COVID-19 vaccine fact-checking posts influence (1) the public's*

XSL•FO
RenderX

attitude toward the COVID-19 vaccine and (2) social media engagement with fact-checking posts?

Emotional Trends and Linguistic Features in COVID-19 Fact-Checking Posts

In addition to the public's valenced attitude toward the COVID-19 vaccine, varied discrete emotions may reveal more about the specific attitudes or concerns. Different discrete emotions have different effects on the vaccine-hesitant. For example, vaccine-hesitant users are more likely than provaccine users to express anger in posts and replies [31]. For health promotion information, heightened anxiety in protective health messages can encourage the hesitant to vaccinate [32]. Thus, we are interested in the specific discrete emotions that emerge in posts and comments over time. Thus, we posit RQ3: *Which discrete emotions manifest in COVID-19 vaccine fact-checking posts and comments over time?*

Although staunch antivaxxers exist, a prominent group of Americans who understand vaccine importance are hesitant to take it because of uncertainty about the safety of the vaccine's rushed development [33]. Thus, correcting misinformation with confident messages (ie, low tentativeness, high certainty) is an essential component of restoring public trust in the vaccine's safety and efficacy. Furthermore, health threat messages with logical, actionable steps for alleviating the health threat can encourage protective behaviors, such as encouraging vaccination [32]. Yet, the extent to which credible sources of health information projected confidence and logic in messages throughout the pandemic is unclear, particularly at the beginning of the vaccine rollout, when even health experts were uncertain about aspects of the vaccine (eg, how the vaccine works in those with COVID-19 mortality risk factors, which vaccine is most effective) [34]. In addition, how such subtle linguistic features are present in average social media users' remarks informs us how the public's attitude toward the COVID-19 vaccines has evolved over time. As such, we raise RQ4: *How do linguistic features of confidence, tentativeness, and analytical thinking in the COVID-19 vaccine fact-checking posts and comments vary over time?*

Methods

Data Set

To fill these gaps in the literature, we collected a data set of 12,553 COVID-19 vaccine fact-checking posts and their associated comments (N=122,362) from January 2020 to March 2022. Facebook was selected because it is 1 of the most popular social media platforms worldwide with a significant presence of both misinformation and fact-checking information [35].

Collecting Facebook Posts Using CrowdTangle

We leveraged Meta's *CrowdTangle* tool to (1) identify relevant sources of COVID-19 vaccine information and (2) collect Facebook posts related to the COVID-19 vaccine created between January 1, 2020, and March 10, 2022 [36]. CrowdTangle is a data-tracking platform owned by Meta (Facebook's parent company) that monitors social media public conversations and related data.

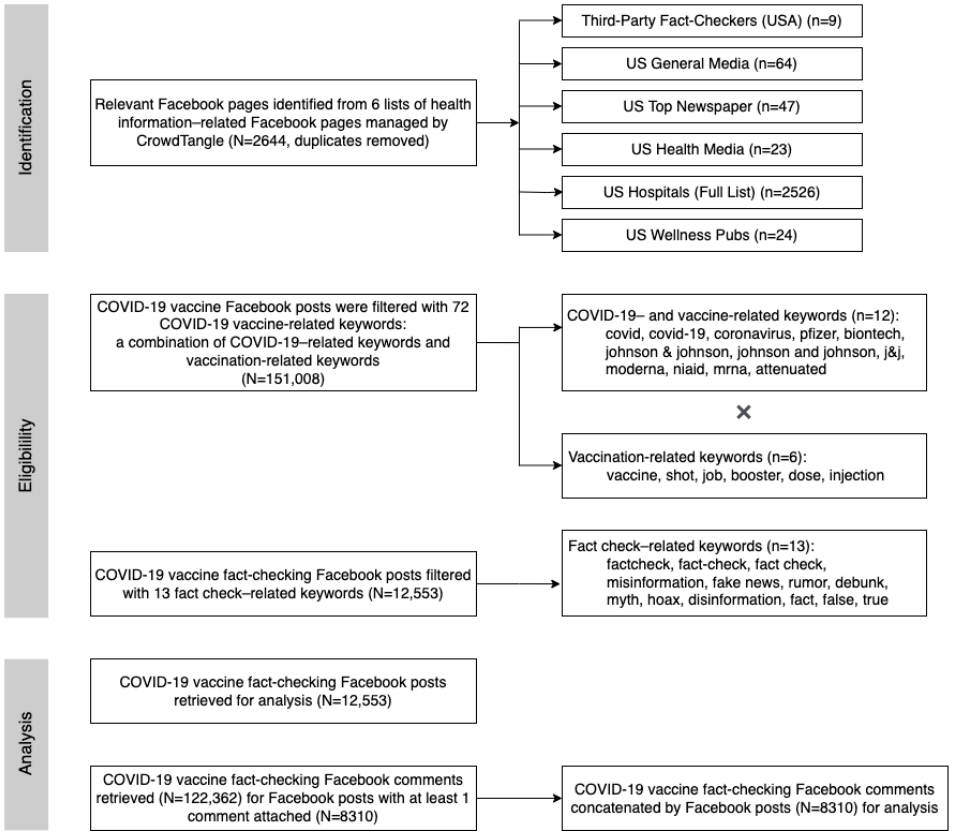
CrowdTangle tags public Facebook pages based on several attributes, including the primary language of the content, the country the content is geared toward, and the type of entity that owns the page (eg, health influencer, top newspaper). We curated a list of pages belonging to categories related to health information sharing, namely third-party fact checkers, general media sources, top newspaper sources, health influencers, health media, hospitals, and wellness publications (Figure 1). To keep the framework parsimonious, the 6 categories were further aggregated into 4 category types: (1) US news media (including US general media and the top newspaper), (2) third-party fact checkers, (3) US health media (including US health media and wellness publications), and (4) US hospitals. Only English Facebook pages geared toward US audiences were retained. In total, there were 2644 unique pages obtained from the categories, with duplicates removed (n=49).

Second, we mined the total number of COVID-19 vaccine–related posts (N=151,008) and the accompanying post metadata provided with Crowdtangle (ie, date, number of shares, comments, and emoticon reactions) from the curated list of health-related Facebook pages (N=2644) [37]. Posts were retained if they contained both (1) at least 1 COVID-19– or vaccine-related keyword (eg, "mRNA," "coronavirus") and (2) at least 1 vaccine-related keyword (eg, "vaccines," "booster," "dose") in either the body of the post or a Uniform Resource Locator (URL) shared in the post. See Table S1 in Multimedia Appendix 1 for an overview.

Further, we extrapolated posts containing at least 1 (8%) of 13 fact-checking keywords (eg, "debunk," "hoax") to distill fact-checking posts (N=12,553) from the larger data set (Multimedia Appendix 1, Table S2). COVID-19 vaccine fact-checking posts (N=12,553) from 1226 different Facebook pages were retained for further analysis.

XSL•FO
RenderX

Figure 1. Process to identify relevant lists of Facebook pages, collect Facebook posts and comments, and filter relevant Facebook posts.



Collecting Facebook Comments Using Facepager

Although CrowdTangle makes some metadata available (ie, post date, number of shares, and number of Facebook emoticon reactions), it does not provide access to comment data. We used an automatic data collection software called *Facepager* to retrieve up to 25 of the highest-ranked comments (ie, "top 25 comments") attached to each Facebook post in the data set [37,38]. We retrieved 122,362 comments associated with 12,553 COVID-19 vaccine fact-checking posts. Comments attached to a post were concatenated as a single textual observation; not all posts had comments, leaving us with 8310 comment threads for further analysis.

Measures

Attitude Toward COVID-19 Vaccines: Google Cloud Natural Language AI

We used Google Cloud Natural Language AI, a machine learning-based natural language–understanding tool to retrieve the public's attitude toward the COVID-19 vaccine by (1) identifying all entities that were discussed in a given post or comment, (2) using COVID-19 vaccine keywords to distill entities specifically related to the COVID-19 vaccine (eg, COVID-19 vaccine, Pfizer Booster, etc; see Table S3 in Multimedia Appendix 1 for a full list), and (3) measuring attitudes toward each COVID-19 vaccine–related entity. We extracted 23,636 distinct entities from fact-checking posts and 71,418 entities from the comments [20]. To identify entities specifically related to COVID-19 vaccines (eg, Pfizer), as opposed to off-topic entities (eg, President Donald Trump), we only retained entities containing vaccine-related keywords (Multimedia Appendix 1, Table S3). We retrieved 3014 distinct entities in the posts and 3641 entities in the comments that were related to COVID-19 vaccines.

Specifically, we focused on 3 dimensions of the attitude: COVID-19 vaccine entity salience (ie, the salience of COVID-19 vaccine–related entities of all entities in a given text), attitudinal valence (ie, the positive or negative attitude toward vaccine-related entities), and attitudinal magnitude of each COVID-19 vaccine entity (ie, how strong the attitude is).

COVID-19 vaccine *entity salience* is the extent to which COVID-19 vaccine–related entities are discussed in texts relative to discussions of entities that veer off the topic of COVID-19 vaccines; it reflects the importance of vaccine entities in posts/comments (min.=0, max.=1) [20]. For example, a comment exchange might begin with a single comment about Pfizer vaccine misinformation and then veer off the topic of vaccines to a lengthy debate about voter fraud in the US presidential election; in this case, the extent to which a COVID-19 vaccine–related entity (ie, the Pfizer vaccine) was discussed would be relatively low compared to the extent to which off-topic entities (eg, former President Trump) were.

COVID-19 vaccine *attitudinal valence* was operationalized as a sentiment score provided by Google Cloud Natural Language AI, which indicates the overall emotional and attitudinal valence of a text toward a specific entity, ranging from –1 (extremely negative) to 1 (extremely positive), with 0 representing a neutral attitude [20]. Although emotional valence measures the difference between the positive and negative emotions in texts, it does not capture discrete positive and negative emotions in

XSL•FO
RenderX

texts. In other words, texts that are considered sad or anxious would both be considered negatively valenced texts.

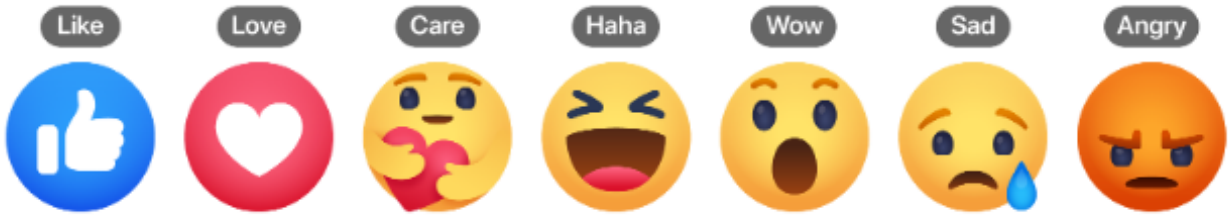
COVID-19 vaccine *attitudinal magnitude* was operationalized as a separate magnitude score that indicates the extent to which a text is emotionally charged, ranging from 0 (neutral) to positive infinity (extremely emotional) [20]. In other words, unlike attitudinal valence, attitudinal magnitude is nonnormalized and each emotionally valenced expression within a given text (irrespective of the direction of the emotional valence) contributes to the attitudinal magnitude revolving around an entity.

Inferences can be made regarding attitude by combining these 2 dimensions. For example, a comment thread that yields a close-to-zero attitudinal valence and close-to-zero attitudinal magnitude indicates that the thread is unemotional. However, a comment thread that yields a close-to-zero attitudinal valence but high attitudinal magnitude suggests the existence of both highly positive and highly negatively valenced attitudes in the comments that cancel each other out [20].

Social Media Engagement

Social media engagement was operationalized as the number of comments and shares each post received, as well as the number of Facebook reactions a given post received. Facebook reactions are a series of 6 emoticons that enable users to express their emotional responses to posts [38]. See Figure 2 for an overview.

Figure 2. Facebook reaction emoticons.



Discrete Emotions and Linguistic Features: IBM Watson Tone Analyzer

We used IBM Watson Tone Analyzer, a classifier of discrete emotions and linguistic features based on cognitive linguistic analysis to extract discrete emotions and linguistic features. In contrast to Google Cloud Natural Language AI [20], IBM Watson Tone Analyzer captures specific positive and negative emotions in a given text (ie, joy, anger, fear, sadness) [39]. Regarding linguistic features, we extracted the levels of confidence, tentativeness, and analytical thinking in texts. These variables of discrete emotions and linguistic features range from 0 to 1, with a larger value representing a stronger existence of an attribute.

Ethics Consideration

This study did not involve human subjects and therefore did not need an institutional review board (IRB) review. The data involved was public data with no identifiable information [40].

Results

Analytical Strategy

To answer RQ1 on the changes of the public's attitude toward the COVID-19 vaccine, we aggregated the COVID-19 vaccine fact-checking posts by time and conducted a series of correlation tests between the attitude variables (ie, vaccine entity salience, attitudinal valence, and magnitude) and time.

To answer RQ2 on the effects of different COVID-19 vaccine fact-checking information sources, we conducted multiple linear regressions with the public's attitude toward the COVID-19 vaccine as the dependent variable and negative binomial regressions with social media engagement as the dependent variable. In these regression models, the word count in the posts and comments, discrete emotions and linguistic features in the posts, and Facebook page followers were controlled.

To answer RQ3 and RQ4 on the changes in discrete emotions and linguistic features over time, we conducted a series of correlation tests between discrete emotions, linguistic features, and time.

The Prevalence of COVID-19 Vaccine Information on Social Media

The number and percentage of COVID-19 fact-checking posts are shown in Figures 3 and 4, respectively. Notably, the percentage of fact-checking posts relative to all COVID-19 vaccine posts steadily decreased as the pandemic progressed (Figure 4).

Interestingly, we found that the 2 peaks of the COVID-19 vaccine posts corresponded with the key time points of COVID-19 vaccination (Figures 3 and 5). This pattern was shown by the number of COVID-19 and fact-checking posts peaking in December 2020, when the first COVID-19 vaccines were available to the public. In addition, the number of posts experienced another major soar in August 2021, when the public started to receive COVID-19 booster shots.

XSL•FO
RenderX

Figure 3. Number of COVID-19 vaccine fact-checking Facebook posts compared to all COVID-19 vaccine–related posts aggregated by month, from January 1, 2020, to March 10, 2022. The line represents the number of new COVID-19 cases, aggregated by month, from January 1, 2020, to March 10, 2022.

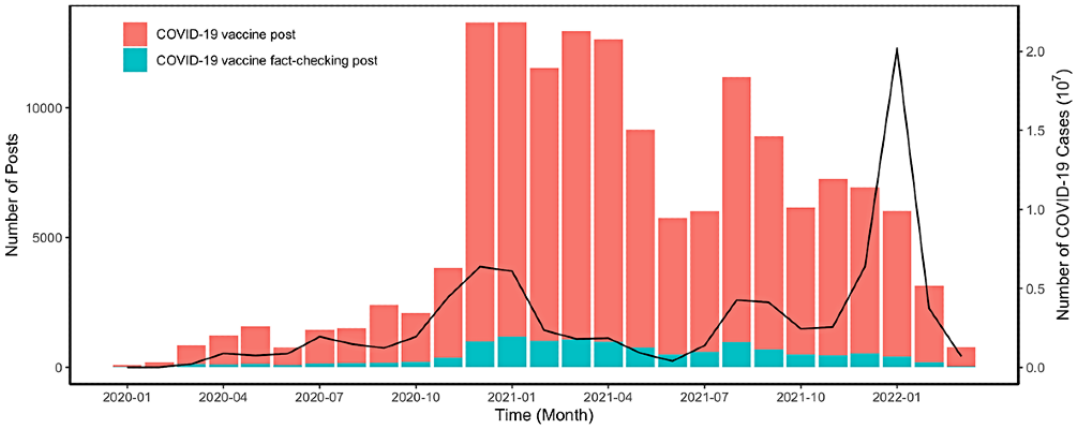


Figure 4. Changes in the percentage of COVID-19 vaccine fact-checking Facebook posts in all COVID-19 vaccine–related Facebook posts over time, aggregated by month, from January 1, 2020, to March 10, 2022.

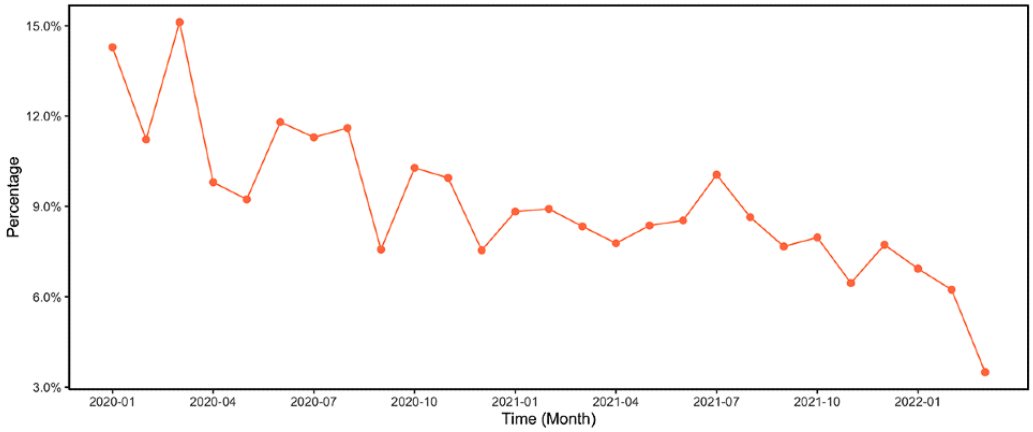
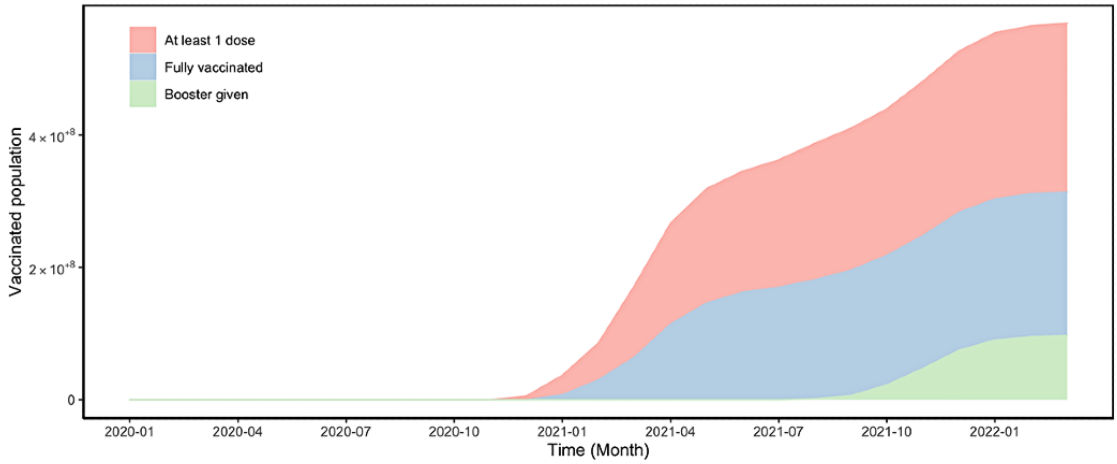


Figure 5. Vaccination population in the United States over time: the number of individuals receiving at least 1 dose, the number of fully vaccinated individuals, and the number of booster shots issued, aggregated by month, from January 1, 2020, to March 10, 2022.



Attitude Toward COVID-19 Vaccine Fact-Checking Posts and Comments Over Time

To answer RQ1, we aggregated vaccine entity salience and attitudinal valence and magnitude by month. The salience of COVID-19 vaccine–related entities relative to off-topic entities is depicted in Figure 6. The salience of COVID-19 vaccine entities was greater in fact-checking posts (mean 0.09, SD 0.11) compared to comments (mean 0.03, SD 0.09, t=39.28, P<.001). COVID-19 vaccine salience in the posts started to increase and peaked around May 2020 and continued to decrease since then (r=–0.92, df=21, t=–10.94, 95% CI –0.97 to –0.82, P<.001). This may reflect less public concern over vaccine misinformation relative to other COVID-19 topics over time.

The mean of COVID-19 vaccine attitudinal valence in posts (mean 0.01, SD 0.10) and comments (mean –0.004, SD 0.12)

XSL•FO
RenderX

was close to 0. Since the average attitudinal magnitude of posts (mean 0.05, SD 0.11) and comments (mean 0.09, SD 0.13) was close to 0, results revealed that the texts were relatively neutral. Notably, COVID-19 vaccine attitudinal magnitude in comments increased during the pandemic (r=0.52, df=25, t=3.02, 95% CI 0.17-0.75, P<.001), which indicates that the public's attitude toward COVID-19 vaccines was becoming more extreme (Figure 7).

Therefore, we noticed a decrease in vaccine entity salience in posts over time (RQ1a) and an increase in attitudinal magnitude in comments over time (RQ1b). Attitudinal valence did not change significantly over time for both posts and comments.

Figure 6. Salience of COVID-19 vaccine–related entities over time, aggregated by month, from January 1, 2020, to March 10, 2022.

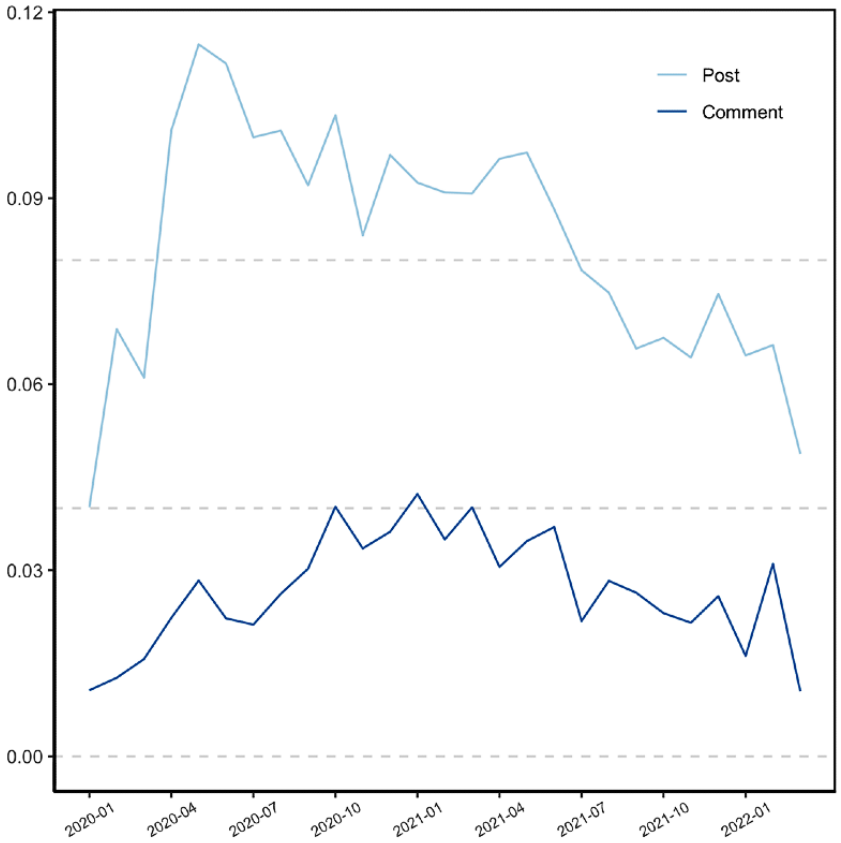


Figure 7. COVID-19 vaccine attitudinal valence and magnitude in COVID-19 vaccine fact-checking posts and comments aggregated by month, from January 1, 2020, to March 10, 2022.

XSL•FO
RenderX

Information Sources of COVID-19 Vaccine Fact-Checking Posts

The majority of COVID-19 vaccine fact-checking Facebook posts were generated by news media (n=5821, 46.4%) and hospitals (n=4921, 39.2%), while relatively fewer posts were posted by third-party fact checkers (n=1523, 12.1%) and US health media (n=288, 2.3%); see Table 1 and Figure 8. Notably, third-party fact checkers have been playing a more important role in more fact-checking over time (r=0.63, df=25, t=4.06, 95% CI 0.33-0.82, P<.001).

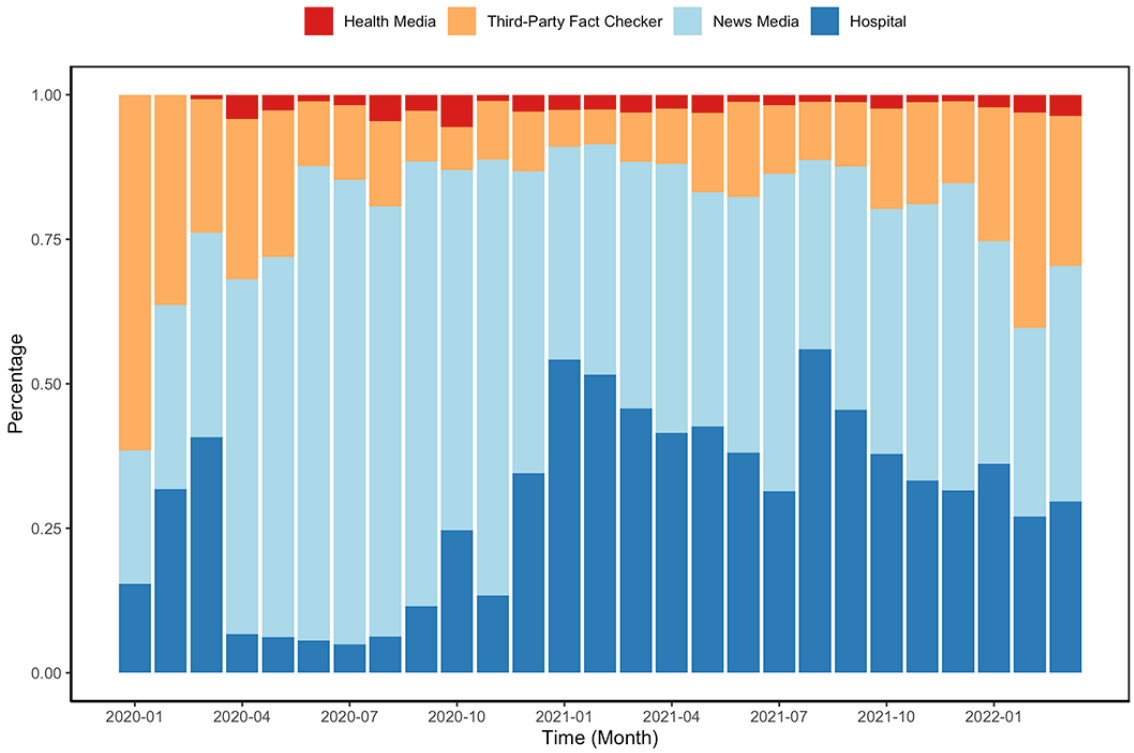
Social media engagement was operationalized as 9 different metrics, namely, the number of comments (mean 146.76, SD 561.07, median 9) and shares each post received (mean 92.28, SD 594.93, median 8), as well as the number of Facebook reactions a given post received (see Table 2 for an overview). Different information sources have different levels of popularity and social media engagement (Table 3).

Health media have more followers than other 3 sources, while hospitals on average have the least number of followers. COVID-19 vaccine fact-checking posts created by news media were most popular, with the highest number of likes, comments, and shares on average, and posts by hospitals were least popular.

Table 1. Summary statistics for sources of COVID-19 fact-checking posts.

Information source	Facebook pages (N=2644), n (%)	Posts (N=12,553), n (%)
News media	95 (3.6)	5821 (46.4)
Hospitals	1096 (41.5)	4921 (39.2)
Third-party fact checkers	9 (0.3)	1523 (12.1)
US health media	26 (1.0)	288 (2.3)

Figure 8. Percentage of COVID-19 vaccine fact-checking Facebook posts by 4 information sources in the United States, aggregated by month, from January 1, 2020, to March 10, 2022.



XSL•FO
RenderX

Table 2. Summary statistics for Facebook reactions.

Reaction metric	Mean (SD)	Median
Like	427.65 (1946.95)	41
Love	42 (346.86)	1
Wow	26.04 (199.94)	0
Haha	64.96 (406)	2
Sad	26.95 (218.49)	0
Angry	76.33 (554.58)	0
Care	3.96 (19.84)	0

Table 3. Average social media engagement of COVID-19 vaccine fact-checking Facebook posts across information sources.

Category	Third-party fact checker	Health media	Hospital	News media
Average followers	538,028	2,193,198	15,639	6,619,852
Posts (N=12,553), n (%)	1523 (12.1)	288 (2.3)	4921 (39.2)	5821 (46.4)
Likes, mean (SD)	231.55 (587.35)	252.07 (1039.86)	24 (162.96)	828.88 (2,774.68)
Comments, mean (SD)	125.1 (228.83)	144.38 (320)	9.39 (150.71)	268.67 (781.04)
Shares, mean (SD)	100.14 (476.54)	125.02 (398.62)	27.07 (650.73)	143.72 (576.06)
Love, mean (SD)	7.66 (58.81)	10.77 (31.86)	3.34 (33.04)	85.22 (504.1)
Wow, mean (SD)	9.7 (33.28)	10.3 (44.35)	0.16 (1.62)	52.96 (290.61)
Haha, mean (SD)	50.08 (184.47)	33.14 (89.59)	1.32 (42.64)	124.23 (581.05)
Sad, mean (SD)	10.67 (42.26)	8.4 (38.94)	0.5 (6.86)	54.49 (317.71)
Angry, mean (SD)	43.56 (194.11)	10.24 (32.22)	0.51 (10.83)	152.27 (801.37)
Care, mean (SD)	1.2 (5.91)	2.75 (10.14)	0.98 (11.66)	7.26 (26.45)

Information Sources and COVID-19 Vaccine Attitude

To answer RQ2a, we conducted multiple linear regressions to investigate the effects of information sources on COVID-19 vaccine attitude, with word counts in the posts and comments and the number of Facebook page followers included as control variables (Multimedia Appendix 1, Tables S13-S15). Results showed that the public's attitudinal valence on COVID-19 vaccines significantly increased with fact-checking posts from hospitals (b=0.06, 95% CI 0.00-0.12, t=2.03, P=.04), though we found no significant effects of information sources on the salience or attitudinal magnitude related to COVID-19 vaccines in the comments. This suggested that hospitals' fact-checking posts on Facebook significantly improved the public's valence attitude toward COVID-19 vaccines.

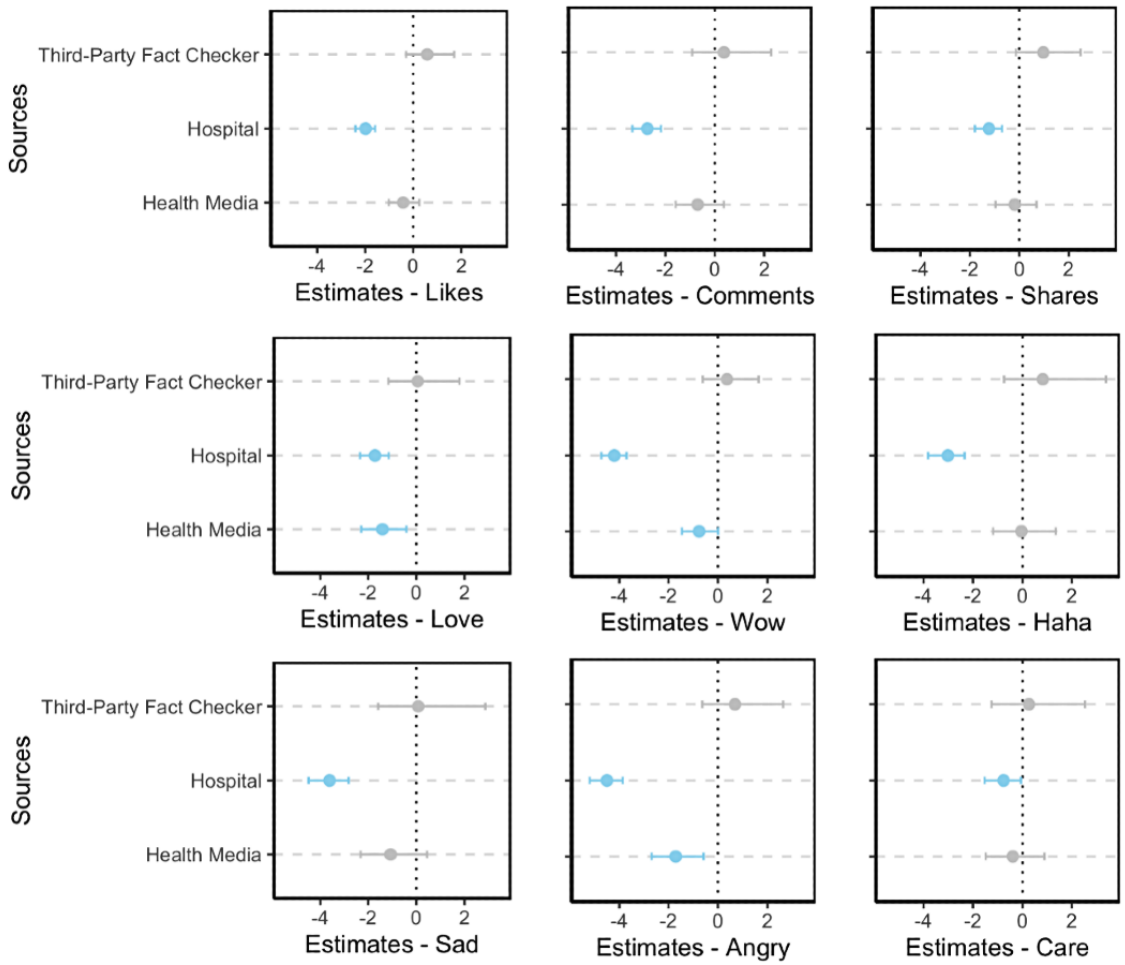
Information Sources and Social Media Engagement

To answer RQ2b, we aggregated fact-checking posts by Facebook page information source and created negative binomial regression models to assess whether the type of information source significantly predicted the 9 metrics of social media engagement, while controlling for post word count and follower count (Multimedia Appendix 1, Tables S4-S12).

Results revealed that hospitals have a significantly lower social media engagement for all engagement metrics than news media. Similarly, health media had significantly fewer wow and angry reactions than news media (P>.05; Figure 9). Results revealed that hospitals also have a smaller audience than other sources of health information. Additionally, although health media and news media posts had similar levels of engagement, health media evoked fewer wow and angry reactions from the public.

XSL•FO
RenderX

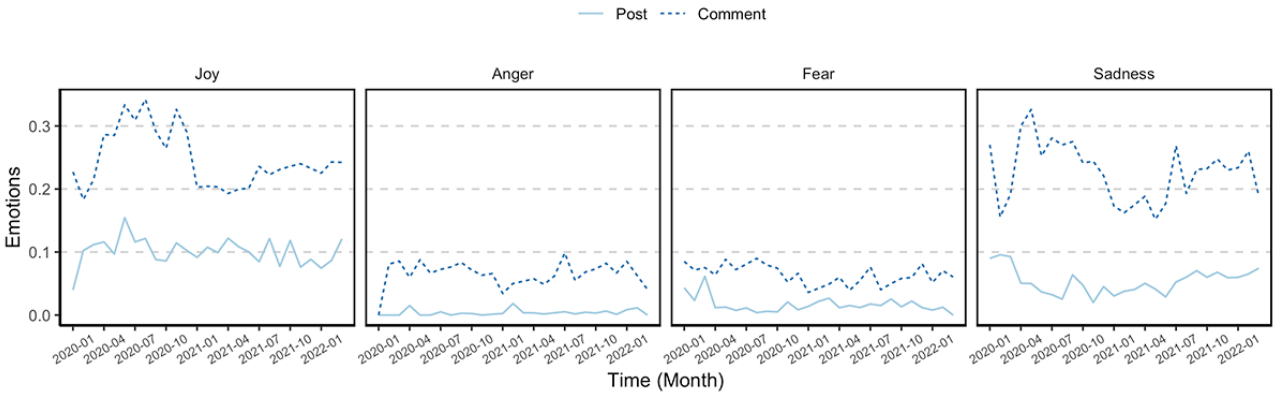
Figure 9. Regression coefficients (95% CI) of information sources with significant effects on social media engagement in negative binomial models. Blue dots and blue error bars show significant coefficients and 95% CIs (P>.05); gray dots and gray error bars show insignificant coefficients and 95% CIs (P>.05). Exact coefficients and P values can be accessed in Multimedia Appendix 1, Tables S4-S12.



Emotional Trends in COVID-19 Vaccine Fact-Checking Posts and Comments

To answer RQ3, we used IBM Watson Tone Analyzer to extract 4 discrete emotions and 3 linguistic features in COVID-19 vaccine fact-checking posts and comments (Figure 10) [39]. Findings revealed that comments were more emotionally charged than fact-checking posts in terms of joy ($mean_{post}$ 0.10, SD_{post} 0.23, $mean_{comment}$ 0.23, $SD_{comment}$ 0.30, t=–38.90), sadness ($mean_{post}$ 0.05, SD_{post} 0.16, $mean_{comment}$ 0.21, $SD_{comment}$ 0.28, t=–56.16), anger ($mean_{post}$ 0.01, SD_{post} 0.05, $mean_{comment}$ 0.06, $SD_{comment}$ 0.18, t=–34.53), and fear ($mean_{post}$ 0.02, SD_{post} 0.10, $mean_{comment}$ 0.06, $SD_{comment}$ 0.17, t=–21.96, P<.001). This suggests that fact-checking posts tended to maintain a neutral tone, whereas the public comments were more emotionally charged. Results also revealed that the dominant emotions in both comments and posts were joy and sadness.

Figure 10. Trends in discrete emotions in COVID-19 vaccine fact-checking posts and comments, aggregated by month, from January 1, 2020, to March 10, 2022.



XSL•FO
RenderX

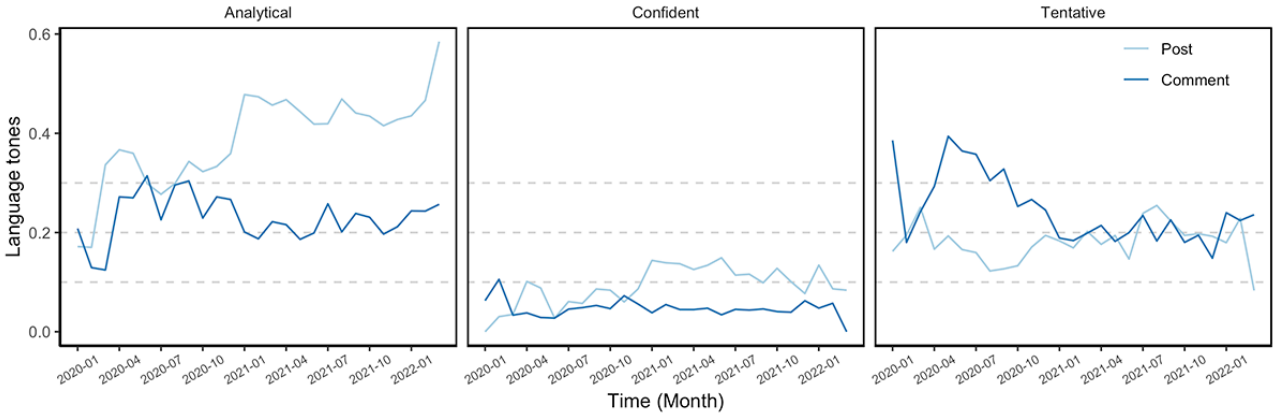
Linguistic Features in COVID-19 Vaccine Fact-Checking Posts and Comments

To answer RQ4, we used IBM Watson Tone Analyzer to extract 3 linguistic feature attributes of COVID-19 vaccine fact-checking posts and comments, namely post analytical thinking, confidence, and tentativeness (Figure 11) [39]. Fact-checking posts were more analytical ($mean_{post}$ 0.43, SD_{post} 0.40, $mean_{comment}$ 0.22, $SD_{comment}$ 0.33, t=44.09), more confident ($mean_{post}$ 0.11, SD_{post} 0.28, $mean_{comment}$ 0.05, $SD_{comment}$ 0.19, t=22.50), and less tentative ($mean_{post}$ 0.19, SD_{post} 0.34, $mean_{comment}$ 0.21, $SD_{comment}$ 0.34, t=–5.19, P<.001) than in the comments.

In addition, linguistic features changed over time (Figure 11). COVID-19 vaccine fact-checking posts continued to be more analytical (r=0.81, df=25, t=6.88, 95% CI 0.62-0.91, P<.001) and more confident (r=0.59, df=25, t=3.68, 95% CI 0.27-0.79, P=.001) over time. This suggests that as public health officials gained more information about the COVID-19 vaccine, they expressed heightened confidence in and reduced tentativeness about the vaccine.

Although comments did not exhibit a significant increase in confidence over time, tentativeness in comments decreased significantly (r=–0.62, df=25, t=–3.94, 95% CI –0.81 to –0.31, P=.001). Results suggested that both public health officials and the public expressed more certain attitudes, in terms of tentativeness and confidence, toward the vaccine as more information became available.

Figure 11. Changes in linguistic features in COVID-19 vaccine fact-checking posts and comments, aggregated by month, from January 1, 2020, to March 10, 2022.



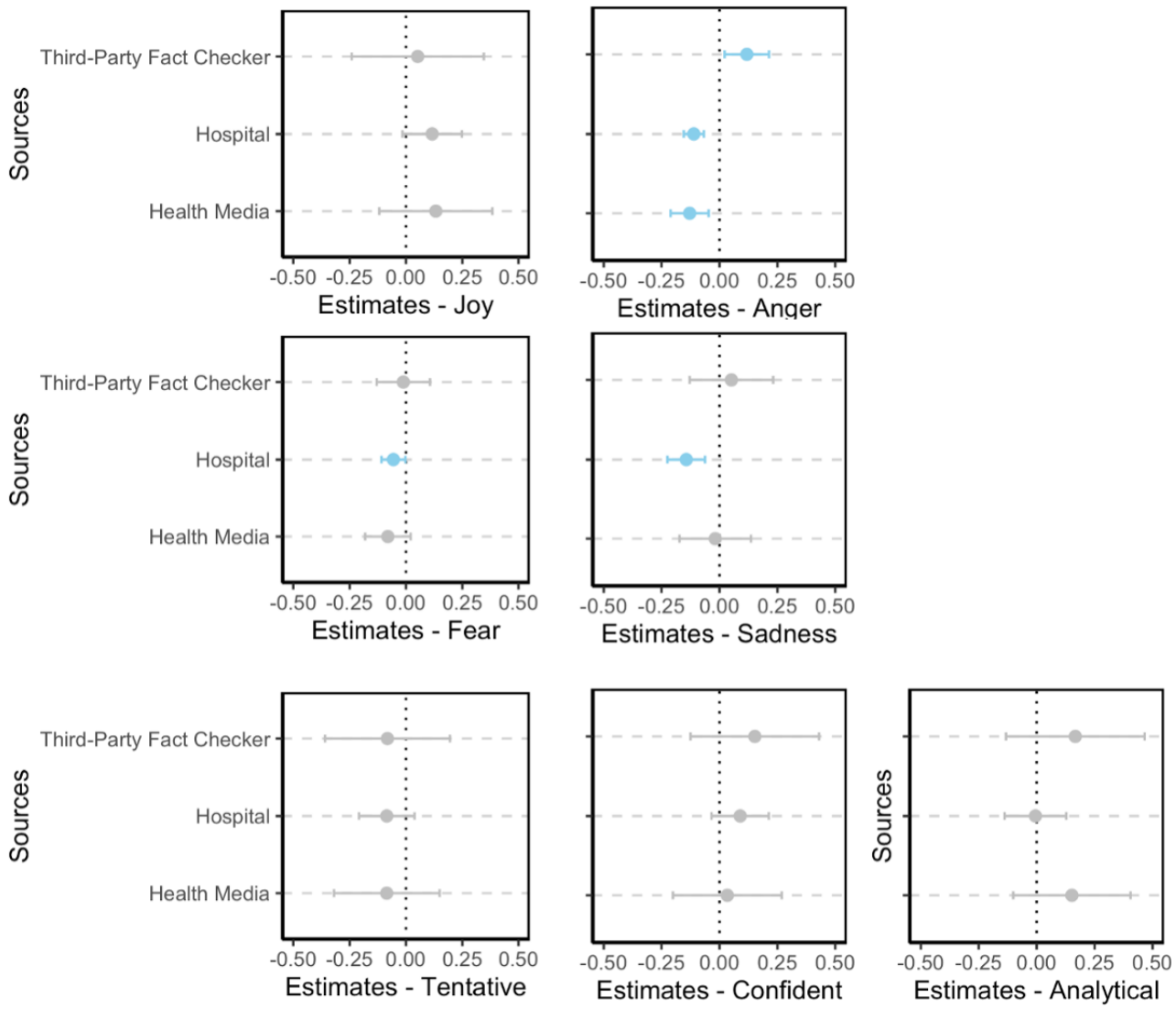
Discrete Emotions and Linguistic Features by Information Sources

In addition to the proposed RQs, we explored how comments might respond differently across information sources regarding discrete emotions and linguistic features with multiple linear regression analyses. The word count of posts and comments, the Facebook page follower count, and discrete emotions and linguistic features of posts were controlled (Multimedia Appendix 1, Tables S16-S19). Regression coefficients are shown in Figure 12.

Results revealed that fact-checking posts from hospitals were associated with lower levels of anger, fear, and sadness in comments, while posts from third-party fact checkers were associated with heightened comment anger. In other words, third party fact checkers tended to evoke heightened comment anger, whereas comments on posts from health media and hospitals expressed less negative emotion. However, no significant effects were found for linguistic features.

XSL•FO
RenderX

Figure 12. Regression coefficients (95% CIs) of information sources with significant effects on emotions and language tones in comments in linear regression models. Blue dots and blue error bars show significant coefficients and 95% CIs (P<.05); gray dots and gray error bars show insignificant coefficients and 95% CIs (P≥.05). Actual coefficients and P values can be accessed in Multimedia Appendix 1, Tables S16-S22.



Discussion

Principal Findings

This study examined the US COVID-19 vaccine fact-checking information on Facebook and analyzed the effects of different fact-checking posts' information sources on the public's attitude toward COVID-19 vaccines and social media engagement. We observed the prevalence and trend of COVID-19 vaccine fact-checking information on Facebook. Findings revealed health information Facebook pages responded to the COVID-19 infodemic by posting most frequently at 2 key vaccine time points in the United States: (1) when the vaccine first became available in December 2020 and (2) when the booster shot became available in August 2021 [41].

Notably, the percentage of fact-checking posts relative to all COVID-19 vaccine posts steadily decreased as the pandemic progressed. This may be because the frequency of COVID-19 vaccine posts increased at a higher rate than fact-checking vaccine posts. Another explanation is that public health organizations' efforts to promote accurate COVID-19 vaccine information reduced COVID-19 vaccine misinformation, necessitating that users fact-check misinformation. Likewise, posts tended to focus more on COVID-19 vaccine entities than comments did, reflecting the public concern over a more diverse set of topics relative to the vaccine itself. This may be because as the pandemic progressed, Facebook and sources of health information took actions to mitigate vaccine misinformation, necessitating less misinformation corrections and vaccine discourse over time. For example, Facebook began removing COVID-19 health misinformation and attaching various warnings to misleading posts, and public health initiatives promoted accurate COVID-19 vaccine information [42,43].

The Role of Hospitals in Communicating COVID-19 Vaccine Information

Our most prominent finding was that hospitals play a key role in disseminating facts and correcting misinformation. Although hospitals receive less engagement than other information sources, the comments expressed more positive emotions compared to other information sources. This suggests that

XSL•FO
RenderX

hospitals should invest more in generating engaging public health campaigns on social media.

Regarding overall emotions in the comments, fact-checking posts from health media and hospitals were associated with lower levels of anger, fear, and sadness in the comments, while posts from third-party fact checkers were associated with higher levels of anger in the comments. These negative emotions are crucial heuristic cues to the public's attitude and therefore should be acknowledged by information and health influencers in communicating facts and correcting misinformation. Empathetic communication enables fact-checking practitioners to better connect with the audience and counterbalance the negative emotions and hesitancy evoked by COVID-19 vaccine misinformation [25].

In addition, the majority of COVID-19 vaccine fact-checking Facebook posts were generated by news media and hospitals, while relatively few were from third-party fact checkers and US health media. Notably, third-party fact checkers posted more COVID-19 vaccine posts as the pandemic progressed. Although health media were the smallest source COVID-19 vaccine fact-checking posts, they have more followers than the other 3 sources, and although hospitals generated more fact-checking posts, they have the fewest followers. Although health media posts had similar levels of engagement as news media, they elicited few wow and angry reactions, likely reflecting a less negative attitude amongst followers of health media compared to news media. This may be because news media communicate with the general public, while health-concerned people follow health media and tend to have consistent health views. COVID-19 vaccine fact-checking posts created by news media were most popular, with the highest number of likes, comments, and shares on average, whereas users engaged with posts from hospitals the least.

Evolution of the Public's Attitude Over Time

Posts and comments tended to be relatively neutral in nature with low levels of attitudinal valence. However, as the pandemic progressed, the salience of COVID-19 vaccine entities in posts kept decreasing, and the publics' comments became more extreme, with higher levels of attitudinal magnitude. This suggests that fact-checking posts tend to report news and communicate facts objectively and have shifted the focus from the COVID-19 vaccine itself to other related subjects. However, the public's attitude became increasingly extreme over time. This supports extant findings that early interventions, such as inoculation against misinformation before attitude becomes increasingly extreme, may be more effective in the long term [44].

In addition, the salience of COVID-19 vaccine entities was significantly lower in comments than in posts. This suggested that the public is more concerned with issues other than the COVID-19 vaccine. The discrepancy in posts and comments further suggests the need for responsive and empathetic communication that might be more effective in improving the vaccine attitude and confidence.

Discrete Emotions and Linguistic Features

In line with our conclusion that public comments became more extreme as the pandemic progressed, fact-checking comments exhibited heightened joy, anger, fear, and sadness than posts. Although the presence of heightened positive emotions (eg, joy) in COVID-19 health messages has shown to predict compliance with COVID-19 public health guidelines [26], it is also true that individuals who have a positive attitude toward the vaccine may opt to generate and seek out COVID-19 misinformation corrections to reinforce their positive attitude [45]. Thus, these messages may not be reaching vaccine-hesitant individuals.

Likewise, users may seek and engage with sad content to manage negative emotions [45]. Notably, just as sadness can protect against initial belief in misinformation [46], it also seems to facilitate attitude change when encountered by misinformation corrections [47]. Thus, heightened sadness in fact-checking messages may hold promise for mitigating vaccine hesitancy.

Emotions and linguistic features in both COVID-19 vaccine fact-checking posts and comments evolved over time. The posts adopted a more analytical and confident tone over time, while we observed a significant drop in fear and tentativeness in the comments. Both trends suggest that with more information we know about the pandemic and the COVID-19 vaccine, the confidence related to the COVID-19 vaccine increases for both information sources and the general public.

Limitations

Although our findings shed light on COVID-19 vaccine fact-checking in a naturalistic setting, this study is not without limitations. By focusing on a sample of US posts, we neglected to explore how fact-checking manifests in other countries. Additionally, as different platforms have different behavioral norms [48], it is reasonable for user behaviors to vary by platform. Furthermore, our natural language processing sentiment analysis tools do not allow for the more refined coding established by human coders. However, we used a machine learning approach, which has shown to yield increased explanatory power, to reduce this limitation [49]. Relatedly, as with all observational studies, we cannot infer the direction of causality. Lastly, Facebook data usage restrictions prohibit researchers from collecting all comments; however, researchers are permitted to mine the top 25 comments for posts. The more user engagement (ie, likes, reactions, replies) a comment has, the higher up Facebook algorithmically ranks it in the comment thread [50]; thus, although we acknowledge not having all comments as a limitation, we believe the top 25 comments for each post are sufficiently informative and representative of public opinion.

Conclusion

This study has broad implications for public health practitioners and social media managers. First, although hospitals play a large role in fact-checking COVID-19 vaccine misinformation, they should work to design posts that will better engage the public. Hospitals are perceived as credible information sources with authority as health institutions, which makes them credible sources that are highly likely to elicit attitude and behavior change on health issues. Second, as fact-checking posts evoked

XSL•FO
RenderX

increasingly extreme public attitude over time, early interventions (ie, social media campaigns that inoculate against misinformation before it becomes mainstream) are critical. Additionally, fact-checking information sources should engage in empathetic communication to better address the concerns of the public and empathize with the public. For example, sadness can both protect against belief in misinformation and facilitate attitude change when confronted with misinformation corrections [46,47]. Expression of sadness in fact-checking messages holds promise for mitigating vaccine hesitancy. Distinct emotions are crucial heuristic cues for public attitude formation and therefore should be acknowledged by social media managers tasked with communicating facts and correcting misinformation, and, ultimately, countering COVID-19 vaccine hesitancy [25]. Finally, in contrast to health media and hospitals, who evoked less anger in public responses to fact-checking COVID-19 vaccine posts, third-party fact checkers tended to evoke heightened anger in responses; fact-checking agencies should be mindful of this when communicating with the public.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental material.
[DOCX File , 47 KB - jmir_v24i6e38423_app1.docx]

References

1. Mapping COVID-19. Center for Systems Science and Engineering (CSSE), Johns Hopkins University (JHU). URL: https://systems.jhu.edu/research/public-health/ncov/ [accessed 2022-06-14]
2. Centers for Disease Control and Prevention (CDC). Rates of COVID-19 Cases or Deaths by Age Group and Vaccination Status and Booster Dose. URL: https://data.cdc.gov/Public-Health-Surveillance/Rates-of-COVID-19-Cases-or-Deaths-by-Age-Group-and/3rge-nu2a/data [accessed 2022-06-14]
3. Alagoz O, Sethi AK, Patterson BW, Churpek M, Alhanaee G, Scaria E, et al. The impact of vaccination to control COVID-19 burden in the United States: a simulation modeling approach. PLoS One 2021;16(7):e0254456 [FREE Full text] [doi: 10.1371/journal.pone.0254456] [Medline: 34260633]
4. Tenforde MW, Self WH, Adams K, Gaglani M, Ginde AA, McNeal T, InfluenzaOther Viruses in the Acutely Ill (IVY) Network. Association between mRNA vaccination and COVID-19 hospitalization and disease severity. JAMA 2021 Nov 23;326(20):2043-2054. [doi: 10.1001/jama.2021.19499] [Medline: 34734975]
5. Iboi EA, Ngonghala CN, Gumel AB. Will an imperfect vaccine curtail the COVID-19 pandemic in the U.S.? Infect Dis Model 2020;5:510-524 [FREE Full text] [doi: 10.1016/j.idm.2020.07.006] [Medline: 32835142]
6. Wardle C, Singerman E. Too little, too late: social media companies' failure to tackle vaccine misinformation poses a real threat. BMJ 2021 Jan 21;372:n26. [doi: 10.1136/bmj.n26] [Medline: 33478950]
7. The New York Times. See Where 12 Million U.S. Employees Are Affected by Government Vaccine Mandates. URL: https://www.nytimes.com/interactive/2021/12/18/us/vaccine-mandate-states.html [accessed 2022-06-14]
8. Sabahelzain MM, Hartigan-Go K, Larson HJ. The politics of Covid-19 vaccine confidence. Curr Opin Immunol 2021 Aug;71:92-96 [FREE Full text] [doi: 10.1016/j.coi.2021.06.007] [Medline: 34237648]
9. Hitlin P, Olmstead K. The Science People See on Social Media. URL: https://www.pewresearch.org/science/2018/03/ [accessed 2022-06-14]
10. Eysenbach G. How to fight an infodemic: the four pillars of infodemic management. J Med Internet Res 2020 Jun 29;22(6):e21820 [FREE Full text] [doi: 10.2196/21820] [Medline: 32589589]
11. Kreps SE, Kriner DL. The COVID-19 infodemic and the efficacy of interventions intended to reduce misinformation. Public Opin Q 2022;86(1):162-175. [doi: 10.1093/poq/nfab075] [Medline: 34260633]
12. Carey JM, Guess AM, Loewen PJ, Merkley E, Nyhan B, Phillips JB, et al. The ephemeral effects of fact-checks on COVID-19 misperceptions in the United States, Great Britain and Canada. Nat Hum Behav 2022 Feb;6(2):236-243. [doi: 10.1038/s41562-021-01278-3] [Medline: 35115678]
13. Loomba S, de Figueiredo A, Piatek SJ, de Graaf K, Larson HJ. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. Nat Hum Behav 2021 Mar;5(3):337-348. [doi: 10.1038/s41562-021-01056-1] [Medline: 33547453]
14. Garett R, Young SD. Online misinformation and vaccine hesitancy. Transl Behav Med 2021 Dec 14;11(12):2194-2199. [doi: 10.1093/tbm/ibab128] [Medline: 34529080]
15. Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. Hum Vaccin Immunother 2020 Nov 01;16(11):2586-2593. [doi: 10.1080/21645515.2020.1780846] [Medline: 32693678]
16. Vraga EK, Bode L. Using expert sources to correct health misinformation in social media. Sci Commun 2017 Sep 14;39(5):621-645. [doi: 10.1177/1075547017731776]

XSL•FO
RenderX

17. Oeldorf-Hirsch A, Schmierbach M, Appelman A, Boyle MP. The ineffectiveness of fact-checking labels on news memes and articles. Mass Commun Soc 2020 Mar 23;23(5):682-704. [doi: 10.1080/15205436.2020.1733613]
18. Zhang J, Featherstone JD, Calabrese C, Wojcieszak M. Effects of fact-checking social media vaccine misinformation on attitudes toward vaccines. Prev Med 2021 Apr;145:106408. [doi: 10.1016/j.ypmed.2020.106408] [Medline: 33388335]
19. van der Meer TGLA, Jin Y. Seeking formula for misinformation treatment in public health crises: the effects of corrective information type and source. Health Commun 2020 May;35(5):560-575. [doi: 10.1080/10410236.2019.1573295] [Medline: 30761917]
20. Google Cloud. Cloud Natural Language: Natural Language AI. URL: https://cloud.google.com/natural-language [accessed 2022-06-14]
21. Sherman SM, Smith LE, Sim J, Amlôt R, Cutts M, Dasch H, et al. COVID-19 vaccination intention in the UK: results from the COVID-19 vaccination acceptability study (CoVAccS), a nationally representative cross-sectional survey. Hum Vaccin Immunother 2021 Jun 03;17(6):1612-1621. [doi: 10.1080/21645515.2020.1846397] [Medline: 33242386]
22. Gerend MA, Shepherd JE. Predicting human papillomavirus vaccine uptake in young adult women: comparing the health belief model and theory of planned behavior. Ann Behav Med 2012 Oct;44(2):171-180 [FREE Full text] [doi: 10.1007/s12160-012-9366-5] [Medline: 22547155]
23. Lee J, Bissell K. Assessing COVID-19 Vaccine Misinformation Interventions among Rural, Suburban and Urban Residents. Natural Hazards Center Quick Response Grant Report Series. Boulder, CO: Natural Hazards Center; 2022.
24. Betsch C, Ulshöfer C, Renkewitz F, Betsch T. The influence of narrative v. statistical information on perceiving vaccination risks. Med Decis Making 2011;31(5):742-753. [doi: 10.1177/0272989X11400419] [Medline: 21447730]
25. Chou WS, Budenz A. Considering emotion in COVID-19 vaccine communication: addressing vaccine hesitancy and fostering vaccine confidence. Health Commun 2020 Dec;35(14):1718-1722. [doi: 10.1080/10410236.2020.1838096] [Medline: 33124475]
26. Heffner J, Vives M, FeldmanHall O. Emotional responses to prosocial messages increase willingness to self-isolate during the COVID-19 pandemic. Pers Individ Dif 2021 Feb 15;170:110420. [doi: 10.1016/j.paid.2020.110420] [Medline: 33082614]
27. MacFarlane D, Hurlstone MJ, Ecker UKH. Protecting consumers from fraudulent health claims: a taxonomy of psychological drivers, interventions, barriers, and treatments. Soc Sci Med 2020 Aug;259:112790. [doi: 10.1016/j.socscimed.2020.112790] [Medline: 32067757]
28. Lerner JS, Keltner D. Fear, anger, and risk. J Pers Soc Psychol 2001;81(1):146-159. [doi: 10.1037/0022-3514.81.1.146]
29. Stevens HR, Oh YJ, Taylor LD. Desensitization to fear-inducing COVID-19 health news on Twitter: observational study. JMIR Infodemiol 2021 Jul 16;1(1):e26876 [FREE Full text] [doi: 10.2196/26876] [Medline: 34447923]
30. Sbaffi L, Rowley J. Trust and credibility in web-based health information: a review and agenda for future research. J Med Internet Res 2017 Jun 19;19(6):e218 [FREE Full text] [doi: 10.2196/jmir.7579] [Medline: 28630033]
31. Mitra T, Counts S, Pennebaker J. Understanding anti-vaccination attitudes in social media. Proceedings of the International AAAI Conference on Web and Social Media. 2016. URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14729 [accessed 2022-06-14]
32. Witte K, Allen M. A meta-analysis of fear appeals: implications for effective public health campaigns. Health Educ Behav 2000 Oct;27(5):591-615. [doi: 10.1177/109019810002700506] [Medline: 11009129]
33. Newhagen JE, Bucy EP. Overcoming resistance to COVID-19 vaccine adoption: how affective dispositions shape views of science and medicine. HKS Misinfo Review 2020 Oct 23:1-16. [doi: 10.37016/mr-2020-44]
34. Subbarao K. COVID-19 vaccines: time to talk about the uncertainties. Nature 2020 Oct 20;586(7830):475. [doi: 10.1038/d41586-020-02944-8] [Medline: 33082537]
35. Yang A, Shin J, Zhou A, Huang-Isherwood KM, Lee E, Dong C, et al. The battleground of COVID-19 vaccine misinformation on Facebook: fact checkers vs. misinformation spreaders. HKS Misinfo Review 2021 Aug 30:1-15. [doi: 10.37016/mr-2020-78]
36. CrowdTangle. URL: http://www.crowdtangle.com [accessed 2022-05-11]
37. Jünger J, Keyling T. strohne/Facepager. URL: https://github.com/strohne/Facepager [accessed 2022-06-14]
38. facebook. Reactions. URL: https://www.facebook.com/brand/resources/facebookapp/reactions [accessed 2022-06-14]
39. IBM. Watson Natural Language Understanding. URL: https://www.ibm.com/cloud/watson-tone-analyzer [accessed 2022-05-11]
40. UCDavis Office of Research. Does My Project Need Review by the IRB. URL: https://research.ucdavis.edu/policiescompliance/irb-admin/researchers/irb-submissions/new-projects/do-i-need-irb-review/ [accessed 2022-06-17]
41. Mathieu E, Ritchie H, Ortiz-Ospina E, Roser M, Hasell J, Appel C, et al. A global database of COVID-19 vaccinations. Nat Hum Behav 2021 Jul;5(7):947-953 [FREE Full text] [doi: 10.1038/s41562-021-01122-8] [Medline: 33972767]
42. Brennen J, Simon F, Howard P, Kleis NR. Types, Sources, and Claims of COVID-19 Misinformation. URL: https://tinyurl.com/58dak255 [accessed 2022-05-11]
43. The White House. Remarks by President Biden on the COVID-19 Response and the Vaccination Program. URL: https://tinyurl.com/2sdjmntn [accessed 2022-05-11]
44. Maertens R, Roozenbeek J, Basol M, van der Linden S. Long-term effectiveness of inoculation against misinformation: three longitudinal experiments. J Exp Psychol Appl 2021 Mar;27(1):1-16. [doi: 10.1037/xap0000315] [Medline: 33017160]

XSL•FO
RenderX

45. Allen Catellier JR, Yang ZJ. Trust and affect: how do they impact risk information seeking in a health context? J Risk Res 2012 Sep;15(8):897-911. [doi: 10.1080/13669877.2012.686048]
46. Forgas JP. Don't worry, be sad! On the cognitive, motivational, and interpersonal benefits of negative mood. Curr Dir Psychol Sci 2013 Jun 04;22(3):225-232. [doi: 10.1177/0963721412474458]
47. Trevors G, Bohn-Gettler C, Kendeou P. The effects of experimentally induced emotions on revising common vaccine misconceptions. Q J Exp Psychol (Hove) 2021 Nov;74(11):1966-1980. [doi: 10.1177/17470218211017840] [Medline: 33926324]
48. Choudhury D, Sushovan M. Mental health discourse on reddit: self-disclosure, social support, and anonymity. 2014 Presented at: Eighth International AAAI Conference on Weblogs and Social Media; 2014; Ann Arbor, MI.
49. Frankel R, Jennings J, Lee J. Disclosure sentiment: machine learning vs. dictionary methods. Manag Sci 2021 Nov 11:1-19. [doi: 10.1287/mnsc.2021.4156]
50. Facebbook Help Center. How Do I Turn Comment Ranking off for My Facebook Page?. URL: https://www.facebook.com/help/www/1494019237530934 [accessed 2022-06-14]

Abbreviations

AI: artificial intelligence
RQ: research question

Edited by M Gisondi, M Gottlieb; submitted 31.03.22; peer-reviewed by J Barnett, SF Tsao, M Lotto, K Ahmed; comments to author 23.04.22; revised version received 12.05.22; accepted 30.05.22; published 21.06.22.

Please cite as:
Xue H, Gong X, Stevens H
COVID-19 Vaccine Fact-Checking Posts on Facebook: Observational Study
J Med Internet Res 2022;24(6):e38423
URL: https://www.jmir.org/2022/6/e38423
doi:10.2196/38423
PMID:35671409

©Haoning Xue, Xuanjun Gong, Hannah Stevens. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.06.2022. 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