Journal of Medical Internet Research

Journal Impact Factor (JIF) (2022): 7.4
Volume 24 (2022), Issue 4 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Original Papers

Viewpoints

Tutorials

Review

The Efficacy of Health Information Technology in Supporting Health Equity for Black and Hispanic Patients With Chronic Diseases: Systematic Review

Charles Senteio[1*], PhD, MSW, MBA; Paul Joseph Murdock[2*], BA, MS

[1]Department of Library and Information Science, School of Communication and Information, Rutgers University, New Brunswick, NJ, United States
[2]Division of Health Sciences Informatics, School of Medicine, Johns Hopkins University, Baltimore, MD, United States
[*]all authors contributed equally

Corresponding Author:
Charles Senteio, PhD, MSW, MBA
Department of Library and Information Science
School of Communication and Information
Rutgers University
4 Huntington St
New Brunswick, NJ, 08901-1071
United States
Phone: 1 8489327586
Email: charles.senteio@rutgers.edu

Abstract

Background: Racial inequity persists for chronic disease outcomes amid the proliferation of health information technology (HIT) designed to support patients in following recommended chronic disease self-management behaviors (ie, medication behavior, physical activity, and dietary behavior and attending follow-up appointments). Numerous interventions that use consumer-oriented HIT to support self-management have been evaluated, and some of the related literature has focused on racial minorities who experience disparate chronic disease outcomes. However, little is known about the efficacy of these interventions.

Objective: This study aims to conduct a systematic review of the literature that describes the efficacy of consumer-oriented HIT interventions designed to support self-management involving African American and Hispanic patients with chronic diseases.

Methods: We followed an a priori protocol using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)-Equity 2012 Extension guidelines for systematic reviews that focus on health equity. Themes of interest included the inclusion and exclusion criteria. We identified 7 electronic databases, created search strings, and conducted the searches. We initially screened results based on titles and abstracts and then performed full-text screening. We then resolved conflicts and extracted relevant data from the included articles.

Results: In total, there were 27 included articles. The mean sample size was 640 (SD 209.5), and 52% (14/27) of the articles focused on African American participants, 15% (4/27) of the articles focused on Hispanic participants, and 33% (9/27) included both. Most articles addressed 3 of the 4 self-management behaviors: medication (17/27, 63%), physical activity (17/27, 63%), and diet (16/27, 59%). Only 15% (4/27) of the studies focused on follow-up appointment attendance. All the articles investigated HIT for use at home, whereas 7% (2/27) included use in the hospital.

Conclusions: This study addresses a key gap in research that has not sufficiently examined what technology designs and capabilities may be effective for underserved populations in promoting health behavior in concordance with recommendations.

(J Med Internet Res 2022;24(4):e22124) doi:10.2196/22124

KEYWORDS

chronic disease; minority health; technology assessment; biomedical; self-management; systematic review; mobile phone

Introduction

Background

Nearly half of all adults in the United States are living with 1 or more of the *Big Five* chronic conditions—diabetes mellitus (*diabetes*), cardiovascular disease, chronic respiratory disease, cancer, and stroke [1]. Racial inequity persists for outcomes under these conditions [2]. For example, African American individuals continue to experience greater disease prevalence than non-Hispanic White individuals for hypertension (25%) and diabetes (49%); likewise, Hispanic individuals' diabetes rates are 20% higher than those of White individuals [3]. Furthermore, nearly 5 decades of literature details racial and ethnic inequity in diabetes prevalence and risk factors for diabetes-related complications and following recommended self-management behavior [4].

Chronic disease self-management is challenging because the treatment regimens often demand much from the patient and their families; recommended self-management frequently includes regular meal planning, consistent physical activity, monitoring and tracking (eg, fluid intake and blood glucose), and daily medication behavior [5]. Following the recommended self-management behavior is vital because these behaviors are associated with health outcomes. For example, following the recommended medication behavior, physical activity, dietary behavior, and blood sugar testing are all associated with glycemic control [6]. Comorbidity can exacerbate the burden associated with following self-management recommendations. For example, a cancer survivor with diabetes who must take medication as part of their cancer treatment (eg, prednisone) may experience difficulty in maintaining the recommended glucose levels, which can in turn impact medication behavior [7]. Chemotherapy can also cause adverse side effects, including pain and cognitive impairment. Both can present barriers to following recommended self-management for years following cancer treatment [8], and cancer survivors from racial minority groups experience poorer outcomes for other chronic conditions diagnosed after a cancer diagnosis [9].

Patients with chronic diseases may use information technology (eg, mobile apps) as sources of health information to help answer questions regarding symptoms and treatment options [10-12]. However, racial inequity also characterizes access to information and communication technologies (ICTs). Most White individuals own a laptop or desktop computer (83%), whereas only about two-thirds of African American individuals (66%) reported owning either. There is also racial and ethnic inequity in access to broadband at home, with 78% of the White population reporting access compared with 65% of African American individuals and 58% of Hispanic individuals [13]. African American individuals and Hispanic individuals own smartphones and tablets at similar rates as White individuals; however, smartphones represent the only web-based access for 12% of the African American population and 22% of the Hispanic population, whereas only 4% of White individuals only access the internet via smartphones [13]. Furthermore, African American individuals experience disruptions in access, as they are twice as likely as White individuals to cancel or suspend mobile phone services because of cost [13]. These interruptions are particularly vital because African American individuals are more likely to use smartphones for web-based access than White individuals [14,15]. To use health information technology (HIT) to help support following the recommended self-management health behavior, individuals must both have access to ICTs and possess the requisite skills to use them [16,17]. African American individuals experience barriers to HIT use because of inequitable access and disparities in skills required to use technology designed to support chronic disease self-management [16]. Consequently, the extent to which this technology is effective in supporting Hispanic and African American patients for chronic disease self-management is unclear. Understanding efficacy is imperative given persistent disparities in health outcomes and in HIT access and use.

Sociocultural factors also influence individuals from ethnic minority groups' use of consumer-oriented HIT. Trust, perceived credibility, attitudes, and perceptions predict health technology acceptance and use [17]. For example, over a decade of research describes how African American individuals have different attitudes than White individuals regarding technology innovations in health care, and these factors predict HIT acceptance [18]. Trust is an important consideration in the design of health informatics interventions to promote health and wellness [19]. Sociocultural barriers (eg, unwanted attention) are among the barriers Hispanic populations report for consumer-oriented HIT [20].

Sociocultural factors present barriers that contribute to intervention-generated inequality [21,22]. Intervention-generated inequality occurs when technology-enabled health informatics approaches disproportionally benefit most populations [17]. Therefore, these interventions are less effective for minority populations and can essentially exacerbate population disparities that contribute to health inequity [23]. HIT-enabled health promotion can be enhanced by developing HIT that considers sociocultural factors that influence use (eg, levels of health literacy and digital literacy, lack of access to, or knowledge of digital tools) [24]. Systematic reviews of consumer-oriented HIT to support health and wellness find that articles do not adequately consider sociotechnical factors [25].

Objectives

HIT research describes the potential benefit from the use of technologies designed to track and report health behaviors, along with the acknowledgment of sparse insights to guide researchers concerning specific barriers to use for ethnically diverse populations [20]. However, no systematic review has been published describing the efficacy of consumer-oriented HIT designed to support following recommended self-management behavior for African American or Hispanic patients with chronic diseases. Therefore, we conducted this study of efficacy of consumer-oriented HIT in these patients. For this study, we classify *consumer-oriented HIT* as a technology designed to support recommended chronic disease self-management. It includes a myriad of mobile, tablet, and computer apps designed to support following recommended chronic disease self-management behaviors, such as electronic journals to track physical activity and prompts and reminders

to support medication behavior. HIT also includes technology that enables access to health information, such as podcasts and disease-specific discussion boards. Given that ethnic minority populations experience both persistent inequity in chronic disease outcomes and barriers to access of consumer-oriented HIT designed to support following recommended self-management behavior, this study was guided by the following research question: what is the impact on clinical outcomes of consumer-oriented HIT interventions on self-management behavior or health outcomes for Black or Hispanic patients with chronic diseases?

Methods

Overview

After confirming health equity as the focus of this study, we followed an a priori protocol with equity as the focus, the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)-Equity 2012 Extension was selected as a guideline for conducting systematic reviews that focus on health equity [26,27].

Inclusion Criteria

We developed a rationale for eligible study designs and inclusion of outcomes, per the PRISMA-Equity 2012 Extension for systematic reviews [27]. First, we identified foundational articles based on the refined research question [28-31]. We then reviewed papers from several journals that published the foundational articles and published the foundational articles and journals that published papers that the foundational articles cited. Given the interdisciplinary nature of health equity research, we selected established journals, with an emphasis on health equity (eg, *Social Science & Medicine*, the *Journal of Racial and Ethnic Health Disparities*, and the *Journal of Health Care for the Poor and Underserved*) and from medical informatics (eg, the *Journal of American Medical Informatics Association* and the *Journal of Medical Internet Research*). We chose a systematic review based on the types of articles appearing in these journals. Given that outcomes are germane for describing inequity, we selected journals that reported outcomes.

Information Sources

Next, we crafted themes of interest, again per the PRISMA-Equity 2012 Extension for systematic reviews [26], which formed the foundation of our inclusion and exclusion criteria: health technology designed for patients, Unified Theory of Acceptance and Use of Technology, theme (eg, acceptance, usability, readiness, satisfaction, and preference), self-management (eg, self-management behavior, health behavior, adherence, and compliance), health conditions (eg, chronic disease and physical health), and demographics.

To evaluate and select databases, we again reviewed the 4 foundational articles. We also consulted with a health sciences librarian to evaluate and finalize the databases. We selected seven electronic databases: PubMed, Cumulative Index of Nursing and Allied Health Literature, Web of Science, Cochrane, Compendex, Institute of Electrical and Electronics Engineers, and Computers and Applied Sciences Complete.

Search Strategy

We created search strings based on our themes of interest (eg, acceptance, usability, readiness, satisfaction, and preference), according to the specific database format, to locate articles that met our inclusion criteria. We consulted with health science librarians to ensure adherence to the database string format. Information regarding the search strategy (eg, search strings) is given in Multimedia Appendix 1 [30,32-57]. When the database permitted, all results were limited to peer-reviewed journal articles published after 1990 as the World Wide Web was introduced during this period. All database searches were conducted on November 26, 2018. In addition, PJM hand-searched references of the included articles to ensure all pertinent articles were included.

Study Selection

Articles were included if they met specific inclusion criteria and excluded if they fulfilled the exclusion criteria (Textbox 1).

Rayyan (Rayyan Inc), an internet-based software package, was used to facilitate article screening [58]. CRS and PJM blindly completed the title and abstract and full-text screening. They resolved conflicts together after the blind screening feature in Rayyan was turned off.

Textbox 1. Article inclusion and exclusion criteria.

Inclusion criteria

- Articles included patients with chronic diseases or caregivers who specified they were of Black or African American, or Hispanic origin.
- The patient or caregiver must be the end user or direct benefactor of technology.
- Technology gives personalized information to patients and or caregivers.
- Technology was designed to support self-management recommended for chronic conditions (ie, medication behavior, physical activity, dietary behavior, and attending follow-up appointments).
- The article is in English in a peer-reviewed journal.
- The article has been published since 1990.

Exclusion criteria

- Intervention targets providers.
- No electronic technologies (ie, technology using electricity) examined in the article.
- Technology is not designed to support self-management recommended for chronic conditions (ie, medication behavior, physical activity, dietary behavior, and attending follow-up appointments). Technology designed to prevent falls was not included.
- A systematic review of technology.

Data Collection Process and Data Items

Once conflicts were resolved, we analyzed the included articles and extracted relevant information (Table 1). Given the focus of our review is technology designed for chronic disease self-management for African American and Hispanic patients, we detailed information concerning race or ethnicity and cultural tailoring, type of technology used, behavior targeted, and specific chronic disease and clinical outcomes measured. We used content analysis to classify themes and totaled the frequency of self-management activities reported.

We analyzed the risk of bias in each included article using the Cochrane Collaboration Risk of Bias Tool [59]. The tool was developed in 2005 based on the following seven principles for assessing risk of bias in randomized trials: (1) avoiding use of quality scales (eg, because scales, and resulting scores, are inappropriate appraisals of clinical trials, their use increases risk of bias), (2) focusing on internal validity (eg, a small trial with high internal validity may have high risk of bias, whereas a large trial, while having high precision may have high risk of bias if internal validity is low), (3) assessing the risk of bias in trial results (eg, the quality of the reporting—which may be assessed by evaluating level of detail—helps determine the risk of bias; methodology used in conducting the trial—such as not calculating the sample size with power analysis, not including ethical review board approval, or limiting participants' knowledge of intervention received can all increase the risk of bias), (4) using judgment when assessing risk of bias (eg, omitting bias assessments from aspects of the trail methodology or interpretation of results may increase risk of bias), (5) choosing domains to be assessed (eg, if detail is not described for how incomplete data were accounted for, or aspects of blinding for participants and practitioners, can increase the risk of bias), (6) focusing on the risk of bias in the data as represented in the article (eg, the exclusion of certain participants in trial results who are then reinstated for other results increases the risk of bias), and (7) reporting outcome-specific evaluations of the risk of bias (eg, describing randomized allocation to control or experimental group during participation may influence the risk of bias in other aspects of the trial, such as physicians' knowledge of the specific intervention and its usual effects). The tool contains six domains for assessing potential bias, with sources of bias in each domain: (1) selection bias (inadequate generation of a randomized sequence and inadequate concealment of allocations before an assignment increase the risk of bias), (2) performance bias (inadequate blinding of participants and study personnel increases the risk of bias), (3) detection bias (inadequate blinding of outcome assessment increases the risk of bias), (4) attrition bias (incomplete outcome data for outcomes reported increases the risk of bias), (5) reporting bias (selective reporting increases the risk of bias), and (6) other bias (ie, any bias not included in the other 5 named domains).

Table 1. General characteristics (N=27).

Characteristics[a]	Values, n (%)
Self-management area	
Medication behavior	17 (62)
Follow-up appointment attendance	4 (14)
Physical activity	17 (62)
Dietary behavior	16 (59)
Care setting	
Home (capability to access or use from home)	27 (100)
Hospital[b]	2 (7)
Technology	
Computer, laptop, or tablet[c]	3 (11)
Telephone (landline)	0 (0)
Mobile phone	17 (62)
Mobile app	1 (3)
Text	15 (55)
Web-based	8 (29)
Bluetooth device	2 (7)
Specialized telemedicine device	2 (7)
Nintendo Wii	1 (3)
Voice-enabled device	1 (3)
Social media	1 (3)
Function	
Collecting personal health data[d]	13 (48)
Goal setting and tracking	17 (62)
Integrated survey and assessment	19 (70)

[a]Articles may be included within multiple categories.

[b]We did not include articles in which users could use videos to chat or communicate with providers.

[c]Telemedicine units or devices were included.

[d]Tracking of patient's personal health data (data logs) and tracking of patient data by providers were included.

Results

Study Characteristics

A total of 25 eligible articles involving African American participants and 13 articles with Hispanic participants were identified. Of these, only 27 met our final criteria, as not all articles discussed technology use and design for patients (see PRISMA flowchart in Figure 1). All 27 articles were published between 1996 and 2018. The mean participant sample size was 640 (SD 209.5; 26/27, 96% of articles). Of the 27 included articles, 14 (52%) focused exclusively on African American patients, 4 (15%) focused on Hispanic patients, and 9 (33%) focused on both African American and Hispanic patients.

Each of the 27 included articles was examined for the risk of potential bias according to each of the 6 domains (Table 2).

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

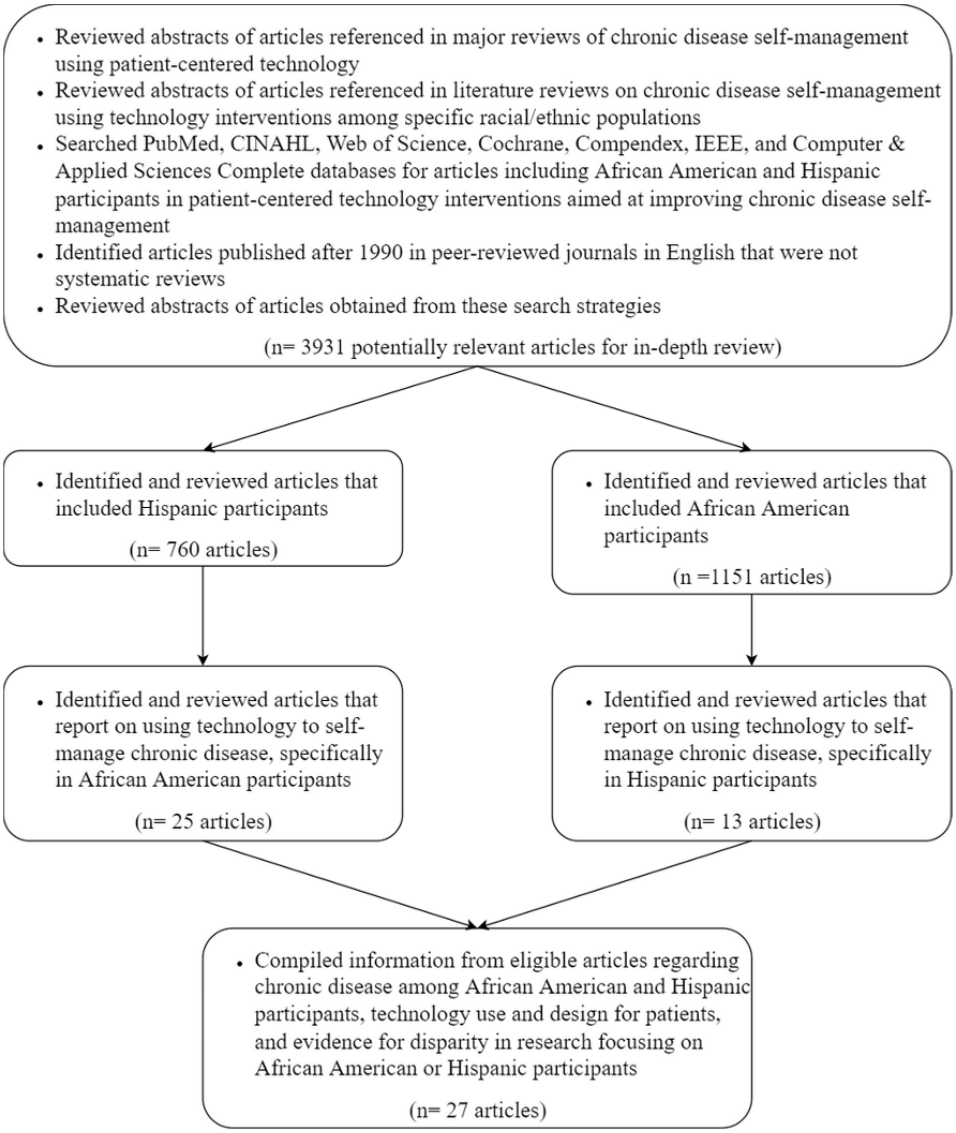


Table 2. Risk of bias in individual articles (N=27).

Study	Participants, n	Patients or caregivers involved in the design of technology	Incomplete outcome data[a]	Blinding of participants or personnel[b]	Other bias[c]
Almeida et al [32]	452	High	Low	Low	Not reported
Collins and Champion [33]	15	Not reported	Low	Low	Not reported
Davidson et al [34]	50	Low	Low	Low	Low
Davis et al [35]	51	Low	Low	High	Not reported
Finkelstein et al [36]	30	Not reported	Low	Low	Not reported
Finkelstein and Wood [37]	N/A[d]	High	High	Not reported	Low
Fortmann et al [38]	414	Low	Low	Low	Not reported
Friedman et al [39]	267	Not reported	Low	Low	Not reported
Gerber et al [40]	95	Not reported	Not reported	High	High
Green et al [41]	9298	Low	Low	Low	Low
Grimes et al [42]	12	Low	Not reported	High	Not reported
Heitkemper et al [30]	220	Low	High	Not reported	Not reported
Joseph et al [43]	29	Low	Low	Low	Not reported
Kline et al [44]	123	Not reported	Low	High	Not reported
MacDonell et al [45]	48	Low	High	Low	Low
Lin et al [46]	124	High	Low	Low	High
Mayberry et al [47]	19	Low	Low	High	Not reported
McGillicuddy et al [48]	12	Low	Low	Low	Not reported
Newton et al [49]	97	Not reported	High	Low	High
Nundy et al [50]	15	Not reported	Low	Not reported	Not reported
Reese et al [51]	14	Low	High	High	Not reported
Reininger et al [52]	71	Not reported	Low	High	Not reported
Rosal et al [53]	89	Low	Low	Low	Low
Shea [54]	1665	High	High	Low	Low
Skolarus et al [55]	94	Low	High	Low	Low
Trief et al [56]	1665	Low	Low	Low	Not reported
Weinstock et al [57]	1665	Low	Low	Low	Not reported

[a]Outcome data.

[b]Randomization or blinding of patients.

[c]Any other bias identified by the reviewers.

[d]N/A: not applicable.

Additional Analyses: Qualitative Synthesis

Articles that reported technology interventions and included self-management aimed at improving chronic disease outcomes using either clinical or behavioral outcomes were eligible for systematic review inclusion (Table 3). We chose content analysis for categorizing data into themes and counting their frequency based on our decision to count frequency of self-management behaviors [60]. We reported the following four specific self-management activities: medication behavior, physical activity, dietary behavior, and follow-up appointment attendance. The frequency of each self-management behavior was totaled by analyzing the included articles.

Table 3. Self-management behaviors in the included articles (N=27).

Study	Medication behavior	Follow-up appointment attendance	Physical activity	Dietary behavior
Almeida et al [32]	No	No	Yes	No
Collins and Champion [33]	No	No	Yes	Yes
Davidson et al [34]	Yes	No	No	No
Davis et al [35]	Yes	No	Yes	Yes
Finkelstein et al [36]	No	No	Yes	No
Finkelstein and Wood [37]	No	No	Yes	No
Fortmann et al [38]	Yes	No	Yes	Yes
Friedman et al [39]	Yes	No	No	No
Gerber et al [40]	No	No	Yes	Yes
Green et al [41]	Yes	Yes	No	No
Grimes et al [42]	No	No	No	Yes
Heitkemper et al [30]	Yes	No	Yes	Yes
Joseph et al [43]	No	No	Yes	No
Kline et al [44]	Yes	No	Yes	Yes
MacDonell et al [45]	Yes	No	No	No
Lin et al [46]	No	No	Yes	Yes
Mayberry et al [47]	Yes	No	Yes	Yes
McGillicuddy et al [48]	Yes	No	No	No
Newton et al [49]	No	No	Yes	Yes
Nundy et al [50]	Yes	Yes	No	Yes
Reese et al [51]	No	No	Yes	No
Reininger et al [52]	Yes	No	Yes	Yes
Rosal et al [53]	Yes	No	Yes	Yes
Shea [54]	Yes	No	No	No
Skolarus et al [55]	Yes	No	Yes	Yes
Trief et al [56]	Yes	Yes	No	Yes
Weinstock et al [57]	Yes	Yes	No	Yes

Study Selection

Other data recorded from the articles included the technology functions (Table 4), the type of technology used, the effectiveness of the technology, the number of participants enrolled, and the first author's last name (Multimedia Appendix 1). These were grouped together to reveal findings such as which technology was effective based on population and chronic disease type (Multimedia Appendix 1) [29,30,32-57].

Table 4. Technology functions in the included articles (N=27).

Study	Tracking by a patient or caregiver using technology	Tracking or viewing patient data by a patient or caregiver	Tracking of patient data by providers	Goal setting or tracking	Integrated surveys or assessments
Almeida et al [32]	No	No	Yes	Yes	No
Collins and Champion [33]	No	No	No	No	Yes
Davidson et al [34]	Yes	Yes	Yes	Yes	Yes
Davis et al [35]	No	No	No	Yes	Yes
Finkelstein et al [36]	No	Yes	Yes	Yes	No
Finkelstein and Wood [37]	No	Yes	Yes	Yes	No
Fortmann et al [38]	No	No	No	Yes	No
Friedman et al [39]	No	No	No	Yes	Yes
Gerber et al [40]	No	No	No	No	No
Green et al [41]	Yes	Yes	Yes	No	Yes
Grimes et al [42]	No	No	No	Yes	No
Heitkemper et al [30]	No	No	No	Yes	Yes
Joseph et al [43]	No	Yes	Yes	Yes	Yes
Kline et al [44]	No	No	No	No	Yes
MacDonell et al [45]	No	No	No	No	Yes
Lin et al [46]	No	Yes	Yes	Yes	No
Mayberry et al [47]	No	Yes[a]	No	Yes	Yes
McGillicuddy et al [48]	No	Yes	Yes	No	No
Newton et al [49]	No	Yes	Yes	No	Yes
Nundy et al [50]	Yes	No	Yes	No	Yes
Reese et al [51]	No	No	No	Yes	Yes
Reininger et al [52]	No	No	No	No	Yes
Rosal et al [53]	No	Yes	Yes	Yes	Yes
Shea [54]	No	Yes	Yes	No	Yes
Skolarus et al [55]	No	Yes	No	Yes	Yes
Trief et al [56]	No	Yes	Yes	Yes	Yes
Weinstock et al [57]	No	Yes	Yes	Yes	Yes
Total, n (%)	4 (14)	14 (51)	15 (55)	17 (62)	19 (70)

[a]Coaching of family members via phone was also conducted.

Health Outcomes Described in the Included Articles

Diabetes, hypertension, and heart failure were the three chronic conditions included in the resultant studies (N=27). Diabetes was the most common chronic disease among these studies. Of the total number of studies, 8 specifically tracked hemoglobin A_{1c} (HbA_{1c}) and blood pressure (BP) levels [30,35,38,44,54,56,57]. Only a study by Weinstock et al [57] reported statistical significance for both clinical outcomes. A study by Davis et al [35] reported increases in medication adherence and self-efficacy for diabetes medication behavior in African American patients. However, none of the results were statistically significant, and they did not report any clinical significance. The study by Weinstock et al [57] targeted A_{1c} reduction in Hispanic and African American patients. The intervention included a home telemedicine unit with a web-enabled camera for a videoconference consultation, which provided educational information. Results showed HbA_{1c} improvement for Hispanic patients; however, improvement was not statistically significant and clinical significance was not specified. A study by Shea [54] used a home telemedicine–specialized device for videoconferencing with a nurse to support HbA_{1c} and BP monitoring. The intervention reported clinically significant A_{1c} improvement (8.35%-7.42%), but the results were not statistically significant. A study by Heitkemper et al [30] used a website and SMS text messages

for diabetes management targeting African American and Hispanic patients. Use was low among participants because they rarely used the internet to search for health information; consequently, outcomes were not reported. In a study by Kline et al [44], the intervention was a culturally tailored guide for diabetes management targeting the Hispanic population. It included a telenovela with learning modules and games. However, specific clinical outcomes (eg, A_{1c}) were not reported, as the focus was on the development and feasibility of the intervention.

Hypertension was the next most common condition specified (ie, they focused on hypertension vs BP reporting). Three studies specified the goals of reducing hypertension [39,48,55]. However, none reported statistical or clinical significance. A study by Skolarus et al [55] used an SMS text messaging intervention with a faith-based collaborator and reported both systolic and diastolic BP for African American patients. Half of the participants reached BP targets. A study by McGillicuddy et al [48] used a mobile health intervention that targeted Hispanic patients to promote BP improvement through medication self-management support. The study reported statistically significant increases in medication adherence. A study by Davidson et al [34] reported statistically significant results in systolic and diastolic BP reduction in Hispanic and African American participants. The system used electronic medication dispensers and SMS text messages. It included Bluetooth-enabled BP monitors.

Heart failure was the third chronic disease that was the focus of one of the resultant studies. A study by Finkelstein and Wood [37] assessed the feasibility of an intervention that used a laptop and Nintendo Wii to support medication behavior in African American patients with heart failure. Although clinical outcomes were not reported, participants reported a high level of acceptance of the technology.

Discussion

Principal Findings

Given the development of HIT apps and considerable research in this area, a relatively small number of resultant articles (N=27) investigated associations between the use of HIT and chronic disease outcomes among African American and Hispanic patients. This is a vital gap because of persistent inequity in chronic disease outcomes for racial minority populations and because intention to use HIT designed for chronic disease self-management is most predicted by performance expectancy, followed by social influence [61]. Researchers of HIT acceptance and use for chronic disease self-management should incorporate health outcomes in investigations, particularly outcomes commonly used to report racial inequity. Of the 27 articles, a majority addressed 3 of the 4 self-management behaviors investigated: medication behavior (17/27, 63%), physical activity (17/27, 63%), and dietary behavior (16/27, 59%). However, only a few (4/27, 15%) focused on follow-up appointment attendance. This is an area that warrants investigation and development of capabilities because HIT may be well-positioned to mitigate known causes of missed appointments, which is a persistent issue among racial minority populations who experience persistent inequity, such as Latinx immigrants and low-income African American patients in urban settings: forgetfulness, transportation barriers, family and employer obligations, and anticipated long clinic wait times [62-64]. Investigating and alleviating barriers to appointment attendance is important because ethnic minority patients are more likely to have low income and live in urban areas, two factors that are associated with the frequency of missing primary care appointments [65-67].

All the articles investigated HIT designed with capabilities to access or use at the patient's home, whereas only 2 articles also included use in the hospital. This is concordant with the movement of developing HIT for use in patients' homes versus hospitals [68]. The risk of misuse of HIT is segmented according to environmental, human, and technological factors [69]. The number of different users is associated with the risk of misuse because users may have various levels of education, instruction, or training. Thus, the risk of misuse is higher in home care settings when compared with hospital care settings [68]. Developers should consider the known risks of misuse and the number of users in home care settings for African American and Hispanic patients, given that individuals from racial minority populations have different attitudes than White patients regarding technology innovations in health care, and these factors predict HIT acceptance [18].

Various technologies are included in the resultant articles, except for the landline telephone, in which none of the articles were investigated. This follows the broad trend that more than half of US households are reliant on mobile phones and do not have landlines. In addition, Hispanic and Black adults are more likely than White adults to live in households with only mobile phones [70].

The collecting and tracking of personal health data, which over 10% of users are doing on behalf of someone else (eg, caregivers), and goal setting and evaluation are pertinent capabilities that are closely related to self-management behavior [71,72]. In addition, given the association of social influence with intention to use HIT, caregivers and other members of the patient's network should be incorporated into the design of HIT. This may be especially pertinent for Black women who report feeling responsible for providing emotional and tangible support to homebound parents who may live in the home [73]. In fact, incorporating shared tracking use should inform HIT design, and models have been created that reflect the interplay of social context and health tracking [71].

Insights derived from this study of the 27 resultant articles reveal the potential for future development and evaluation of HIT tools in two distinct areas—known barriers faced by members of ethnic minority groups in using HIT and the unique barriers they may face in following self-management recommendations. For example, in a limited sample size, a mobile phone–based intervention that combined SMS text messaging with nursing care showed improvement in following recommended self-management behavior (ie, medication behavior, glucose monitoring, foot care, physical activity, and dietary behavior) for Black adults with diabetes [74]. In addition, a diabetes self-management education intervention for medically

underserved populations showed specific impact on outcomes that are characterized by racial disparities and HbA_{1c} improvement at 6 and 12 months [30].

Despite these important findings, more specific research is needed to elucidate the sociocultural factors that in particular are known to impact HIT acceptance and use [75]. For example, the level of trust is associated with HIT acceptance and use in the context of diabetes self-management [76-78]. Moreover, factors for older adults from racial minority groups should be specifically investigated because they are less likely than individuals who do not belong to racial minority groups to use health management sites and search the web for health information to support chronic disease self-management [79]. Finally, investigations of HIT acceptance, use, and impact on self-management and outcomes should be conducted with larger samples. Despite considerable literature on drivers of inequity and the emergent literature describing the potential for HIT to support chronic disease self-management, the literature suggests that persistent disparities in chronic disease outcomes are in part because of the lack of large-scale, HIT-enabled interventions that support following self-management recommendations and report impact on outcomes [75,80-82]. In addition, given the limited reporting of clinical outcomes that inform equity measures (eg, HbA_{1c}), more research is needed to understand if or how access and then use may impact following recommended self-management behavior and subsequent outcomes. Doing so may reveal critical insights to associate HIT access with outcomes, particularly imperative given persistent barriers to technology acceptability and use [75].

This study has a key limitation. We only examined articles that specified Black and Hispanic users. Specific cultural factors may emerge from a broader examination, given that various cultural factors influence both technology acceptance and use (eg, practices, customs, language, and communication) [82,83]. Understanding cultural factors is essential because they can influence the way an individual interprets health information, how they define symptoms, and if and who they decide should provide them care [75]. Therefore, individuals' sociocultural factors must be considered in the design and use of *culturally informed HIT* [84]. This insight is vital because cultural competence is specified as a critical aspect in developing technology to help reduce health inequity globally; in fact, this has become a popular concept in various countries for improving quality of care, specifically access to respectful and responsive health care [85].

Conclusions

The proliferation of technology-enabled tools designed to support people in following recommendations for chronic disease self-management has outpaced the research describing the degree to which the Black and Hispanic populations use this technology to support self-management behavior. Although factors driving the general use among the Black and Hispanic populations continue to be investigated, little is known about their impact on health outcomes because of their use. In this paper, we have helped to address this important gap because various technology skills are required to use consumer-oriented HIT designed to support recommended self-management and doing so may require considerable effort from the patients [86]. For example, deciphering the vast and growing amount of information requires that individuals access, assess, and organize various health information. To help elucidate gaps in the literature, we conducted this systematic review to understand the extant literature concerning the use of ICTs among Hispanic and Black people to support chronic disease self-management and highlight potential gaps.

Acknowledgments

The authors would like to acknowledge the health sciences librarians from Rutgers University who assisted in refining our search strategy, specifically Lilyana Todorinova.

Conflicts of Interest

None declared.

Multimedia Appendix 1
A list of search strings and their corresponding databases, the care setting and self-management behavior for each included article, and technology effectiveness in managing chronic disease for each included article.
[DOCX File , 29 KB - jmir_v24i4e22124_app1.docx]

References

1. Wildevuur SE, Simonse LW. Information and communication technology-enabled person-centered care for the "big five" chronic conditions: scoping review. J Med Internet Res 2015;17(3):e77 [FREE Full text] [doi: 10.2196/jmir.3687] [Medline: 25831199]
2. Colen CG, Ramey DM, Cooksey EC, Williams DR. Racial disparities in health among nonpoor African Americans and Hispanics: the role of acute and chronic discrimination. Soc Sci Med 2018;199:167-180 [FREE Full text] [doi: 10.1016/j.socscimed.2017.04.051] [Medline: 28571900]
3. National Center for Health Statistics (US). Health, United States, 2015: with special feature on racial and ethnic health disparities. Hyattsville, MD: National Center for Health Statistics (US); 2016.

4. Osborn CY, de Groot M, Wagner JA. Racial and ethnic disparities in diabetes complications in the northeastern United States: the role of socioeconomic status. J Natl Med Assoc 2013;105(1):51-58 [FREE Full text] [doi: 10.1016/s0027-9684(15)30085-7] [Medline: 23862296]
5. Carpenter R, DiChiacchio T, Barker K. Interventions for self-management of type 2 diabetes: an integrative review. Int J Nurs Sci 2019;6(1):70-91 [FREE Full text] [doi: 10.1016/j.ijnss.2018.12.002] [Medline: 31406872]
6. Mayberry LS, Harper KJ, Osborn CY. Family behaviors and type 2 diabetes: what to target and how to address in interventions for adults with low socioeconomic status. Chronic Illn 2016;12(3):199-215 [FREE Full text] [doi: 10.1177/1742395316644303] [Medline: 27099387]
7. Ritchie CS, Kvale E, Fisch MJ. Multimorbidity: an issue of growing importance for oncologists. J Oncol Pract 2011;7(6):371-374 [FREE Full text] [doi: 10.1200/JOP.2011.000460] [Medline: 22379419]
8. Bluethmann SM, Sciamanna CN, Winkels RM, Sturgeon KM, Schmitz KH. Healthy living after cancer treatment: considerations for clinical and community practice. Am J Lifestyle Med 2018;12(3):215-219 [FREE Full text] [doi: 10.1177/1559827618755681] [Medline: 30283252]
9. Leach CR, Weaver KE, Aziz NM, Alfano CM, Bellizzi KM, Kent EE, et al. The complex health profile of long-term cancer survivors: prevalence and predictors of comorbid conditions. J Cancer Surviv 2015;9(2):239-251. [doi: 10.1007/s11764-014-0403-1] [Medline: 25319681]
10. Subhi Y, Bube SH, Rolskov Bojsen S, Skou Thomsen AS, Konge L. Expert involvement and adherence to medical evidence in medical mobile phone apps: a systematic review. JMIR Mhealth Uhealth 2015;3(3):e79 [FREE Full text] [doi: 10.2196/mhealth.4169] [Medline: 26215371]
11. Magnusson L, Hanson E, Borg M. A literature review study of Information and Communication Technology as a support for frail older people living at home and their family carers. Technol Disabil 2004;16(4):223-235. [doi: 10.3233/tad-2004-16404]
12. Mirkovic J, Kaufman DR, Ruland CM. Supporting cancer patients in illness management: usability evaluation of a mobile app. JMIR Mhealth Uhealth 2014;2(3):e33 [FREE Full text] [doi: 10.2196/mhealth.3359] [Medline: 25119490]
13. Smith A. U.S. smartphone use in 2015. Pew Research Center. 2015. URL: https://www.pewresearch.org/internet/2015/04/01/us-smartphone-use-in-2015/ [accessed 2022-03-11]
14. Anderson M. Racial and ethnic differences in how people use mobile technology. Pew Research Center. 2015. URL: https://www.pewresearch.org/fact-tank/2015/04/30/racial-and-ethnic-differences-in-how-people-use-mobile-technology/ [accessed 2022-03-11]
15. Anderson M, Perrin A. 15% of Americans don't use the internet. Who are they? Pew Research Center. 2015. URL: https://www.pewresearch.org/fact-tank/2021/04/02/7-of-americans-dont-use-the-internet-who-are-they/ [accessed 2022-03-11]
16. Veinot T, Mitchell H, Ancker JS. Good intentions are not enough: how informatics interventions can worsen inequality. J Am Med Inform Assoc 2018;25(8):1080-1088 [FREE Full text] [doi: 10.1093/jamia/ocy052] [Medline: 29788380]
17. Tao D, Shao F, Liu S, Wang T, Qu X. Predicting factors of consumer acceptance of health information technologies: a systematic review. Proc Hum Factors Ergon Soc Annu Meet 2016;60(1):598-602. [doi: 10.1177/1541931213601137]
18. Groeneveld PW, Sonnad SS, Lee AK, Asch DA, Shea JE. Racial differences in attitudes toward innovative medical technology. J Gen Intern Med 2006;21(6):559-563 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00453.x] [Medline: 16808736]
19. Veinot TC, Campbell TR, Kruger DJ, Grodzinski A. A question of trust: user-centered design requirements for an informatics intervention to promote the sexual health of African-American youth. J Am Med Inform Assoc 2013;20(4):758-765 [FREE Full text] [doi: 10.1136/amiajnl-2012-001361] [Medline: 23512830]
20. Nebeker C, Murray K, Holub C, Haughton J, Arredondo EM. Acceptance of mobile health in communities underrepresented in biomedical research: barriers and ethical considerations for scientists. JMIR Mhealth Uhealth 2017;5(6):e87 [FREE Full text] [doi: 10.2196/mhealth.6494] [Medline: 28659258]
21. Gell NM, Grover KW, Humble M, Sexton M, Dittus K. Efficacy, feasibility, and acceptability of a novel technology-based intervention to support physical activity in cancer survivors. Support Care Cancer 2017;25(4):1291-1300. [doi: 10.1007/s00520-016-3523-5] [Medline: 27957621]
22. Puszkiewicz P, Roberts AL, Smith L, Wardle J, Fisher A. Assessment of cancer survivors' experiences of using a publicly available physical activity mobile application. JMIR Cancer 2016;2(1):e7 [FREE Full text] [doi: 10.2196/cancer.5380] [Medline: 28410168]
23. Lorenc T, Petticrew M, Welch V, Tugwell P. What types of interventions generate inequalities? Evidence from systematic reviews. J Epidemiol Community Health 2013;67(2):190-193. [doi: 10.1136/jech-2012-201257] [Medline: 22875078]
24. Lupton D. Health promotion in the digital era: a critical commentary. Health Promot Int 2015;30(1):174-183. [doi: 10.1093/heapro/dau091] [Medline: 25320120]
25. Stowell E, Lyson MC, Saksono H, Wurth RC, Jimison H, Pavel M, et al. Designing and evaluating mHealth interventions for vulnerable populations: a systematic review. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18; April 21-26, 2018; Montreal, Canada p. 1-17. [doi: 10.1145/3173574.3173589]

26. Welch V, Petticrew M, Petkovic J, Moher D, Waters E, White H, PRISMA-Equity Bellagio group. Extending the PRISMA statement to equity-focused systematic reviews (PRISMA-E 2012): explanation and elaboration. J Clin Epidemiol 2016;70:68-89 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.09.001] [Medline: 26348799]
27. Welch V, Petticrew M, Tugwell P, Moher D, O'Neill J, Waters E, PRISMA-Equity Bellagio group. PRISMA-Equity 2012 extension: reporting guidelines for systematic reviews with a focus on health equity. PLoS Med 2012;9(10):e1001333 [FREE Full text] [doi: 10.1371/journal.pmed.1001333] [Medline: 23222917]
28. Chaet AV, Morshedi B, Wells KJ, Barnes LE, Valdez R. Spanish-language consumer health information technology interventions: a systematic review. J Med Internet Res 2016;18(8):e214 [FREE Full text] [doi: 10.2196/jmir.5794] [Medline: 27511437]
29. Dorr D, Bonner LM, Cohen AN, Shoai RS, Perrin R, Chaney E, et al. Informatics systems to promote improved care for chronic illness: a literature review. J Am Med Inform Assoc 2007;14(2):156-163 [FREE Full text] [doi: 10.1197/jamia.M2255] [Medline: 17213491]
30. Heitkemper EM, Mamykina L, Tobin JN, Cassells A, Smaldone A. Baseline characteristics and technology training of underserved adults with type 2 diabetes in the mobile diabetes detective (MoDD) randomized controlled trial. Diabetes Educ 2017;43(6):576-588 [FREE Full text] [doi: 10.1177/0145721717737367] [Medline: 29059017]
31. Nelson LA, Mulvaney SA, Gebretsadik T, Ho YX, Johnson KB, Osborn CY. Disparities in the use of a mHealth medication adherence promotion intervention for low-income adults with type 2 diabetes. J Am Med Inform Assoc 2016;23(1):12-18 [FREE Full text] [doi: 10.1093/jamia/ocv082] [Medline: 26186935]
32. Almeida FA, Smith-Ray RL, Dzewaltowski DA, Glasgow RE, Lee RE, Thomas DS, et al. An interactive computer session to initiate physical activity in sedentary cardiac patients: randomized controlled trial. J Med Internet Res 2015;17(8):e206 [FREE Full text] [doi: 10.2196/jmir.3759] [Medline: 26303347]
33. Collins JL, Champion JD. Assessment of mobile device and SMS use for diet and exercise information among rural Mexican-American adolescents. J Pediatr Nurs 2014;29(6):493-502 [FREE Full text] [doi: 10.1016/j.pedn.2014.03.020] [Medline: 24704179]
34. Davidson TM, McGillicuddy J, Mueller M, Brunner-Jackson B, Favella A, Anderson A, et al. Evaluation of an mHealth medication regimen self-management program for African American and Hispanic uncontrolled hypertensives. J Pers Med 2015;5(4):389-405 [FREE Full text] [doi: 10.3390/jpm5040389] [Medline: 26593951]
35. Davis SA, Carpenter D, Cummings DM, Lee C, Blalock SJ, Scott JE, et al. Patient adoption of an internet based diabetes medication tool to improve adherence: a pilot study. Patient Educ Couns 2017;100(1):174-178. [doi: 10.1016/j.pec.2016.07.024] [Medline: 27480530]
36. Finkelstein J, Bedra M, Li X, Wood J, Ouyang P. Mobile app to reduce inactivity in sedentary overweight women. Stud Health Technol Inform 2015;216:89-92. [Medline: 26262016]
37. Finkelstein J, Wood J. Delivering chronic heart failure telemanagement via multiple interactive platforms. In: Proceedings of the 15th World Multi-Conference on Systemics, Cybernetics and Informatics. 2011 Presented at: WMSCI '11; July 19-22, 2011; Orlando, FL p. 207-212 URL: http://www.iiisci.org/journal/pdv/sci/pdfs/8BA056LN.pdf
38. Fortmann AL, Gallo LC, Garcia MI, Taleb M, Euyoque JA, Clark T, et al. Dulce digital: an mHealth SMS-based intervention improves glycemic control in Hispanics with type 2 diabetes. Diabetes Care 2017;40(10):1349-1355 [FREE Full text] [doi: 10.2337/dc17-0230] [Medline: 28600309]
39. Friedman R, Kazis LE, Jette A, Smith MB, Stollerman J, Torgerson J, et al. A telecommunications system for monitoring and counseling patients with hypertension. Impact on medication adherence and blood pressure control. Am J Hypertens 1996;9(4 Pt 1):285-292. [doi: 10.1016/0895-7061(95)00353-3] [Medline: 8722429]
40. Gerber BS, Stolley MR, Thompson AL, Sharp LK, Fitzgibbon ML. Mobile phone text messaging to promote healthy behaviors and weight loss maintenance: a feasibility study. Health Informatics J 2009;15(1):17-25 [FREE Full text] [doi: 10.1177/1460458208099865] [Medline: 19218309]
41. Green BB, Anderson ML, Ralston JD, Catz S, Fishman PA, Cook AJ. Patient ability and willingness to participate in a web-based intervention to improve hypertension control. J Med Internet Res 2011;13(1):e1. [doi: 10.2196/jmir.1625] [Medline: 21371993]
42. Grimes A, Bednar M, Bolter JD, Grinter RE. EatWell: sharing nutrition-related memories in a low-income community. In: Proceedings of the 2008 ACM conference on Computer supported cooperative work. 2008 Presented at: CSCW '08; November 8-12, 2008; San Diego, CA p. 87-96. [doi: 10.1145/1460563.1460579]
43. Joseph RP, Keller C, Adams MA, Ainsworth BE. Print versus a culturally-relevant Facebook and text message delivered intervention to promote physical activity in African American women: a randomized pilot trial. BMC Womens Health 2015;15:30 [FREE Full text] [doi: 10.1186/s12905-015-0186-1] [Medline: 25886945]
44. Kline KN, Montealegre JR, Rustveld LO, Glover TL, Chauca G, Reed BC, et al. Incorporating cultural sensitivity into interactive entertainment-education for diabetes self-management designed for Hispanic audiences. J Health Commun 2016;21(6):658-668. [doi: 10.1080/10810730.2016.1153758] [Medline: 27166651]
45. Kolmodin MacDonell K, Naar S, Gibson-Scipio W, Lam P, Secord E. The Detroit young adult asthma project: pilot of a technology-based medication adherence intervention for African-American emerging adults. J Adolesc Health 2016;59(4):465-471 [FREE Full text] [doi: 10.1016/j.jadohealth.2016.05.016] [Medline: 27475032]

46. Lin M, Mahmooth Z, Dedhia N, Frutchey R, Mercado CE, Epstein DH, et al. Tailored, interactive text messages for enhancing weight loss among African American adults: the TRIMM randomized controlled trial. Am J Med 2015;128(8):896-904. [doi: 10.1016/j.amjmed.2015.03.013] [Medline: 25840035]
47. Mayberry LS, Berg CA, Harper KJ, Osborn CY. The design, usability, and feasibility of a family-focused diabetes self-care support mHealth intervention for diverse, low-income adults with type 2 diabetes. J Diabetes Res 2016;2016:7586385 [FREE Full text] [doi: 10.1155/2016/7586385] [Medline: 27891524]
48. McGillicuddy JW, Gregoski MJ, Brunner-Jackson BM, Weiland AK, Patel SK, Rock RA, et al. Facilitating medication adherence and eliminating therapeutic inertia using wireless technology: proof of concept findings with uncontrolled hypertensives and kidney transplant recipients. In: Proceedings of the Conference on Wireless Health. 2012 Presented at: WH '12; October 23-25, 2012; San Diego, CA p. 1-9. [doi: 10.1145/2448096.2448108]
49. Newton Jr RL, Carter LA, Johnson W, Zhang D, Larrivee S, Kennedy BM, et al. A church-based weight loss intervention in African American adults using text messages (LEAN study): cluster randomized controlled trial. J Med Internet Res 2018;20(8):e256 [FREE Full text] [doi: 10.2196/jmir.9816] [Medline: 30143478]
50. Nundy S, Razi RR, Dick JJ, Smith B, Mayo A, O'Connor A, et al. A text messaging intervention to improve heart failure self-management after hospital discharge in a largely African-American population: before-after study. J Med Internet Res 2013;15(3):e53 [FREE Full text] [doi: 10.2196/jmir.2317] [Medline: 23478028]
51. Reese JM, Joseph RP, Cherrington A, Allison J, Kim YI, Spear B, et al. Development of participant-informed text messages to promote physical activity among African American women attending college: a qualitative mixed-methods inquiry. J Transcult Nurs 2017;28(3):236-242 [FREE Full text] [doi: 10.1177/1043659616644959] [Medline: 27093904]
52. Reininger B, Mecca LP, Stine KM, Schultz K, Ling L, Halpern D. A type 2 diabetes prevention website for African Americans, Caucasians, and Mexican Americans: formative evaluation. JMIR Res Protoc 2013;2(2):e24 [FREE Full text] [doi: 10.2196/resprot.2573] [Medline: 23846668]
53. Rosal MC, Heyden R, Mejilla R, Capelson R, Chalmers KA, Rizzo DePaoli M, et al. A virtual world versus face-to-face intervention format to promote diabetes self-management among African American women: a pilot randomized clinical trial. JMIR Res Protoc 2014;3(4):e54 [FREE Full text] [doi: 10.2196/resprot.3412] [Medline: 25344620]
54. Shea S, IDEATel Consortium. The informatics for diabetes and education telemedicine (IDEATel) project. Trans Am Clin Climatol Assoc 2007;118:289-304 [FREE Full text] [Medline: 18528511]
55. Skolarus LE, Cowdery J, Dome M, Bailey S, Baek J, Byrd JB, et al. Reach out churches: a community-based participatory research pilot trial to assess the feasibility of a mobile health technology intervention to reduce blood pressure among African Americans. Health Promot Pract 2018;19(4):495-505 [FREE Full text] [doi: 10.1177/1524839917710893] [Medline: 28583024]
56. Trief PM, Izquierdo R, Eimicke JP, Teresi JA, Goland R, Palmas W, et al. Adherence to diabetes self care for White, African-American and Hispanic American telemedicine participants: 5 year results from the IDEATel project. Ethn Health 2013;18(1):83-96. [doi: 10.1080/13557858.2012.700915] [Medline: 22762449]
57. Weinstock RS, Teresi JA, Goland R, Izquierdo R, Palmas W, Eimicke JP, IDEATel Consortium. Glycemic control and health disparities in older ethnically diverse underserved adults with diabetes: five-year results from the Informatics for Diabetes Education and Telemedicine (IDEATel) study. Diabetes Care 2011;34(2):274-279 [FREE Full text] [doi: 10.2337/dc10-1346] [Medline: 21270184]
58. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
59. 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]
60. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
61. Zhang Y, Liu C, Luo S, Xie Y, Liu F, Li X, et al. Factors influencing patients' intentions to use diabetes management apps based on an extended unified theory of acceptance and use of technology model: web-based survey. J Med Internet Res 2019;21(8):e15023 [FREE Full text] [doi: 10.2196/15023] [Medline: 31411146]
62. Starnes JR, Slesur L, Holby N, Rehman S, Miller RF. Predicting no-shows at a student-run comprehensive primary care clinic. Fam Med 2019;51(10):845-849 [FREE Full text] [doi: 10.22454/FamMed.2019.406053] [Medline: 31722103]
63. Ofei-Dodoo S, Kellerman R, Hartpence C, Mills K, Manlove E. Why patients miss scheduled outpatient appointments at urban academic residency clinics: a qualitative evaluation. Kans J Med 2019;12(3):57-61 [FREE Full text] [doi: 10.17161/kjm.v12i3.11793] [Medline: 31489100]
64. Lacy NL, Paulman A, Reuter MD, Lovejoy B. Why we don't come: patient perceptions on no-shows. Ann Fam Med 2004;2(6):541-545 [FREE Full text] [doi: 10.1370/afm.123] [Medline: 15576538]
65. Brown E, Schwartz M, Shi C, Carter T, Shea JA, Grande D, et al. Understanding why urban, low-income patients miss primary care appointments: insights from qualitative interviews of west Philadelphians. J Ambul Care Manage 2020;43(1):30-40. [doi: 10.1097/JAC.0000000000000316] [Medline: 31770184]

66. Brim C. A descriptive analysis of the non-urgent use of emergency departments. Nurse Res 2008;15(3):72-88. [doi: 10.7748/nr2008.04.15.3.72.c6458] [Medline: 18459488]
67. Zeber JE, Pearson D, Smith D. Analysis of health appointment no-shows. Prim Health Care 2009;19(2):25-29. [doi: 10.7748/phc2009.03.19.2.25.c6895]
68. Ten Haken I, Ben Allouch S, van Harten WH. The use of advanced medical technologies at home: a systematic review of the literature. BMC Public Health 2018;18(1):284 [FREE Full text] [doi: 10.1186/s12889-018-5123-4] [Medline: 29482550]
69. Henriksen K, Joseph A, Zayas-Cabán T. The human factors of home health care: a conceptual model for examining safety and quality concerns. J Patient Saf 2009;5(4):229-236. [doi: 10.1097/PTS.0b013e3181bd1c2a] [Medline: 22130216]
70. Blumberg SJ, Luke JV. Wireless substitution: early release of estimates from the national health interview survey, July-December 2017. National Center for Health Statistics. 2018. URL: https://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201806.pdf [accessed 2019-12-29]
71. Vizer LM, Eschler J, Koo BM, Ralston J, Pratt W, Munson S. "It's Not Just Technology, It's People": constructing a conceptual model of shared health informatics for tracking in chronic illness management. J Med Internet Res 2019;21(4):e10830 [FREE Full text] [doi: 10.2196/10830] [Medline: 31033452]
72. Figueiredo M, Caldeira C, Chen Y, Zheng K. Routine self-tracking of health: reasons, facilitating factors, and the potential impact on health management practices. AMIA Annu Symp Proc 2017;2017:706-714 [FREE Full text] [Medline: 29854136]
73. Samuel-Hodge CD, Headen SW, Skelly AH, Ingram AF, Keyserling TC, Jackson EJ, et al. Influences on day-to-day self-management of type 2 diabetes among African-American women: spirituality, the multi-caregiver role, and other social context factors. Diabetes Care 2000;23(7):928-933. [doi: 10.2337/diacare.23.7.928] [Medline: 10895842]
74. Nundy S, Mishra A, Hogan P, Lee SM, Solomon MC, Peek ME. How do mobile phone diabetes programs drive behavior change? Evidence from a mixed methods observational cohort study. Diabetes Educ 2014;40(6):806-819 [FREE Full text] [doi: 10.1177/0145721714551992] [Medline: 25278512]
75. Anderson-Lewis C, Darville G, Mercado RE, Howell S, Di Maggio S. mHealth technology use and implications in historically underserved and minority populations in the United States: systematic literature review. JMIR Mhealth Uhealth 2018;6(6):e128 [FREE Full text] [doi: 10.2196/mhealth.8383] [Medline: 29914860]
76. Gomersall T, Madill A, Summers LK. A metasynthesis of the self-management of type 2 diabetes. Qual Health Res 2011;21(6):853-871. [doi: 10.1177/1049732311402096] [Medline: 21429946]
77. Lehrer HM, Dubois SK, Brown SA, Steinhardt MA. Resilience-based diabetes self-management education: perspectives from African American participants, community leaders, and healthcare providers. Diabetes Educ 2017;43(4):367-377. [doi: 10.1177/0145721717714894] [Medline: 28614997]
78. Williams IC, Clay OJ, Ovalle F, Atkinson D, Crowe M. The role of perceived discrimination and other psychosocial factors in explaining diabetes distress among older African American and White adults. J Appl Gerontol 2020;39(1):99-104 [FREE Full text] [doi: 10.1177/0733464817750273] [Medline: 29291677]
79. Mitchell UA, Chebli PG, Ruggiero L, Muramatsu N. The digital divide in health-related technology use: the significance of race/ethnicity. Gerontologist 2019;59(1):6-14. [doi: 10.1093/geront/gny138] [Medline: 30452660]
80. Grande SW, Sherman LD. Too important to ignore: leveraging digital technology to improve chronic illness management among Black men. J Med Internet Res 2018;20(5):e182 [FREE Full text] [doi: 10.2196/jmir.9434] [Medline: 29759956]
81. Claudel SE, Bertoni AG. Exploring the use of personal technology in type 2 diabetes management among ethnic minority patients: cross-sectional analysis of survey data from the lifestyle intervention for the treatment of diabetes study (LIFT Diabetes). JMIR Diabetes 2018;3(1):e5 [FREE Full text] [doi: 10.2196/diabetes.8934] [Medline: 30291086]
82. Jang M, Johnson CM, D'Eramo-Melkus G, Vorderstrasse AA. Participation of racial and ethnic minorities in technology-based interventions to self-manage type 2 diabetes: a scoping review. J Transcult Nurs 2018;29(3):292-307. [doi: 10.1177/1043659617723074] [Medline: 28826353]
83. Valdez RS, Brennan PF. Embracing complexity: rethinking culturally informed design in human factors/ergonomics and consumer health informatics. Int J Hum–Comput Interact 2017;33(4):322-332. [doi: 10.1080/10447318.2016.1269429]
84. Montague E, Winchester W, Valdez R, Vaughn-Cooke M, Perchonok J. Considering culture in the design and evaluation of health IT for patients. Proc Hum Factors Ergon Soc Annu Meet 2013;57(1):1088-1092. [doi: 10.1177/1541931213571242]
85. Jongen C, McCalman J, Bainbridge R, Clifford A. Cultural competence in health: a review of the evidence. Berlin, Germany: Springer; 2018.
86. Senteio C, Veinot T. Trying to make things right: adherence work in high-poverty, African American neighborhoods. Qual Health Res 2014;24(12):1745-1756. [doi: 10.1177/1049732314549027] [Medline: 25212857]

Abbreviations

BP: blood pressure
HbA_{1c}: hemoglobin A_{1c}
HIT: health information technology
ICT: information and communication technology
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by G Eysenbach; submitted 03.07.20; peer-reviewed by R Newton, M Peeples; comments to author 18.08.20; revised version received 25.09.20; accepted 23.01.22; published 04.04.22.

Please cite as:
Senteio C, Murdock PJ
The Efficacy of Health Information Technology in Supporting Health Equity for Black and Hispanic Patients With Chronic Diseases: Systematic Review
J Med Internet Res 2022;24(4):e22124
URL: https://www.jmir.org/2022/4/e22124
doi:10.2196/22124
PMID:35377331

Review

From Digital Health to Digital Well-being: Systematic Scoping Review

Merlijn Smits[1], MSci, EDAC; Chan Mi Kim[2], MSci; Harry van Goor[1], MD, Prof Dr, PhD; Geke D S Ludden[2], Prof Dr, PhD, IR

[1]Department of Surgery, Radboud University Medical Center, Nijmegen, Netherlands

[2]Department of Design, Production, and Management, Faculty of Engineering Technology, University of Twente, Enschede, Netherlands

Corresponding Author:
Merlijn Smits, MSci, EDAC
Department of Surgery
Radboud University Medical Center
Geert Grooteplein Zuid 10
Nijmegen, 6525 GA
Netherlands
Phone: 31 634165922
Email: merlijnsmits@hotmail.com

Abstract

Background: Digital health refers to the proper use of technology for improving the health and well-being of people and enhancing the care of patients through the intelligent processing of clinical and genetic data. Despite increasing interest in well-being in both health care and technology, there is no clear understanding of what constitutes well-being, which leads to uncertainty in *how* to create well-being through digital health. In an effort to clarify this uncertainty, Brey developed a framework to define problems in technology for well-being using the following four categories: epistemological problem, scope problem, specification problem, and aggregation problem.

Objective: This systematic scoping review aims to gain insights into how to define and address well-being in digital health.

Methods: We followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist. Papers were identified from 6 databases and included if they addressed the design or evaluation of digital health and reported the enhancement of patient well-being as their purpose. These papers were divided into design and evaluation papers. We studied how the 4 problems in technology for well-being are considered per paper.

Results: A total of 117 studies were eligible for analysis (n=46, 39.3% design papers and n=71, 60.7% evaluation papers). For the *epistemological problem*, the thematic analysis resulted in various definitions of well-being, which were grouped into the following seven values: *healthy body*, *functional me*, *healthy mind*, *happy me*, *social me*, *self-managing me*, and *external conditions*. Design papers mostly considered well-being as *healthy body* and *self-managing me*, whereas evaluation papers considered the values of *healthy mind* and *happy me*. Users were rarely involved in defining well-being. For the *scope problem*, patients with chronic care needs were commonly considered as the main users. Design papers also regularly involved other users, such as caregivers and relatives. These users were often not involved in evaluation papers. For the *specification problem*, most design and evaluation papers focused on the provision of care support through a digital platform. Design papers used numerous design methods, whereas evaluation papers mostly considered pre-post measurements and randomized controlled trials. For the *aggregation problem*, value conflicts were rarely described.

Conclusions: Current practice has found pragmatic ways of circumventing or dealing with the problems of digital health for well-being. Major differences exist between the design and evaluation of digital health, particularly regarding their conceptualization of well-being and the types of users studied. In addition, we found that current methodologies for designing and evaluating digital health can be improved. For optimal digital health for well-being, multidisciplinary collaborations that move beyond the common dichotomy of design and evaluation are needed.

(J Med Internet Res 2022;24(4):e33787) doi:10.2196/33787

KEYWORDS

well-being; design; evaluation; technology assessment; digital health; eHealth; mHealth; telehealth; mobile phone

Introduction

Background

Digital health technologies are increasingly being used to monitor, manage, and support health and well-being. The use of digital health promises to increase access to health information, improve the quality of care, reduce errors, and stimulate healthy behavior [1]. Digital health delivery is also often referred to as *eHealth* [2]. Subsets of eHealth are the application of a specific technology in health delivery (eg, mobile health technologies [3]) or the use of digital technology for a specific purpose (eg, remote delivery of health care services: *telehealth* or *telemedicine* [4,5]). The COVID-19 crisis has stimulated the use of digital health, and its application is expected to increase in the coming years [6-8].

The term *digital health* is defined as the "proper use of technology for improving the health and wellbeing of people at individual and population levels, as well as enhancing the care of patients through intelligent processing of clinical and genetic data" [9]. In this paper, we aim to shed light on the term *well-being* in this definition. Studies on the meaning of well-being date back to ancient Greece. Since then, many disciplines have reflected on this term. Well-being is commonly considered as "a state of persons which designates that they are happy or flourishing and that their life is going well for them" [10]. It is seen as "the highest value to which other values can be subsumed" [10]. In this, *values* are considered as everything people consider important in life [11].

Well-being is gaining increasing interest in both health care and technology. In health care, *health* has long been regarded as the absence of disease or infirmity. Its formal definition was changed to *well-being* in 1948. In that year, the World Health Organization redefined health as "a state of complete physical, mental, and social wellbeing" [12]. In technology, well-being is often considered to be a central value in the design process. Many technologies aim to improve the well-being of their users. To that end, an increasing number of design methodologies exist that aim to guide design for well-being processes [10].

However, there is no clear explanation of what (values) constitute well-being, which leads to differences in understanding well-being and much uncertainty on how to enhance well-being through digital health. Philosopher Philip Brey [10] categorizes uncertainty by means of 4 problems that are paramount in technology for well-being. First, he describes the *epistemological problem.* This problem refers to the definition of well-being that should be embedded in technology design. This includes questions on how users consider their own well-being and how designers should obtain an understanding of the conceptions of users. The *scope problem* refers to questions on the *who* and *when* of considering well-being in design. A design can improve the well-being of its main user group but could also include needs from indirect users. In addition, the focus of improving well-being could be short term, long term, or both. The third problem, the *specification problem*, includes all the questions related to the embodiment of well-being in technology design: how should well-being be translated into design requirements, and how can one ensure that users will experience improved well-being when using technology? The fourth problem, the *aggregation problem*, refers to dealing with conflicts that arise between the contradictory values of well-being. Value conflicts can occur in one user (eg, when choices for optimal short-term well-being do not correspond to choices necessary for long-term well-being) or between users of similar or different groups (when an increase in the well-being of one user leads to a decrease in the well-being of another user). The 4 major questions in technology for well-being challenge the ability to create digital health for optimal well-being.

Objective

In this study, we conducted a systematic scoping review to facilitate future practitioners in the process of creating digital health for the well-being of patients. Digital health commonly follows two processes before adoption takes place: *design* and *evaluation*. Design is the process in which technology is created with the objective of solving a specific problem in health care. Evaluation is the process of (scientifically) assessing the added value of a technology to decide whether to adopt it as part of standard care. We will reflect on the current practices in both design and evaluation to find an answer to how to deal with the 4 problems paramount in technology for well-being.

Methods

Eligibility Criteria

The PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist was followed to ensure the reliability of our results [13]. To meet the eligibility criteria, the included papers had to (1) address the design or evaluation of digital health, (2) report increased well-being as the purpose of digital health, (3) address the patient as the main user, and (4) be published in English. Digital health is mainly designed to improve the well-being of patients. We particularly focused on patients as the main users to locate our research in the context of health care. Age or other criteria were not attributed to the type of patient considered. However, we will reflect on the other users, such as caregivers, who are considered in the design and evaluation processes, in addition to the patient. All types of original peer-reviewed research papers were considered. Reviews were excluded from the search as they did not provide in-depth insights into the 4 guiding problems for each digital health case. To obtain a complete overview of all digital health papers, we did not include any inclusion restrictions on the publication period (see Textbox 1 for an overview of the eligibility criteria).

Textbox 1. Eligibility criteria for the systematic scoping review.

Inclusion criteria

- Study types: any type of original peer-reviewed research paper
- Period: any paper published before February 24, 2021
- Language: English
- Population: any patient receiving care for prevention, cure, or rehabilitation
- Intervention: the design or evaluation of digital health technology
- Outcome: the well-being of the patient

Exclusion criteria

- Study types: conference abstracts, (systematic) reviews, opinion papers, editorials, doctoral theses, workshops, protocols, and textbooks
- Period: papers published after February 24, 2021
- Language: all other languages
- Population: healthy people, health in the workplace, and health of caregivers
- Intervention: no interaction among patient and technology, medical aids (such as electronic wheelchair or intravenous pump), does not address design or evaluation of digital technology, or technical description of the technology
- Outcome: the digital technology does not aim to improve well-being; digital technology only aims to log well-being without considering the impact on well-being and well-being only referred to in the abstract and not in full text

Search Strategy

Papers were collected from six databases on February 24, 2021: ACM Digital Library, PubMed, Web of Science, IEEE Xplore, PhilPapers, and Google Scholar. We used a combination of the following terms appearing in the title or abstract of the papers: (*wellbeing* OR *well-being*) AND (*patient**) AND (*design* OR *moral** OR *ethic**) AND (*technology* OR *digital* OR *ehealth* OR *mhealth* OR *telemedicine* OR *telehealth* OR *electronic health* OR *mobile health* OR *mobile** OR *smart** OR *internet*; Multimedia Appendix 1). The terms *moral* and *ethic* were added to the search strategy as these terms might result in insight into various values related to well-being. The term *design* in this section was used as a synonym for the term *technology* in digital health.

Study Selection

After collecting all papers from the databases, duplicates were removed. The titles and abstracts of the remaining papers were screened for eligibility criteria using Rayyan software [14] by 2 authors independently (MS and CK; Textbox 1). The full texts of the remaining papers were downloaded and independently screened for inclusion by the same authors. In both phases, the authors discussed disagreements until consensus was reached.

Data Analysis

Data were extracted by 1 author (MS), who was guided by the 4 theoretical problems and their subquestions to understand if and how these problems are considered in design and evaluation practices (see Textbox 2 for subquestions for each problem directly derived from Brey [10]). The papers were divided into two groups based on their main focus: design and evaluation of digital health. Sections of each paper were marked when they were related to 1 of the 4 problems and their corresponding questions. All marked sections were grouped by question. Consequently, we compared the results of the various papers for each question to understand how each major problem was considered in practice. To determine what values constitute well-being in the epistemological problem, we followed the principles of thematic content analysis [15]. We coded each definition of well-being and categorized them into 7 overarching values. The results are presented for each problem.

Textbox 2. Categories and guiding questions for data extraction in the systematic scoping review.

General information

- Type of research: design or evaluation
- Year
- Journal

Epistemological problem

- What values constitute well-being?
- Who defined well-being?
- What research tools are used to understand user well-being?

Scope problem

- Who is the main user of the technology?
- What other users are involved?
- What time span of well-being is considered?

Specification problem

- What type of technology is designed?
- What method is used to design or evaluate digital health?
- How is well-being translated into design requirements?

Aggregation problem

- Does the paper refer to conflicts in well-being?
- What type of conflicts are considered?
- How are conflicts solved?

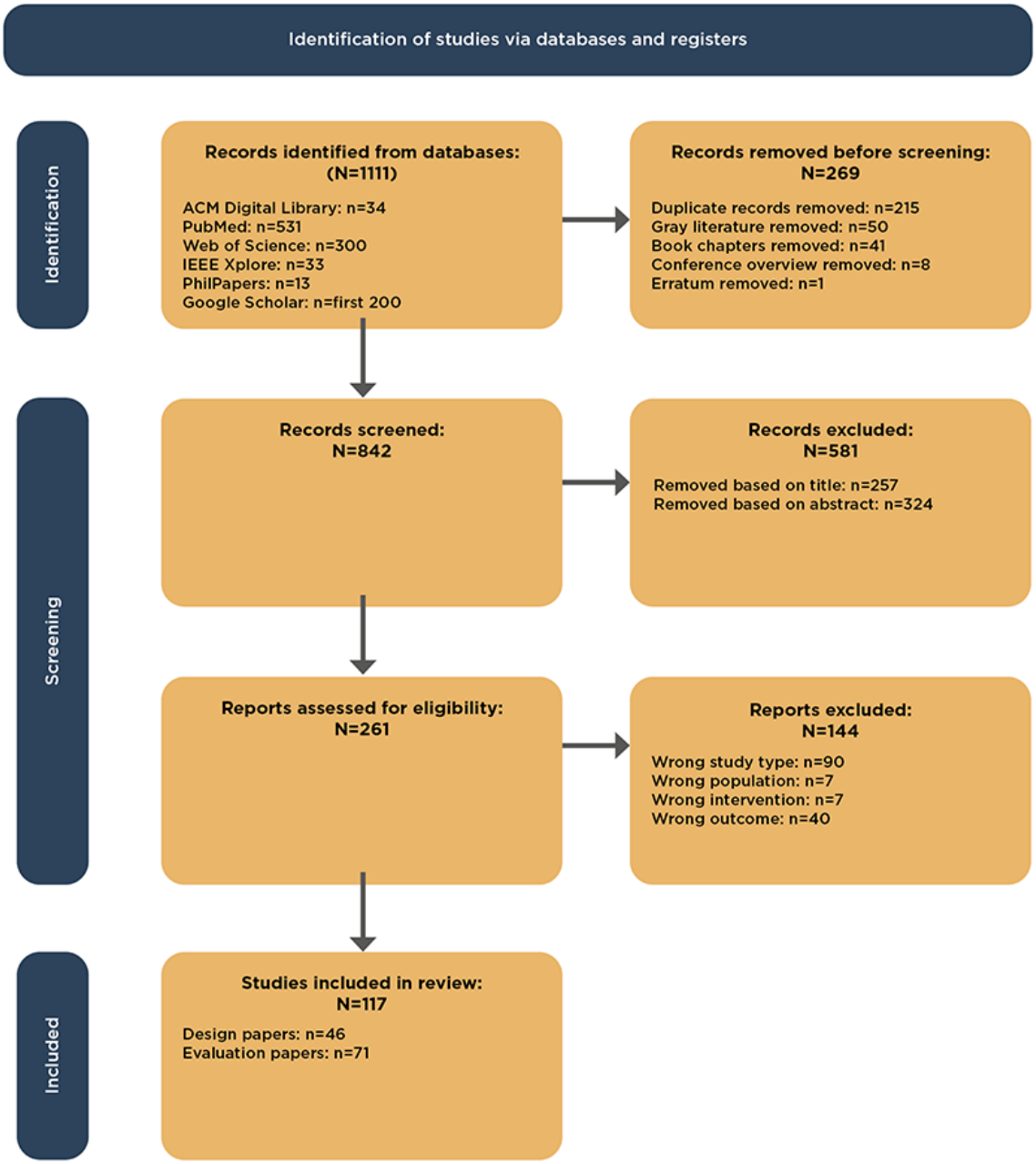
Results

Study Selection

A total of 1111 papers were identified. After removing duplicates and gray literature, 75.79% (842/1111) of papers remained. After title and abstract screening, 69% (581/842) of papers were discarded mainly as they did not study a digital health intervention or did not consider the patient as the main user of the technology. Of the remaining 261 papers, full-text screening resulted in the exclusion of 90 (34.5%) papers for not being original research, 40 (15.3%) papers for only reporting well-being but not considering it in the design or evaluation process, and 14 (5.4%) papers for either not considering patients as the main user or digital health as an intervention (see Figure 1, which was derived from the PRISMA [Preferred Reporting Items for Systematic Reviews and Meta-Analyses] 2020 flow diagram [16]).

Of the 261 papers, 117 (44.8%) papers remained for data extraction, of which 46 (39.3%) papers focused on the design of digital health, and 71 (60.7%) addressed its evaluation. Although some papers described both processes, it was possible to group them according to their main focus. The included papers were published between 1999 and 2021. Approximately 80.3% (94/117) of all the papers were written in or after 2015.

Figure 1. Flowchart of paper selection.



General Results

The results were presented for each of the 4 problems (see Table 1 and Table 2 for a summary of the general results and Multimedia Appendix 2 [17-133] for more detailed information).

Table 1. Overview of results for design papers[a] (N=117).

Categories, review questions, and characteristics	Design papers (n=46), n (%)
General information	
Year	
After 2010	45 (98)
After 2018	24 (52)
Journal	
JMIR or partner journals	10 (22)
Other journals	36 (78)
Epistemological problem	
What is the definition of well-being?	
Healthy body	21 (46)
Healthy mind	15 (33)
Happy me	14 (30)
Social me	15 (33)
Self-managing me	21 (46)
Who defined well-being?	
Author	31 (67)
User	15 (33)
If users defined well-being, what research tools are used?	
Interview	11 (24)
Survey	2 (4)
Data derived from smartphone	1 (2)
Focus group and workshop	4 (9)
Observation	0 (0)
Scope problem	
Who is the main user of the technology?	
Neoplasms	3 (7)
Mental, behavioral, and neurodevelopmental disorder	10 (22)
Endocrine, nutritional, or metabolic diseases	5 (11)
Other	28 (60)
What other users are involved?	
No other users involved	11 (24)
Involvement of caregiver	26 (57)
Involvement of relatives	8 (17)
Involvement of experts	18 (39)
What time span of well-being is considered?	
N/A[b]	N/A
Specification problem	
What type of technology is designed?	
Support platform	30 (65)
Sensor	8 (17)
Phone or video support	0 (0)

Categories, review questions, and characteristics	Design papers (n=46), n (%)
Other	8 (18)
What method is used to design or evaluate digital health?	
Interviews	16 (35)
Focus group	13 (28)
Usability test	15 (33)
Other	2 (4)
How is well-being translated into design requirements?	
Not specified but generally based on insights from user research	—[c]
Aggregation problem	
Does the paper refer to conflicts in well-being?	
Yes	11 (14)
No	35 (76)
What type of conflicts are considered?	
Value conflicts within one user	1 (2)
Value conflicts within users of the same group	2 (4)
Value conflicts between users of different groups	10 (22)
How are conflicts solved?	
Personalization	20 (43)

[a]Some papers are categorized in multiple domains, which might result in percentages of >100%. This table only visualizes the most common outcomes. For more detailed information, see Multimedia Appendix 2 [17-133] and the *Results* section.

[b]N/A: not applicable.

[c]Not available.

Table 2. Overview of results for evaluation papers[a] (N=117).

Categories, review questions, and characteristics	Evaluation papers (n=71), n (%)
General information	
Year (from 1999 to 2021)	
After 2010	67 (94)
After 2018	42 (59)
Journal	
JMIR or partner journals	11 (15)
Other journals	60 (85)
Epistemological problem	
What is the definition of well-being?	
Healthy body	28 (39)
Healthy mind	42 (59)
Happy me	35 (49)
Social me	28 (39)
Self-managing me	20 (28)
Functional me	12 (17)
External conditions	6 (8)
Who defined well-being?	
Author	53 (75)
Without QoL[b] questionnaire	5 (7)
With QoL questionnaire: a total of 154 different QoL questionnaires were used, of which 14 were disease specific)	48 (68)
PHQ-9[c] used most often	8 (11)
User	22 (31)
If users defined well-being, what research tools are used?	
Interview	9 (13)
Survey	10 (14)
Data derived from smartphone	1 (1)
Focus group and workshop	0 (0)
Observation	3 (4)
Scope problem	
Who is the main user of the technology?	
Neoplasms	11 (15)
Mental, behavioral, and neurodevelopmental disorder	10 (14)
Endocrine, nutritional, or metabolic diseases	6 (8)
Other	44 (63)
What other users are involved?	
No other users involved	59 (83)
Involvement of caregiver	9 (13)
Involvement of relatives	3 (4)
Involvement of experts	0 (0)
What time span of well-being is considered?	
During use	13 (18)

Categories, review questions, and characteristics	Evaluation papers (n=71), n (%)
Directly after use	43 (61)
0 to 1 month after use	4 (6)
1 to 3 months after use	11 (15)
3 to 6 months after use	10 (14)
6 to 12 months after use	9 (13)
Specification problem	
What type of technology is designed?	
Support platform	38 (54)
Sensor	7 (10)
Phone or video support	12 (17)
Other	14 (19)
What method is used to design or evaluate digital health?	
Pre-post measurement	23 (32)
Randomized controlled trial	22 (31)
Other	26 (37)
How is well-being translated into design requirements?	
N/A[d]	N/A
Aggregation problem	
Does the paper refer to conflicts in well-being?	
Yes	5 (7)
No	66 (93)
What type of conflicts are considered?	
Value conflicts within one user	1 (1)
Value conflicts within users of the same group	4 (6)
Value conflicts between users of different groups	1 (1)
How are conflicts solved?	
Personalization	31 (44)

[a]Some papers are categorized in multiple domains, which might result in percentages of >100%. This table only visualizes the most common outcomes. For more detailed information, see Multimedia Appendix 2 [17-133] and the *Results* section.

[b]QoL: Quality of Life.

[c]PHQ-9: Patient Health Questionnaire-9.

[d]N/A: not applicable.

Epistemological Problem

Overview

The epistemological problem refers to the definition of well-being in digital health, which defines well-being and the methods used to understand user well-being (Table 3).

Table 3. Epistemological problem of well-being in digital health (N=117).[a]

Values of well-being and who defined well-being	Design methodology (n=46)	Evaluation methodology (n=71)
Healthy body		
Defined by author	• [17-31]	—[b]
Defined by user	• Interview [32-36] • Workshop [37] • Focus group [34]	• Interview [38,39] • Survey [40-42]
Validated questionnaire	—	• Big Five Inventory [43] • Checklist Individual Strength [44] • EQ-5D[c] [45] • HRQoL[d] [46] • HRQoL-MacNew[e] [47] • LTPAQ[f] [48] • PAIS[g] [49] • Parkinson's Disease Questionnaire-39 [50] • Pittsburgh Sleep Quality Index [51] • PWI-A[h] [52] • Quality of Well-being Scale [53] • Multilevel Assessment Instrument [49] • Short Form-12 Health Survey [45] • Timed Up and Go Test [54] • WHO-QOL BREF[i] [55] • 6-Minute Walking Test [54] • QoLAD[j] (Alzheimer disease) [56] • Short Form-36 Health Survey [46,54,57,58] • SwQoR[k] [59] • BCTRI[l] (breast cancer) [49] • FACT-B[m] (breast cancer) [60,61] • QoLBC[n] (breast cancer) [62] • Chronic Respiratory Disease Questionnaire (chronic respiratory disease) [63] • Chronic Respiratory Questionnaire (chronic respiratory disease) [63] • IVI-VLV[o] (low vision) [64,65]
Functional me		
Defined by author	—	—
Defined by user	—	• Survey [66]
Validated questionnaire	—	• EQ-5D [45] • Instrumental activities of daily living [54] • MFHW[p] (German) [67] • MIDAS[q] [68] • HRQoL [46] • HRQoL-MacNew [47] • Parkinson's Disease Questionnaire-39 [50] • Sheehan Disability Scales [69] • Short Form-12 Health Survey [45] • Short Form-36 Health Survey [46,54,57,58] • FACIT[r] (chronic illness) [70]
Healthy mind		
Defined by author	• [17,24,26,31,35,71-73]	—

Values of well-being and who defined well-being	Design methodology (n=46)	Evaluation methodology (n=71)
Defined by user	• Interview [34,74-76] • Focus group [34] • Survey [77,78] • Data derived from smartphone [79]	• Interview [80] • Survey [40,41,81,82] • Observation [83]
Validated questionnaire	—	• Beck Depression Inventory [51,82] • BRIEF-A[s] [84] • Depression, Anxiety, and Stress Scale-21 [52,85] • EQ-5D [45] • EuroQol-5 Dimension [86] • Generalized Anxiety Disorder-7 [69,86,87] • HADS[t] [44,58,88-92] • HRQoL [46] • HRQoL-MacNew [47] • Kessler 10-Item Scale [69] • Paced Auditory Serial Addition Test [84] • PCL-C[u] [51] • Parkinson's Disease Questionnaire-39 [50] • Patient Health Questionnaire-9 [48,69,84-87,93,94] • Patient Health Questionnaire-15 [86] • Perceived Stress Scale [51] • Response to Stressful Experiences Scale [51] • Short Form-12 Health Survey [45] • Short Form-36 Health Survey [46,54,57,58] • The Short Health Anxiety Inventory [86] • Social Readjustment Rating Scale [82] • Simon Task [84] • SwQoR [59] • Wechsler Adult Intelligence Scale-4 [84] • WEMWBS[v] [95,96] • 5-Item Well-Being Index [97,98] • WHO-QOL BREF [55] • Work and Social Adjustment Scale [86] • Zung Self‐Rating Anxiety Scale [99] • Zung Self‐Rating Depression Scale [99] • QoLAD (Alzheimer disease) [56] • BCTRI (breast cancer) [49] • QoLBC (breast cancer) [62] • FACIT (chronic illness) [70] • Chinese 15-item Diabetes Distress Scale (diabetes) [85] • Diabetes Distress Scale (diabetes distress) [93] • IVI-VLV (low vision) [64,65] • SCI-QoL[w] (spinal cord injury) [48,87]
Happy me		

Values of well-being and who defined well-being	Design methodology (n=46)	Evaluation methodology (n=71)
Defined by author	• [20,23,100-104]	—
Defined by user	• Interview [32,74-76] • Focus group [105] • Workshop [37] • Data derived from smartphone [79]	• Interview [38,106,107] • Survey [40,81,82,108-110] • Observation [83,111]
Validated questionnaire	—	• Acceptance and Action Questionnaire-2 [96] • Client Satisfaction Questionnaire [92] • Fordyce Happiness Scale [88] • General Self-efficacy Scale [84] • HeiQ[x] [112] • Herth Hope Scale [99] • Humor Styles Questionnaire [43] • MFHW (German) [67] • Meaning in Life Questionnaire [99] • Orientation to Happiness Scale [52] • PAL-C[y] [49] • Positive and Negative Affect Schedule [43,52] • PWI-A [52] • Self-Compassion Scale [51] • Subjective Happiness Scale [113] • Self‐transcendence Scale [99] • Satisfaction With Life Scale [52] • Visual Analog Scale mood [88] • WEMWBS [95,96] • 5-Item Well-Being Index [97,98] • WHO-QOL BREF [55] • QoLAD (Alzheimer disease) [56] • FACT-B (breast cancer) [60,61] • QoLBC (breast cancer) [62] • FACIT Spiritual Well-being scale–12 (chronic illness) [83] • Diabetes Distress Scale (diabetes) [93] • ICECAP-O[z] (older adults) [45] • IVI-VLV (low vision) [64,65]
Social me		

Values of well-being and who defined well-being	Design methodology (n=46)	Evaluation methodology (n=71)
Defined by author	• [20,25,71,100-102,114-119]	• [120]
Defined by user	• Interview [36,76] • Focus group [105]	• Interview [38,80,121] • SMS text message analysis [122] • Observation [121]
Validated questionnaire	—	• Client Satisfaction Questionnaire [58] • EQ-5D [85] • HeiQ [112] • HRQoL-MacNew [47] • Multidimensional Scale of Perceived Social Support [51,110,113] • PAIS [49] • PAL-C [49] • Parkinson's Disease Questionnaire–39 [50] • PWI-A [52] • Dyadic Adjustment Scale [51] • Short Form-12 Health Survey [45] • Short Form-36 Health Survey [46,54,57,58] • Ability to participate in Social Roles and Activities [48] • University of California Los Angeles Loneliness Scale [110,113] • WHO-QOL BREF [55] • QoLAD (Alzheimer disease) [56] • FACT-B (breast cancer) [60,61] • QoLBC (breast cancer) [62] • FACIT (chronic illness) [70] • Diabetes Distress Scale (diabetes) [93] • ICECAP-O (older adults) [45] • SCI-QoL (spinal cord injury) [48,87]
Self-managing me		
Defined by author	• [18-20,25,26,71,100,101,114-119,123,124]	• [125-128]
Defined by user	• Interview [34,76,129] • Focus group [34,105,130]	• Interview [38,54,80,106,107,121,131] • Observation [121] • Survey [81,132] • Data derived from smartphone [133]
Validated questionnaire	—	• 20-item Diabetes Empowerment Scale [85] • EQ-5D [45,85] • HeiQ [112] • SwQoR [59] • QoLBC (breast cancer) [62] • Diabetes Distress Scale (diabetes) [93] • 14-item Summary for Diabetes Self-care Activities (diabetes) [85] • ICECAP-O (older adults) [45]
External conditions		

Values of well-being and who defined well-being	Design methodology (n=46)	Evaluation methodology (n=71)
Validated questionnaire	—	• PAIS [49] • PWI-A [52] • WHO-QOL BREF [55] • QoLAD (Alzheimer disease) [56] • QoLBC (breast cancer) [62] • ICECAP-O (older adults) [45]

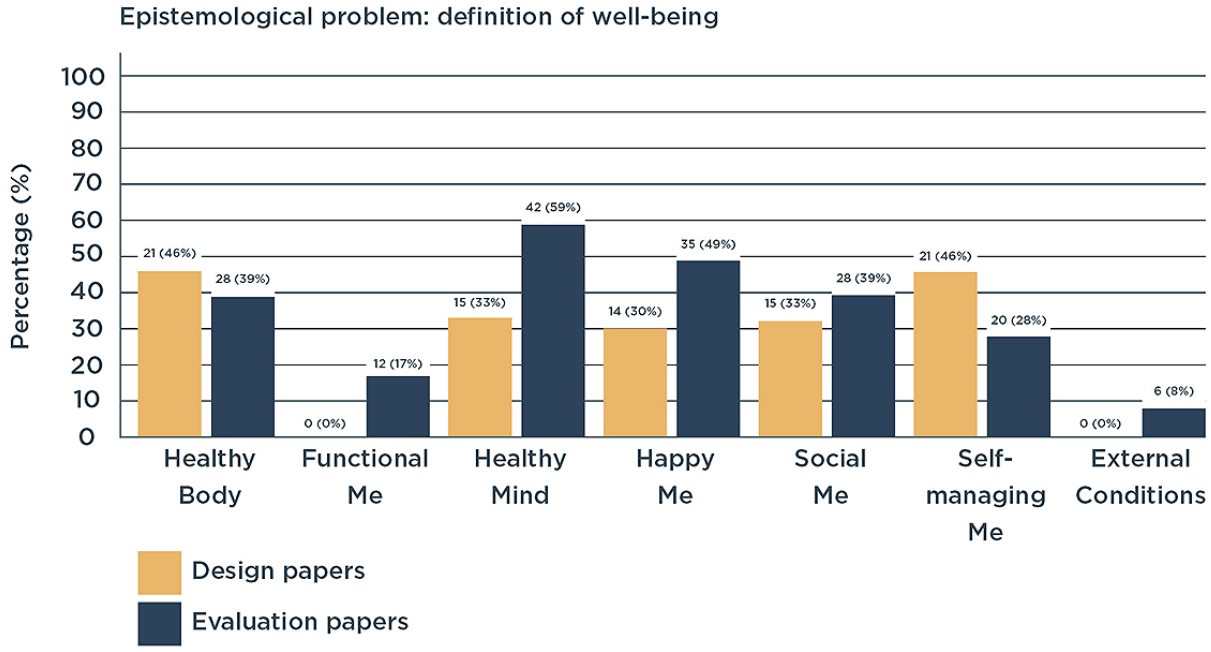
[a]Quality of life questionnaires were downloaded and analyzed by the authors and grouped according to the values of well-being that they considered.
[b]Not available (no papers found in this category).
[c]EQ-5D: EuroQol 5-Dimensions.
[d]HRQoL: Health-Related Quality of Life.
[e]HRQoL-MacNew: The MacNew Heart Disease Health-Related Quality of Life.
[f]LTPAQ: Leisure Time Physical Activity Questionnaire.
[g]PAIS: Psychosocial Adjustment to Illness Scale.
[h]PWI-A: Personal Well-being Index (Australian Version).
[i]WHO-QOL BREF: World Health Organization Quality of Life Questionnaire–Brief.
[j]QoLAD: Quality of Life in Alzheimer's Disease.
[k]SwQoR: Swedish web version of the Quality of Recovery Questionnaire.
[l]BCTRI: Breast Cancer Treatment Response Inventory.
[m]FACT-B: Functional Assessment of Cancer Therapy–Breast.
[n]QoLBC: Quality of Life in Breast Cancer Patients.
[o]IVI-VLV: Vision Impairment–Very Low Vision Questionnaire.
[p]MFHW: Marburger Screening for Habitual Well-being (German).
[q]MIDAS: Migraine Disability Assessment Questionnaire.
[r]FACIT: Functional Assessment of Chronic Illness Therapy.
[s]BRIEF-A: Behavior Rating Inventory of Executive Function–Adult Version.
[t]HADS: Hospital Anxiety and Depression Questionnaire.
[u]PCL-C: Posttraumatic Stress Disorder Checklist–Civilian version.
[v]WEMWBS: Warwick–Edinburgh Mental Well-being Scale.
[w]SCI-QoL: Spinal Cord Injury Quality of Life subscales.
[x]HeiQ: Health Education Impact Questionnaire.
[y]PAL-C: Profile of Adaptation to Life Clinical Scale.
[z]ICECAP-O: ICEpop Capability measure for Older people.

Values Constituting Well-being

Overview

Papers that considered well-being in their study design were often vague in their definitions. For example, Kayrouz et al [69] studied digital health as part of the *well-being course* but did not define well-being. In such cases, we derived the definition from the research tools used in the papers. Most papers applied terms such as *mental well-being*, *emotional well-being*, or *spiritual well-being*. In addition, combinations of these words were used, as follows: *psychosocial well-being*, *psychospiritual well-being*, or *biopsychosocial well-being*. In comparing the different definitions of well-being, we found that similar terms did not always refer to the same content and that 2 varying terms were used to say the same thing. For example, spiritual well-being was used to refer to not only self-transcendence [99] but also self-acceptance [83]. In addition, depression and anxiety were termed interchangeably as belonging to the domains of *emotional well-being* [90], *mental well-being* [87], and *psychophysical well-being* [89]. For these reasons, we refrained from using these terms. Instead, we categorized the various definitions of well-being into values. In this review, we related values to what people consider important for obtaining well-being. After thematic content analysis, we identified seven values: *healthy body*, *functional me*, *healthy mind*, *happy me*, *social me*, *self-managing me*, and *external conditions* (Figure 2).

Figure 2. Values of well-being in design and evaluation papers (in number of papers and percentage of the total amount of papers in each process).



Healthy Body

This value of well-being was commonly considered in design and evaluation papers. It relates to health in its traditional definition: the absence of disease in the body. In the papers, this item was often referred to as *physical well-being*. Within this value, it is common to study the effect of disease on items such as pain experience, obesity, incontinence, sleep quality, vital sign monitoring, delirium, and sexual functioning.

Functional Me

Functional me refers to the ability to execute activities of daily living and reach life goals affected by one's health condition. This value of well-being was identified based on a set of validated questionnaires used in evaluation papers, of which the 36-item Short Form Health Survey was used the most. None of the design papers considered *functional me* explicitly in their design.

Healthy Mind

A *healthy mind* was considered regularly in design papers and even more in evaluation papers. It referred to the absence of mental disease and was, in papers, mostly referred to as *mental well-being* or *emotional well-being*. Depression and anxiety are typical items that relate to this value. In addition, the effects of mental disease belong to this value, such as the effects of bipolar disorder, posttraumatic stress disorder, schizophrenia, and (mild) cognitive decline.

Happy Me

This value was considered by one-third of all design papers and half of all evaluation papers. It comprised the ability to feel happy, flourish, have a meaningful life, and accept one's own body. In the reviewed papers, this value was often referred to as *psychological well-being*, *subjective well-being*, *emotional well-being*, *spiritual well-being*, *happiness*, and *wellness*. Items that are often considered related to this value are hope, satisfaction, positive experiences, pleasure, fulfilling personal potential, feeling needed, self-transcendence, and personal growth. Other important items are the ability to cope with and accept one's own health status, self-confidence, being proud, feeling dignity, and not feeling stigmatized.

Social Me

Social Me was considered in approximately one-third of all design and evaluation papers. It includes all personal relationships that people have and the evaluation of these relationships. This value was consistently termed in the papers as *social well-being*. Items belonging to this value are related to conversations, feelings of partnership and friendship, compassion, trust, empathy, and support. Relationships studied within the papers included those with partners, family members, friends, and health care providers.

Self-managing Me

Approximately half of all design papers considered this item compared with fewer evaluation papers. *Self-managing me* relates to the ability to understand and manage one's health care condition autonomously. Only *psychological well-being* was sometimes related to this value; no other specific terms were used within the papers. Concepts such as autonomy, competence, confidence, free will, decision-making, empowerment, and self-understanding belong to this category. The ability to understand and make use of health care information to make well-informed decisions, or *health literacy*, is also important in this context.

External Conditions

External conditions do not directly refer to a personal ability as other values do. In addition, this item was not studied in design and was rarely studied in evaluation papers. Nonetheless, we

included it as several evaluation questionnaires dedicated questions to these conditions, and these could not be grouped elsewhere. Such questions referred to external conditions that created the setting for the well-being of patients and included, among others, financial security and having a job and a house.

User Definition of Well-being

In most design and evaluation papers, well-being was solely defined by the authors (eg, researchers, physicians, designers, and engineers). Design papers commonly referenced the literature to justify their definitions of well-being. Evaluation papers made use of validated quality of life (QoL) questionnaires, from which the authors derived a definition of well-being. A total of 154 unique questionnaires on well-being were found in the evaluation papers. Of these 154 questionnaires, 14 (9.1%) could only be applied to a specific disease. In only one-third of all design and evaluation papers, the main users of digital health were questioned about their understanding of well-being. In these cases, users were often not *free* to define the several values of well-being but were questioned on how they considered a specific item, such as *physical well-being*. Design papers considered these user definitions in their design requirements. Evaluation papers considered these as a qualitative outcome of the evaluation process. In general, these definitions of users do not seem to vary from the definitions provided by the authors.

Research Tools Used for Understanding User Well-being

In the few papers in which users provided insight into what they considered as well-being, several empirical research tools were used. The most common tools used to understand users' conceptualization of well-being were interviews and open nonvalidated surveys. Design papers commonly considered interviews or workshops. Evaluation papers mostly considered surveys. In 1.7% (2/117) of papers, the concept of well-being was defined through data collected based on smartphone use [79,122].

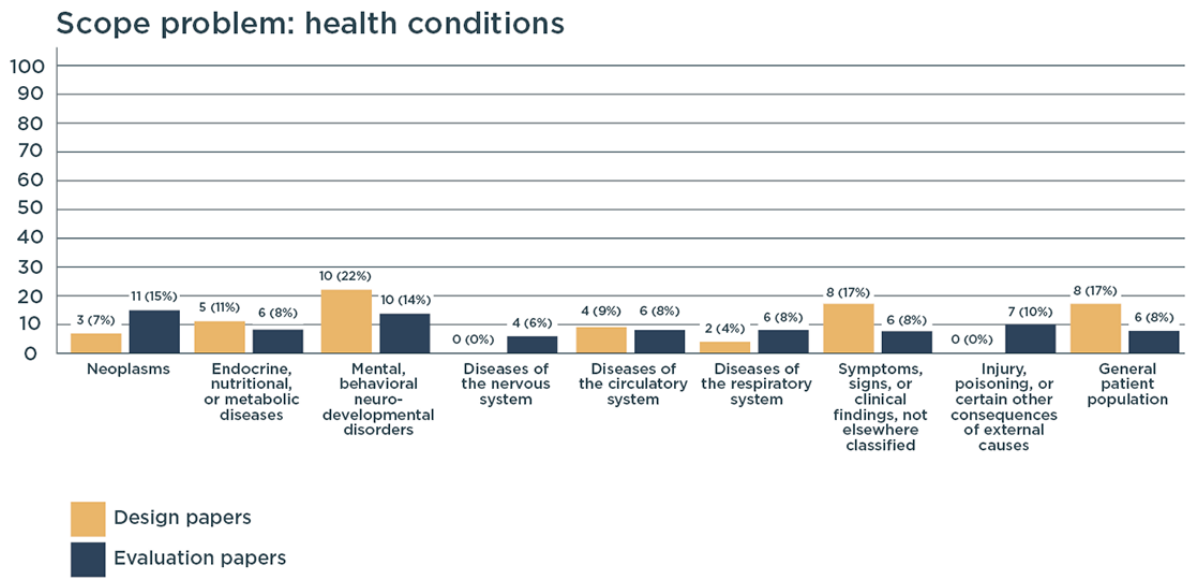
Scope Problem

To understand how the scope problem is considered in the design and evaluation of digital health, we studied the main users of the technology, the additional users, and the time frame considered.

Main User of the Technology

We grouped the health conditions of the main users of digital health (eg, patients) according to the International Classification of Diseases for Mortality and Morbidity [134] (Figure 3). Design papers showed that most digital health solutions were created for patients with mental disorders, in particular depression, anxiety, and cognitive decline. Evaluation papers mostly considered mental disorders and neoplasms.

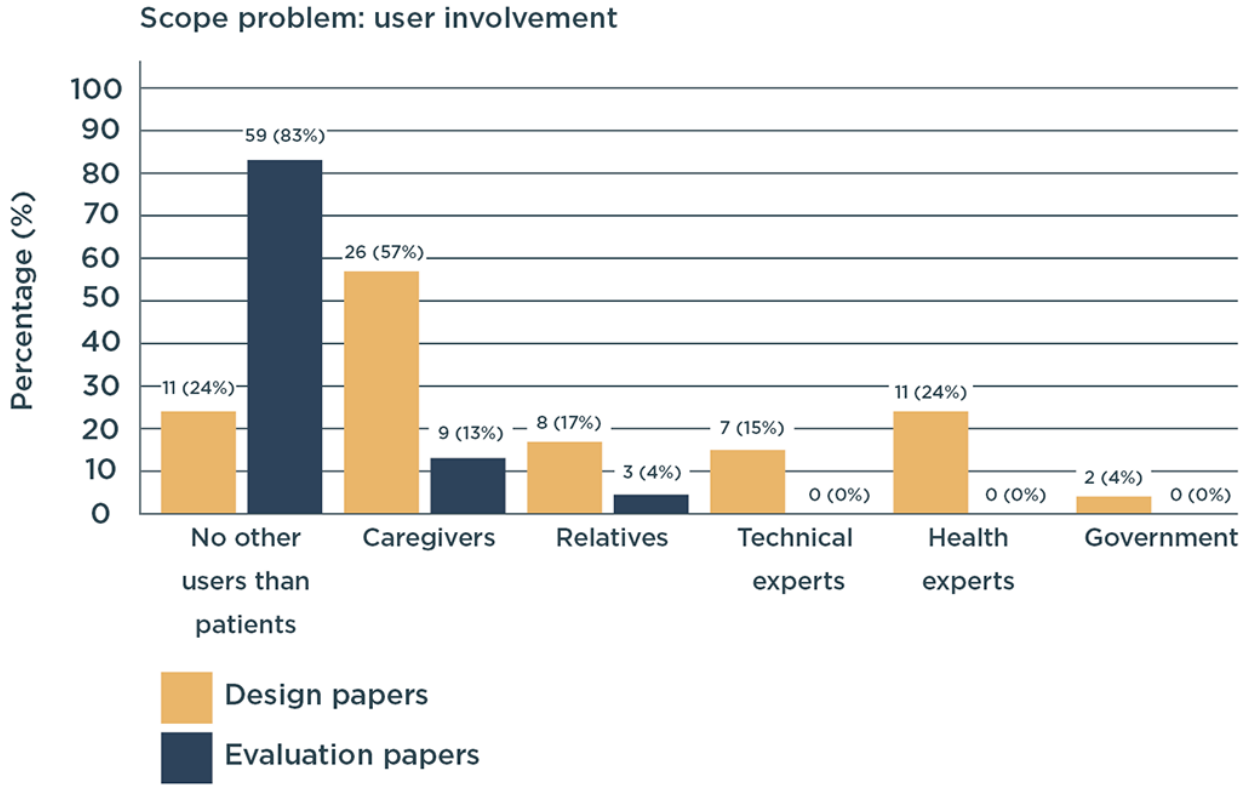
Figure 3. Disease classification of patients in design and evaluation papers (in number of papers and percentage of the total amount of papers in each process).



Other Users Involved

In addition to the main users—patients—we studied what other users were considered in each paper. Design papers commonly considered other users (eg, caregivers and relatives). In contrast, only a few evaluation papers considered other users. The reason for involving other users in the papers was mainly to provide input regarding patients' needs rather than their own. Recognizing the value of involving a wide variety of users in design, 4% (2/46) of design papers conducted an in-depth exploration of the types of users to involve in the design process [35,118] (Figure 4).

Figure 4. Other users considered in design and evaluation papers (in number of papers and percentage of the total amount of papers in each process). Categories with fewer than 4 papers in total were left out of the figure.



Time Span of Well-being

In evaluation papers, the effect of technology on well-being was mostly measured during the use of the technology or on the day directly after its use. Only a few papers conducted an additional measurement after 1 to 12 months from initial use.

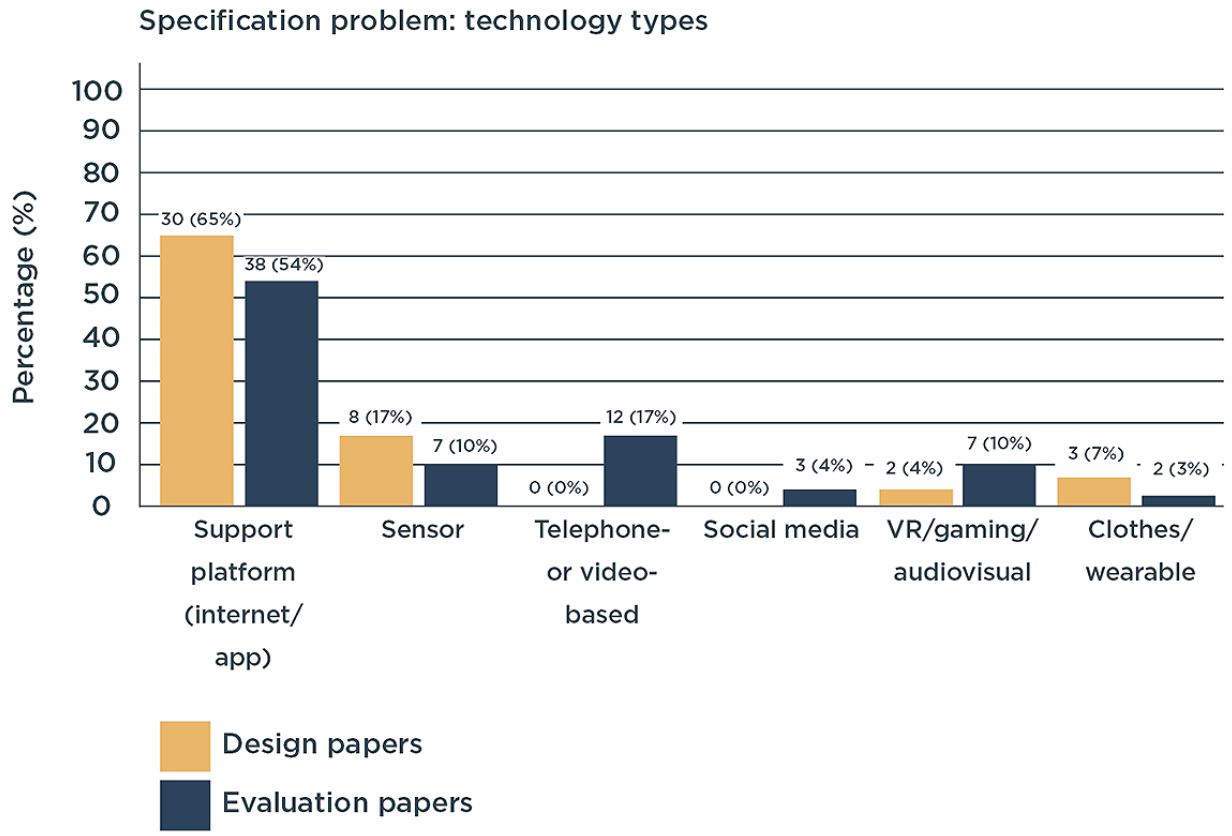
Specification Problem

The specification problem relates to the embodiment of well-being in the design of digital health. To achieve this, we extracted the type of technology, the design or evaluation method used, and the procedure for creating design requirements.

Type of Technology

Most design and evaluation papers studied the application of a supporting digital platform. Such a platform is provided to users through the internet or via a tablet or smartphone app. Few design papers also considered the design of sensors or wearables. Other technologies considered by the evaluation papers were mainly telephone- or video-based consulting and support (Figure 5). The great majority of the digital interventions were designed and evaluated for use at home. Only very few were found to be used within the hospitals, primary care settings, or public spaces.

Figure 5. Technology types considered in design and evaluation papers (in number of papers and percentage of the total amount of papers in each process). VR: virtual reality.



Method for Designing and Evaluating Digital Health

In design papers, a wide range of methodologies was used to create digital health, and most of them were focused on the inclusion of the user in the design process through participatory design. Interviews, focus groups, and workshops were regularly used. In addition, usability testing through small pilot studies and prototype testing were popular tools for the design and refinement of technologies. Methods structuring the design process greatly varied and included, for example, the methods of service design for value networks [118], persona enrichment process [119], social network analysis [35], systems development [19], transformative service research [100], and human factors research [20,101]. In addition, numerous papers applied varying frameworks to design for behavior change [18,22,25,26,32,34,116].

In comparison with design papers, evaluation papers were more consistent in their methodologies. Most papers applied a pre–post study design or a randomized controlled trial (RCT) to evaluate the digital health intervention compared with a baseline or control group. Another type of evaluation paper considered interviews to understand the usability and acceptability of the technology, sometimes in addition to a pre–post study or an RCT. More rare forms of evaluation were population surveys [81,127,128] or analysis of technology use through big data analysis [33,48,91,122]. An alternative to the RCT was explored once, named *partially randomized patient preference*. In partially randomized patient preference, patients were allocated based on their preference in either the intervention or control group. The authors concluded that the intervention has higher efficacy when patients have consciously chosen for its use [58].

Translation Into Design Requirements

Design papers commonly did not explain the methodology used to translate user input into design requirements. We only identified a few papers that illustrated their procedures. For example, requirements were created by coding user input into requirements [101], and the requirements were explicitly elucidated by users in a workshop [37].

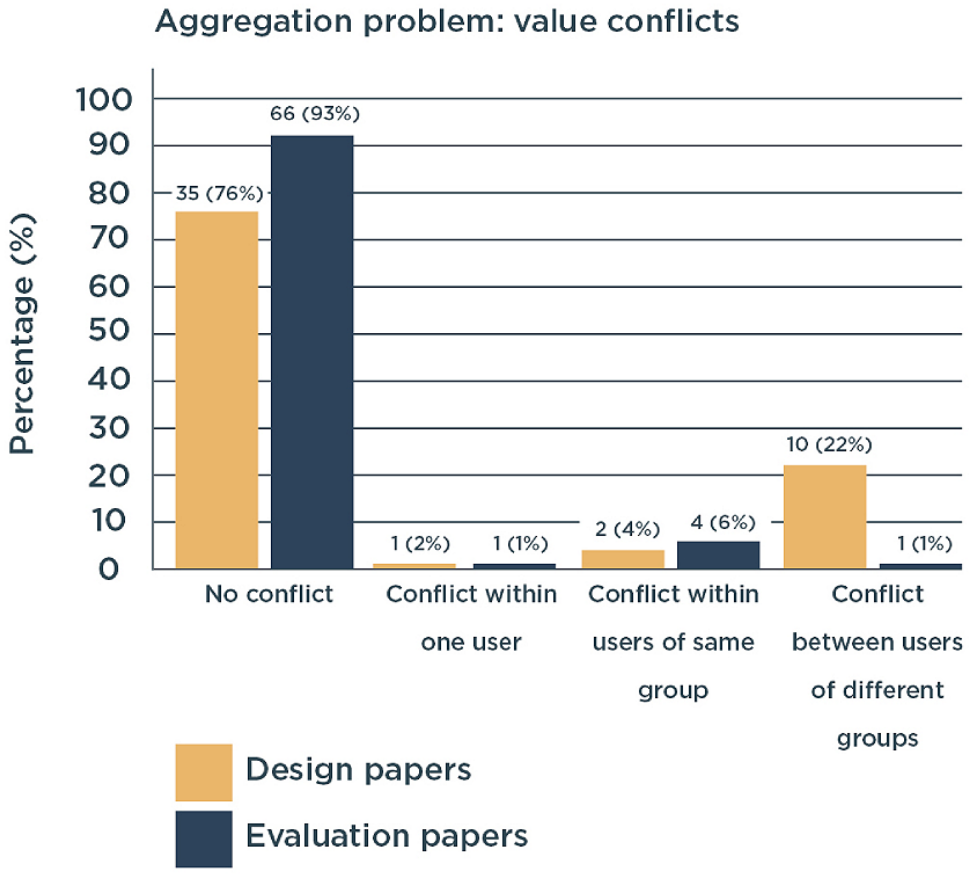
Aggregation Problem

The aggregation problem refers to conflicts within the values of well-being. We aimed to understand if such value conflicts were reported and, if so, the type of conflicts that arose and the solution for solving these conflicts.

Conflicts in Well-being

The great majority of papers did not consider value conflicts. Design papers considered conflicts more often than the evaluation papers (Figure 6). For example, Doherty et al [74] considered a set of value tensions as a source of design inspiration.

Figure 6. Value conflicts considered in design and evaluation papers (in number of papers and percentage of the total amount of papers in each process).



Type of Conflicts

Value conflicts within one user were explicitly considered in 1.7% (2/117) of papers. These papers studied both the benefits and harms of digital health for individual well-being [74,82]. For example, the ability to obtain support through technology versus the reduction in personal contact was contrasted. Approximately 4% (2/46) of design papers [74,130] and 6% (4/71) of evaluation papers [38,81,89,133] considered a conflict arising among users of the same group. Such conflicts generally referred to the differences among individual patients in their desire to apply digital health [38,89,133] or their ability to use it [81,130]. Conflicts among users of different groups were considered by 22% (10/46) of design papers and only 1% (1/71) of evaluation papers [127]. This difference might have resulted from the evaluation papers involving only patients and no other users in the study design. Such conflicts mostly occurred between patients and caregivers. For example, Kujala et al [127] illustrated the use of digital health to improve patient autonomy and decrease caregiver autonomy. Cahill et al [101] showed conflicts among organizational needs (ie, staff costs and keeping residents safe) and patient needs (ie, independence, privacy, and social interaction). In another paper, the same authors explained that although digital health might benefit patients, it might hinder nurses' working processes [20]. Derboven et al [117] addressed a conflict in the autonomy of patients and control over patients by caregivers (Figure 6).

Conflict Resolution

Most studies only referred to conflicts without providing solutions. The few solutions offered were procedural, such as engaging in multidisciplinary collaborations [27], weighing the benefits and harms of conflicting values [117], and being aware of conflicts [102]. Another solution for solving value conflicts within users was offered by Doherty et al [75] and was related to personalizing digital health based on individual needs. The topic of digital health personalization was found in approximately half of all design and evaluation papers but was rarely explicitly related to the topic of value conflicts. Multiple personalization options were addressed. For example, papers recommended providing individualized health advice through digital health [17,23,57,66,77,78,122], adjusting software to patient needs [20,39,41,48,95,98,107,108,110,112,114,122,133], the importance of also supporting nonusers [18,38,44,75,81,89,91,113], and personalizing the support needed to apply digital health [35,49,56,59,100,101,117]. Other personalization options included changing motivational gaming techniques to individual needs [22,41,132], adapting solutions to specific cultures [60,129,133], allowing patients to make their own motivational messages [86], choosing the gender of the digital health assistant [124], and inserting personal memories into the design [99,111].

Discussion

Principal Findings

The enormous growth in papers since 2010 reflects that digital health is increasingly being applied to improve the well-being of patients. In addition, this type of technology is rapidly advancing. Although studies on digital health started with CDs and telephone consulting, today's studies increasingly cover health care provision via web-based supporting platforms, social media, and remote sensing. In all of these cases, we identified how practice deals with the commonly discussed theoretical problems in technology for well-being. We defined common values related to well-being and identified that well-being was rarely defined by the users themselves. In addition, we identified that the current scope of well-being is generally small, involving few users in a short time frame. We illustrated that many methodologies exist on how to embed well-being in design, whereas only a few methods are accepted for evaluation. Finally, we identified that value conflicts, commonly discussed in theory [10,135], are rarely considered in practice. Our results show that simple solutions exist for solving theoretical problems that are paramount in technology for well-being. At the same time, this theory can challenge current practices for continuous improvement.

Definition of Well-being

Within health care, health is considered a state of complete physical, mental, and social well-being. After the introduction of this definition in 1948, it was scarcely applied in practice [136,137] and regularly criticized for being unrealistic. Such criticisms mostly applied to the word *complete*, as this word implies that people cannot feel healthy without absolute physical, mental, and social well-being. In an aging society, *complete* well-being would then only be reserved for a few individuals [138]. To specify the novel definition of health as well-being, the concept of *positive health* was introduced by various authors [139-142]. Positive health focuses on health through well-being. It is the ability to *flourish* despite mental or physical diseases. In addition, as a response to assess health from the perspective of well-being, an entire range of validated questionnaires emerged, often termed QoL measurement scales [137]. Although used often in recent times, the terms of *well-being*, *positive health*, and *QoL* are *rarely defined*, which made Locker and Gibson [143] conclude that the commitment to these concepts is *more rhetorical than real*.

Our review sheds light on how well-being is conceptualized in the context of digital health provision. First, we found that, unfortunately, a large majority of papers only reported on well-being without actually implementing it, which corresponds to the rhetorical commitment to well-being postulated by Locker and Gibson [143]. In the design and evaluation papers that committed to well-being, we were able to identify seven values commonly considered as part of well-being: *healthy body*, *functional me*, *healthy mind*, *happy me*, *social me*, *self-managing me*, and *external conditions*. The often-occurring values of *healthy body*, *healthy mind*, and *social me* reflect the definition of health as physical, mental, and social well-being, as proposed by the World Health Organization. The value of *happy me* was found to greatly reflect the movement of positive health, aiming for health care to make a move from disease toward happiness and flourishing. Remarkably, we also commonly found *self-managing* me to be an important value of well-being related to digital health. Previously, this value was often not considered to be a part of patients' well-being. Digital technologies have the ability to locate the center of health from hospital to home and from health care provider to patient. With this ability, they can increase the autonomy of patients. We believe that with the growing use of digital health, the importance given to the value of *self-managing me* will increase. As a consequence, it might be important to constantly reflect on the interactions among patients, technologies, and values of well-being to optimally design health care for patients' well-being. The identified values can be used as a source of inspiration for digital health designers and evaluators to co-design technologies for optimal well-being with patients.

Differences Between Design and Evaluation

We identified more evaluation papers than design papers. This difference might result from design research not always reporting on its processes [144]. In addition, we found differences between the design and evaluation papers regarding the 4 problems. Design papers mostly considered well-being as *healthy body* and *self-managing me*. Evaluation papers often considered the values of *healthy mind* and *happy me*. Design papers mainly focused on mental disorders. Evaluation papers also studied mental disorders and neoplasms. Other users, such as caregivers and relatives, were regularly involved in design papers but not in evaluation papers. Design papers used numerous design methods, whereas evaluation papers mostly considered pre-post measurements or RCTs as a method. Value conflicts were rarely described in design papers and even less in evaluation papers.

Clearly, differences exist between the processes of design and evaluation, given that they take place in different stages of the digital health development process. Design processes commonly take place outside the care context. These processes are both creative and exploratory. In contrast, evaluation follows the rules of *evidence-based health* and is located in the health care environment [145]. Evaluation is more bound to standardization and reflection. The different characteristics of both processes result in the need for different methodologies for design and evaluation, as discussed in the specification problem. However, the differences seem to be larger than what can be explained by the stage of development. One would expect that the design and evaluation of digital health consider an equal approach to epistemological and scope problems by adopting the same definition of well-being, the same ways of obtaining this definition, and focusing on the same user groups. Nevertheless, this review shows different results. Design and evaluation vary greatly in their definitions of well-being and the main patient groups that are commonly designed for. This is troublesome, considering, for example, an application for postoperative patients to monitor their health. The design team creates this solution to improve the value of *self-managing me* for these patients. During the evaluation, this technology is assessed based on its ability to improve the *healthy body* and the *healthy mind*. The misalignment between the design input and evaluation

outcomes results in suboptimal insight into the potential of the technology, which could hinder its successful implementation. Currently, only a minority of digital health technologies have achieved successful implementation. More digital health technologies would achieve successful implementation when the design is better aligned with evaluation and vice versa [146,147]. In particular, we argue that design and evaluation should consider similar definitions of well-being in the design and evaluation processes and the same user groups.

Methods of Creating Well-being

In most design and evaluation papers, well-being was defined without user input. When users were questioned, they were commonly asked to define only a specific value of well-being instead of explaining the values they considered important. Brey [10] previously noted that well-being could not be objectively determined independently from the user. Our results are worrisome, as the papers did not obtain a real understanding of what users considered to be important values belonging to their well-being. Involving users in defining the values of well-being as a source of design inspiration and evaluation outcome is recommended.

Methods for designing and evaluating digital health clearly differ. In most design papers, a participatory design was applied. However, as Orlowski and Matthews [148] argued, in addition to participatory design, a design method for structuring the process is needed. The papers in this review considered a wide variety of design methods to structure the design process. A review of the methods for usability testing of eHealth showed similar results [144]. No consistency seems to exist in the design papers on what design methods to use to design for well-being. The question of what design method to use to create well-being has been raised in the past [149]. Multiple papers have been written on the potential to embed the values of well-being in technology design. Design methods such as value-sensitive design [150] or values that matter [151] have been introduced. In addition, numerous papers have aimed at creating *design for well-being* methodologies to provide designers with a framework for embodying well-being in design. Examples in the literature are the approaches of emotional design [152]; life-based design [153]; capability-sensitive design [154]; positive design [155]; motivation, engagement, and thriving in user experience model [156]; positive technology [157]; experience design [158]; and positive computing [159]. Given this myriad of design methodologies that particularly aim at our goal of designing for well-being, it is remarkable that none of these methods was found to be used in the design papers. We speculate that this resulted from not knowing about the design methods, finding the methods difficult to apply to health care, or valuing well-known methods over novel ones. Future digital health designers would benefit from heuristics on which method to use in what situation. This can be facilitated by increasing awareness of design methods and transparency and reporting on the reasons for using a certain design method.

The interaction between users and technology and the context of use might affect how well-being is expressed, which does not necessarily correspond to the initial embodiment of well-being in design. This is called the *positivist problem* [160]. For that reason, an evaluation process is necessary to study the actual effects of technology on well-being. Evaluation papers commonly considered the same set of evaluation methodologies: pre-post measurements and RCTs.

The RCT methodology, which has become the gold standard for effectiveness studies in pharmacological interventions, has been transferred one to one to evaluate the effectiveness of nonpharmacological interventions, including digital health [147]. An RCT only provides valuable information on an intervention's effectiveness when (1) the intervention and the way of providing the intervention are stable, (2) the intervention can be applied with fidelity, and (3) when it is expected that the outcomes of the intervention are measurable and meaningful [161]. Given these requisites, the use of RCTs in digital health has been subject to several concerns. An RCT requires that the intervention is stable, implying that the digital health technology has been finalized before the start of the RCT. However, design is an iterative process that requires numerous phases of testing and adaptation. Clinical outcomes obtained through an RCT are often required to justify the use of more resources to improve the design. However, when the design is improved based on the clinical outcomes of RCTs, these outcomes become directly outdated. Similarly, when a digital health solution needs to await clinical outcome measures for improvement, the technology might be outdated once the trial is finished [162]. In addition, a digital health solution cannot be directly applied with high fidelity as the solution mediates clinical processes and might require changes to the context of use to be optimally used. However, these effects require further study. The limited study design of an RCT does not permit the study of such mediating effects, and without knowing the optimal context of use, digital health cannot be applied with high fidelity [163].

Several changes and alternatives to RCTs have been proposed to align scientific evaluations with digital health. These include, among others, a multiphase optimization strategy for the RCT, allowing the design to be adapted during the evaluation process [164], evaluating the principle of a solution rather than the specific technology itself [162], and broadening the set of outcome measures into the inclusion of human–technology interactions [165] and legal and ethical evaluations [166,167]. Another interesting alternative is a single case experimental design [168]. This method illustrates that RCTs only provide an *average* good and not the optimal solution for each individual. They propose to observe a single case over a longer period while manipulating treatment (technology). This study design allows personalizing digital health to an individual's needs, thereby increasing its efficacy. So far, the papers in this review have barely considered moving beyond traditional evaluation methods. Our results resemble previous results from systematic reviews of evaluation methods and health technology assessments of digital health [169,170]. There is a need for the use of evaluation methods that are better aligned with the complexity of digital health.

In addition, not only RCTs but also the many validated QoL questionnaires to assess the effect of digital health via an RCT need reflection. In the past, these questionnaires were studied thoroughly [137,171]. Guyatt et al [171] rightfully posed the concern of how to select, use, and interpret such questionnaires.

We share their concerns and believe that when applying such a questionnaire, it is important to understand and illustrate what values of well-being are measured through the questionnaires and to be aware of the potential and limitations of what is being asked. Similarly, as Blandford et al [147] argued, such questionnaires only have a limited face and construct validity. They do not provide insight into how the social structures of care change and how technology is changing the current values of well-being [172]. In addition, QoL questionnaires were validated for a specific user group in a specific context. When technology is introduced, the entire basis on which such questionnaires are validated might change. Thus, the use of QoL questionnaires requires careful consideration of their selection and application. Our 7 values provide a comprehensive and up-to-date view of well-being. Further work could be done to develop tools (eg, questionnaires) based on the 7 values to complement and support the current evaluation methods to better align with the complexity of digital health.

Recommendations for Best Practice

On the basis of the theoretical and practical reflections of this review, we sketched an *ideal practice* for anyone involved in the design and evaluation of digital health for well-being. This ideal practice begins with the formation of multidisciplinary teams working together from the start of the design process to successful implementation. After the composition of the team, we advise demarcating the scope of the project. Which users will be affected by the technology? What users will be studied, and who will be left out of the study's scope? Furthermore, we encourage the upfront identification of the moments in time during which the effect of technology is evaluated. The process should continue with a clear study on the definition of well-being per user group in which users are closely involved. To prevent well-being from becoming a buzzword, the team should consider constantly aligning the following design and evaluation processes to the found values of well-being. Instead of two linear processes, design and evaluation take place simultaneously. Evaluation outcomes are the source of design input and vice versa. As many methods exist for designing for well-being, the team should decide together which method is most suitable. The reasoning process is reported to facilitate other teams to make their own decisions. To justify the complexity of digital health, the type of evaluation method is carefully considered. The chosen method enables the evaluation of the user–technology–value interaction in an authentic context of use. In addition, the method facilitates obtaining insight into individual experiences for the personalization of the solution.

Strengths and Limitations

In this paper, we aimed to bridge the gaps between the practices of design and evaluation and the theory produced on digital health technologies aimed at improving well-being. This endeavor is both a strength and a limitation. By highlighting the differences among the fields, we enable the design to consider the context of evaluation and vice versa, inspire practice to consider better theoretical insights, and guide theorists to acknowledge the need for pragmatic decision-making in practice. The application of our recommendations would result in individual, social, and economic benefits. First, at the individual level, digital health would foster a culture of *inclusiveness* through personalization. It would better meet the needs and values of individual patients, caregivers, and relatives, regardless of their education, age, gender, or culture. Second, on a social level, digital health would fit within current health care practices and align with the care processes of health care personnel. Third, economic benefits will arise from the alignment of design with evaluation by preventing a waste of resources and leading to a more successful uptake of digital health.

A limitation of our study is that it was difficult to report insights from all disciplines using a common language. The attempt to place all insights from practice into theoretical frameworks might have resulted in missing important items or mistakenly interpreting certain practical phenomena as belonging to certain theoretical concepts. For example, we found that a great majority of design papers aimed to create digital health technologies for lifestyle management of the older population, regardless of their medical condition. By categorizing all design papers into the International Classification of Diseases for Mortality and Morbidity framework, these *older adults* were divided into three different categories (mental disorder, disease of the nervous system, and symptoms not classified elsewhere). Although design papers often considered lifestyle management and older adults, this insight was invisible to those looking at the graph only.

A second limitation is that we aimed to include a wide variety of digital health technologies, although our results mostly identified digital health solutions in the domain of support platforms via apps or the internet. In addition, almost all solutions were designed for use at home. This is remarkable, as our search strategy did not focus solely on the home. Obviously, the terms *telemedicine* and *telehealth* refer to technologies for use at home. Nonetheless, we also included the term *digital* in our search strategy, which we expected would result in papers on digital technologies used in hospitals and primary care settings. Our search strategy had missing terms that would result in papers on technology use in hospitals, technology use in hospitals is not yet common, or it is not discussed in relation to well-being. Although we believe that our insights can also be applied to a wider range of digital health technologies in varying contexts, in the future, it is interesting to expand our research to varying technologies and extend it to the inpatient care context to understand the differences in designing technologies in diverging contexts. Furthermore, as we only included papers focused on well-being improvement, we might have missed the results on in-person value conflicts that are reported in papers studying the harmful effects of digital health on well-being. Finally, we rarely found sets of papers that focused on the design and evaluation of the same technology. The ability to compare the design and evaluation processes of similar technologies would have led to more reliable results. In the future, better aligning design and evaluation would ideally also result in more consistent reporting.

Conclusions

In this review, we have shown how current practices deal with the major problems that are paramount in digital health for

well-being. We identified major gaps in design and evaluation regarding their conceptualization of well-being, types of users studied, and methods used to design and assess well-being. The comparison of empirical practice with theoretical frameworks also showed how both fields have found pragmatic ways of circumventing or dealing with the problems of digital health for well-being. By illuminating the differences between design and evaluation, as well as practice and theory, and providing recommendations for best practice, we expect to have set the first steps to slightly bridge some gaps. As digital health technologies are gaining an increasingly important role in the future, we believe that multidisciplinary collaborations are required to be improved by moving beyond the common dichotomy of design and evaluation. Only then it is possible to transcend from digital health toward digital well-being.

Acknowledgments

This work was supported by the *Europees Fonds voor Regionale Ontwikkeling* for the project R4heal (PROJ-00840, 2018).

Conflicts of Interest

CMK is employed by the University of Twente through the fund, Top Technology Twente Connecting Industry program (TKI Topsector HTSM), which is partially funded (paid to institution) by Philips.

Multimedia Appendix 1
Search strategy.
[DOC File , 43 KB - jmir_v24i4e33787_app1.doc]

Multimedia Appendix 2
Detailed results.
[DOC File , 152 KB - jmir_v24i4e33787_app2.doc]

References

1. Kreps GL, Neuhauser L. New directions in eHealth communication: opportunities and challenges. Patient Educ Couns 2010 Mar;78(3):329-336. [doi: 10.1016/j.pec.2010.01.013] [Medline: 20202779]
2. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res 2005 Feb;7(1):e1 [FREE Full text] [doi: 10.2196/jmir.7.1.e1] [Medline: 15829471]
3. Nacinovich M. Defining mHealth. J Commun Healthc 2013 Jul 18;4(1):1-3 [FREE Full text] [doi: 10.1179/175380611X12950033990296]
4. Dorsey ER, Topol EJ. State of Telehealth. N Engl J Med 2016 Jul 14;375(2):154-161. [doi: 10.1056/NEJMra1601705] [Medline: 27410924]
5. Sood S, Mbarika V, Jugoo S, Dookhy R, Doarn CR, Prakash N, et al. What is telemedicine? A collection of 104 peer-reviewed perspectives and theoretical underpinnings. Telemed J E Health 2007 Oct;13(5):573-590. [doi: 10.1089/tmj.2006.0073] [Medline: 17999619]
6. Smits M, Staal JB, Van Goor H. Could Virtual Reality play a role in the rehabilitation after COVID-19 infection? BMJ Open Sport Exerc Med 2020 Oct 23;6(1):e000943 [FREE Full text] [doi: 10.1136/bmjsem-2020-000943] [Medline: 33178449]
7. Torous J, Wykes T. Opportunities from the coronavirus disease 2019 pandemic for transforming psychiatric care with telehealth. JAMA Psychiatry 2020 Dec 01;77(12):1205-1206. [doi: 10.1001/jamapsychiatry.2020.1640] [Medline: 32391857]
8. Quinn L, Macpherson C, Long K, Shah H. Promoting physical activity via telehealth in people with Parkinson disease: the path forward after the COVID-19 pandemic? Physical therapy. Phys Ther 2020 Sep 28;100(10):1730-1736. [doi: 10.1093/ptj/pzaa128]
9. 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]
10. Brey P. Design for the value of human wellbeing. In: Van Den Hoven MJ, Vermaas PE, Van de Poel I, editors. Handbook of Ethics, Values, and Technological Design - Sources, Theory, Values and Application Domains. Dordrecht, The Netherlands: Springer; 2015:365-382.
11. Friedman B, Kahn P, Borning A. Value sensitive design and information systems. In: Zhang P, Galletta D, editors. Human-Computer Interaction in Management Information Systems: Foundations. New York, United States of America: M. E. Sharpe; 2006:348-372.
12. World Health Organization. Preamble to the Constitution of World Health Organization as Adopted by the International Health Conference. Geneva: World Health Organization; 1946.
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 Sep 04;169(7):467. [doi: 10.7326/M18-0850]

14. 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]
15. Clarke V, Braun V, Hayfield N, Terry G. Thematic analysis. In: Liamputtong P, editor. Handbook of Research Methods in Health Social Sciences. Berlin, Germany: Springer; 2019:843-860.
16. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Br Med J 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
17. Abdullah S. Towards circadian computing: a sensing and intervention framework for BodyClock friendly technology. In: Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing.: Association for Computing Machinery; 2015 Presented at: ACM International Joint Conference on Pervasive and Ubiquitous Computing; 2015; Osaka, Japan p. 515-520. [doi: 10.1145/2800835.2801657]
18. Bentley F, Tollmar K. The power of mobile notifications to increase wellbeing logging behavior. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.: Association for Computing Machinery; 2013 Presented at: SIGCHI Conference on Human Factors in Computing Systems; 2013; Paris, France p. 1095-1098. [doi: 10.1145/2470654.2466140]
19. Berg M, Adolfsson A, Ranerup A, Sparud-Lundin C, University of Gothenburg Centre for Person-Centred Care. Person-centered web support to women with type 1 diabetes in pregnancy and early motherhood--the development process. Diabetes Technol Ther 2013 Jan;15(1):20-25. [doi: 10.1089/dia.2012.0217] [Medline: 23297670]
20. Cahill J, McLoughlin S, O'Connor M, Stolberg M, Wetherall S. Addressing issues of need, adaptability, user acceptability and ethics in the participatory design of new technology enabling wellness, independence and dignity for seniors living in residential homes. In: Proceedings of the Third International Conference on Human Aspects of IT for the Aged Population.: Springer; 2017 Presented at: Third International Conference on Human Aspects of IT for the Aged Population; 2017; Vancouver, Canada. [doi: 10.1007/978-3-319-58530-7_7]
21. Darragh M, Ahn HS, MacDonald B, Liang A, Peri K, Kerse N, et al. Homecare robots to improve health and well-being in mild cognitive impairment and early stage dementia: results from a scoping study. J Am Med Dir Assoc 2017 Dec 01;18(12):1099.e1-1099.e4. [doi: 10.1016/j.jamda.2017.08.019] [Medline: 28974463]
22. De Croon R, Bezabih AM, Geuens J, Wildemeersch D, Oeyen D, Verbert K, et al. Motivational design techniques to increase adherence to a telemonitoring therapy a study with adolescent pectus patients. In: Proceedings of the IEEE International Conference on Healthcare Informatics. 2019 Presented at: IEEE International Conference on Healthcare Informatics; 2019; Xian, China. [doi: 10.1109/ichi.2019.8904661]
23. Faezipour M, Faezipour M. System dynamics modeling for smartphone-based healthcare tools: case study on ECG monitoring. IEEE Syst J 2021 Jun;15(2):3036-3045. [doi: 10.1109/jsyst.2020.3009187]
24. Goodyear-Smith F, Warren J, Elley CR. The eCHAT program to facilitate healthy changes in New Zealand primary care. J Am Board Fam Med 2013 Mar;26(2):177-182 [FREE Full text] [doi: 10.3122/jabfm.2013.02.120221] [Medline: 23471931]
25. Meinert E, Milne-Ives M, Surodina S, Lam C. Agile requirements engineering and software planning for a digital health platform to engage the effects of isolation caused by social distancing: case study. JMIR Public Health Surveill 2020 May 06;6(2):e19297 [FREE Full text] [doi: 10.2196/19297] [Medline: 32348293]
26. Reis J, McGinty B, Jones S. An e-learning caregiving program for prostate cancer patients and family members. J Med Syst 2003;27(1):1-12. [doi: 10.1023/a:1021040911088]
27. Devlin AM, McGee-Lennon M, O'Donnell CA, Bouamrane M, Agbakoba R, O'Connor S, et al. Delivering digital health and well-being at scale: lessons learned during the implementation of the dallas program in the United Kingdom. J Am Med Inform Assoc 2016 Jan 08;23(1):48-59 [FREE Full text] [doi: 10.1093/jamia/ocv097] [Medline: 26254480]
28. Davies A, Mueller J, Hennings J, Caress A, Jay C. Recommendations for developing support tools with people suffering from chronic obstructive pulmonary disease: co-design and pilot testing of a mobile health prototype. JMIR Hum Factors 2020 May 15;7(2):e16289 [FREE Full text] [doi: 10.2196/16289] [Medline: 32410730]
29. Alelaiwi A, Hassan M, Bhuiyan Z. A secure and dependable connected smart home system for elderly. In: Proceedings of the IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech).: IEEE; 2017 Presented at: IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech); 2017; Orlando, FL, United States of America p. 722-727. [doi: 10.1109/dasc-picom-datacom-cyberscitec.2017.126]
30. Pinto M, Marques P. OneCare Spiro: mobile application for monitoring and pre-diagnosis of chronic obstructive pulmonary disease. In: 5th Portuguese Meeting on Bioengineering (ENBENG).: IEEE; 2017 Presented at: 5th Portuguese Meeting on Bioengineering (ENBENG); 2017; Coimbra, Portugal. [doi: 10.1109/enbeng.2017.7889481]
31. Rodriguez I, Herskovic V, Fuentes C, Campos M. B-ePain: a wearable interface to self-report pain and emotions. In: Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2016 Presented at: ACM International Joint Conference on Pervasive and Ubiquitous Computing; 2016; Heidelberg, Germany p. 1120-1125. [doi: 10.1145/2968219.2972719]

32. Cai RA, Beste D, Chaplin H, Varakliotis S, Suffield L, Josephs F, et al. Developing and evaluating JIApp: acceptability and usability of a smartphone app system to improve self-management in young people with juvenile idiopathic arthritis. JMIR Mhealth Uhealth 2017 Aug 15;5(8):e121 [FREE Full text] [doi: 10.2196/mhealth.7229] [Medline: 28811270]
33. Fylan F, Caveney L, Cartwright A, Fylan B. Making it work for me: beliefs about making a personal health record relevant and useable. BMC Health Serv Res 2018 Jun 14;18(1):445 [FREE Full text] [doi: 10.1186/s12913-018-3254-z] [Medline: 29898730]
34. Schoenthaler A, Cruz J, Payano L, Rosado M, Labbe K, Johnson C, et al. Investigation of a mobile health texting tool for embedding patient-reported data into diabetes management (i-Matter): development and usability study. JMIR Form Res 2020 Aug 31;4(8):e18554 [FREE Full text] [doi: 10.2196/18554] [Medline: 32865505]
35. Vos J, Gerling K, Linehan C, Siriwardena AN, Windle K. Understanding care navigation by older adults with multimorbidity: mixed-methods study using social network and framework analyses. JMIR Aging 2018 Nov 14;1(2):e11054 [FREE Full text] [doi: 10.2196/11054] [Medline: 31518236]
36. Ragouzeos D, Gandrup J, Berrean B, Li J, Murphy M, Trupin L, et al. "Am I OK?" using human centered design to empower rheumatoid arthritis patients through patient reported outcomes. Patient Educ Couns 2019 Mar;102(3):503-510 [FREE Full text] [doi: 10.1016/j.pec.2018.10.016] [Medline: 30446358]
37. Ashmore L, Stewart H, Hutton D, Evans K. Digital support for living with and beyond gynaecological cancer. Radiography (Lond) 2020 Nov;26(4):270-276. [doi: 10.1016/j.radi.2020.03.014] [Medline: 32439292]
38. Ando H, Ashcroft-Kelso H, Halhead R, Chakrabarti B, Young CA, Cousins R, et al. Experience of telehealth in people with motor neurone disease using noninvasive ventilation. Disabil Rehabil Assist Technol 2019 Sep 12:1-7. [doi: 10.1080/17483107.2019.1659864] [Medline: 31512539]
39. Hawley CE, Genovese N, Owsiany MT, Triantafylidis LK, Moo LR, Linsky AM, et al. Rapid integration of home telehealth visits amidst COVID-19: what do older adults need to succeed? J Am Geriatr Soc 2020 Nov 04;68(11):2431-2439 [FREE Full text] [doi: 10.1111/jgs.16845] [Medline: 32930391]
40. Lau AY, Dunn AG, Mortimer N, Gallagher A, Proudfoot J, Andrews A, et al. Social and self-reflective use of a Web-based personally controlled health management system. J Med Internet Res 2013;15(9):e211 [FREE Full text] [doi: 10.2196/jmir.2682] [Medline: 24064188]
41. Taft C, Hallberg I, Bengtsson U, Manhem K, Kjellgren K. Links between blood pressure and medication intake, well-being, stress, physical activity and symptoms reported via a mobile phone-based self-management support system: a cohort study in primary care. BMJ Open 2018 Aug 23;8(8):e020849 [FREE Full text] [doi: 10.1136/bmjopen-2017-020849] [Medline: 30139897]
42. Papa A, Mital M, Pisano P, Del Giudice M. E-health and wellbeing monitoring using smart healthcare devices: an empirical investigation. Technol Forecast Soc Chang 2020 Apr;153:119226. [doi: 10.1016/j.techfore.2018.02.018]
43. Spencer R, Alexander V, Eickhoff J, Woo K, Costanzo E, Marx N, et al. A digital media attention diversion improves mood and fear in patients receiving chemotherapy for recurrent gynecologic malignancies: results of a randomized trial. Int J Gynecol Cancer 2020 Apr 01;30(4):525-532 [FREE Full text] [doi: 10.1136/ijgc-2019-001185] [Medline: 32122951]
44. Willems RA, Lechner L, Verboon P, Mesters I, Kanera IM, Bolman CA. Working mechanisms of a web-based self-management intervention for cancer survivors: a randomised controlled trial. Psychol Health 2017 May 21;32(5):605-625. [doi: 10.1080/08870446.2017.1293054] [Medline: 28276741]
45. Hirani SP, Beynon M, Cartwright M, Rixon L, Doll H, Henderson C, et al. The effect of telecare on the quality of life and psychological well-being of elderly recipients of social care over a 12-month period: the Whole Systems Demonstrator cluster randomised trial. Age Ageing 2014 May;43(3):334-341. [doi: 10.1093/ageing/aft185] [Medline: 24333802]
46. Børøsund E, Varsi C, Clark MM, Ehlers SL, Andrykowski MA, Sleveland HR, et al. Pilot testing an app-based stress management intervention for cancer survivors. Transl Behav Med 2019 Apr 24;10(3):770-780. [doi: 10.1093/tbm/ibz062] [Medline: 31330023]
47. Brough C, Boyce S, Houchen-Wolloff L, Sewell L, Singh S. Evaluating the interactive web-based program, activate your heart, for cardiac rehabilitation patients: a pilot study. J Med Internet Res 2014;16(10):e242 [FREE Full text] [doi: 10.2196/jmir.3027] [Medline: 25359204]
48. Mercier HW, Hamner JW, Torous J, Onnela JP, Taylor JA. Digital phenotyping to quantify psychosocial well-being trajectories after spinal cord injury. Am J Phys Med Rehabil 2020 Jun 23;99(12):1138-1144. [doi: 10.1097/phm.0000000000001506]
49. Sherman DW, Haber J, Hoskins CN, Budin WC, Maislin G, Shukla S, et al. The effects of psychoeducation and telephone counseling on the adjustment of women with early-stage breast cancer. Appl Nurs Res 2012 Feb;25(1):3-16. [doi: 10.1016/j.apnr.2009.10.003] [Medline: 20974079]
50. Cikajlo I, Hukic A, Dolinsek I, Zajc D, Vesel M, Krizmanic T, et al. Can telerehabilitation games lead to functional improvement of upper extremities in individuals with Parkinson's disease? Int J Rehabil Res 2018;41(3):230-238. [doi: 10.1097/mrr.0000000000000291]
51. 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 Sep 27;18(9):e255 [FREE Full text] [doi: 10.2196/jmir.5800] [Medline: 27678169]

52. Mitchell J, Stanimirovic R, Klein B, Vella-Brodrick D. A randomised controlled trial of a self-guided internet intervention promoting well-being. Comput Hum Behav 2009 May;25(3):749-760. [doi: 10.1016/j.chb.2009.02.003]
53. Phillips VL, Vesmarovich S, Hauber R, Wiggers E, Egner A. Telehealth: reaching out to newly injured spinal cord patients. Public Health Rep 2001 Nov;116 Suppl 1:94-102 [FREE Full text] [doi: 10.1093/phr/116.S1.94] [Medline: 11889278]
54. Kyriazakos S, Valentini V, Cesario A, Zachariae R. FORECAST - A cloud-based personalized intelligent virtual coaching platform for the well-being of cancer patients. Clin Transl Radiat Oncol 2018 Jan;8:50-59 [FREE Full text] [doi: 10.1016/j.ctro.2017.11.006] [Medline: 29594242]
55. Coulter EH, McLean AN, Hasler JP, Allan DB, McFadyen A, Paul L. The effectiveness and satisfaction of web-based physiotherapy in people with spinal cord injury: a pilot randomised controlled trial. Spinal Cord 2017 Apr;55(4):383-389. [doi: 10.1038/sc.2016.125] [Medline: 27596027]
56. Possin KL, Merrilees JJ, Dulaney S, Bonasera SJ, Chiong W, Lee K, et al. Effect of collaborative dementia care via telephone and internet on quality of life, caregiver well-being, and health care use: the care ecosystem randomized clinical trial. JAMA Intern Med 2019 Dec 01;179(12):1658-1667 [FREE Full text] [doi: 10.1001/jamainternmed.2019.4101] [Medline: 31566651]
57. Cammisuli D, Pietrabissa G, Castelnuovo G. Improving wellbeing of community-dwelling people with mild cognitive impairment: the SENIOR (SystEm of Nudge theory based ICT applications for OldeR citizens) project. Neural Regen Res 2021 May;16(5):963-966 [FREE Full text] [doi: 10.4103/1673-5374.297063] [Medline: 33229736]
58. Klein S, Cumming G, Lee A, Alexander D, Bolsover D. Evaluating the effectiveness of a web-based intervention to promote mental wellbeing in women and partners following miscarriage, using a modified patient preference trial design: an external pilot. Br J Obstet Gynaecol 2012;119(6):762-767. [doi: 10.1111/j.1471-0528.2012.03302.x]
59. Jaensson M, Dahlberg K, Eriksson M, Nilsson U. Evaluation of postoperative recovery in day surgery patients using a mobile phone application: a multicentre randomized trial. Br J Anaesth 2017 Nov 01;119(5):1030-1038 [FREE Full text] [doi: 10.1093/bja/aex331] [Medline: 29077818]
60. Yanez B, Oswald LB, Baik SH, Buitrago D, Iacobelli F, Perez-Tamayo A, et al. Brief culturally informed smartphone interventions decrease breast cancer symptom burden among Latina breast cancer survivors. Psychooncology 2020 Jan 15;29(1):195-203 [FREE Full text] [doi: 10.1002/pon.5281] [Medline: 31693265]
61. Zhou K, Wang W, Zhao W, Li L, Zhang M, Guo P, et al. Benefits of a WeChat-based multimodal nursing program on early rehabilitation in postoperative women with breast cancer: a clinical randomized controlled trial. Int J Nurs Stud 2020 Jun;106:103565. [doi: 10.1016/j.ijnurstu.2020.103565] [Medline: 32272281]
62. Ryhänen AM, Rankinen S, Siekkinen M, Saarinen M, Korvenranta H, Leino-Kilpi H. The impact of an empowering internet-based breast cancer patient pathway program on breast cancer patients' clinical outcomes: a randomised controlled trial. J Clin Nurs 2013 Apr 11;22(7-8):1016-1025. [doi: 10.1111/jocn.12007] [Medline: 23480498]
63. Rehman H, Karpman C, Douglas KV, Benzo RP. Effect of a motivational interviewing-based health coaching on quality of life in subjects with COPD. Respir Care 2017 Aug 13;62(8):1043-1048 [FREE Full text] [doi: 10.4187/respcare.04984] [Medline: 28611230]
64. Nguyen BJ, Kim Y, Park K, Chen AJ, Chen S, Van Fossan D, et al. Improvement in patient-reported quality of life outcomes in severely visually impaired individuals using the aira assistive technology system. Transl Vis Sci Technol 2018 Sep 04;7(5):30 [FREE Full text] [doi: 10.1167/tvst.7.5.30] [Medline: 30386682]
65. Park K, Kim Y, Nguyen BJ, Chen A, Chao DL. Quality of life assessment of severely visually impaired individuals using aira assistive technology system. Transl Vis Sci Technol 2020 Mar 09;9(4):21 [FREE Full text] [doi: 10.1167/tvst.9.4.21] [Medline: 32818108]
66. Cassidy K, Vallis M, Conn D, Massie AS, Checkland C, Parsons D, et al. Comparing a clinician-assisted and app-supported behavioral activation intervention to promote brain health and well-being in frontline care. Int Psychogeriatr 2020 Aug 24;33(6):615-625. [doi: 10.1017/s1041610220001325]
67. Hauser-Ulrich S, Künzli H, Meier-Peterhans D, Kowatsch T. A smartphone-based health care chatbot to promote self-management of chronic pain (SELMA): pilot randomized controlled trial. JMIR Mhealth Uhealth 2020 Apr 03;8(4):e15806 [FREE Full text] [doi: 10.2196/15806] [Medline: 32242820]
68. Dogu OZ, Kaya H. Compliance of the web-based distance training and consultancy on individual's treatment having suffered myocardial infarction and its effects on well-being. J Coll Physicians Surg Pak 2018 Dec 01;28(12):953-959. [doi: 10.29271/jcpsp.2018.12.953] [Medline: 30501835]
69. Kayrouz R, Dear BF, Karin E, Fogliati VJ, Titov N. A pilot study of a clinician-guided internet-delivered cognitive behavioural therapy for anxiety and depression among Arabs in Australia, presented in both English and Arabic languages. Internet Interv 2016 Sep;5:5-11. [doi: 10.1016/j.invent.2016.06.002] [Medline: 30135801]
70. Villani D, Cognetta C, Repetto C, Serino S, Toniolo D, Scanzi F, et al. Promoting emotional well-being in older breast cancer patients: results from an eHealth intervention. Front Psychol 2018 Nov 27;9:2279 [FREE Full text] [doi: 10.3389/fpsyg.2018.02279] [Medline: 30538650]
71. Hauffman A, Alfonsson S, Mattsson S, Forslund M, Bill-Axelson A, Nygren P, et al. The development of a nurse-led internet-based learning and self-care program for cancer patients with symptoms of anxiety and depression—a part of U-CARE. Cancer Nurs 2017;40(5):9-16. [doi: 10.1097/ncc.0000000000000402]

72. Mann S, Pan Z, Tao Y, Gao A, Tao X, Garcia DE, et al. Face recognition and rehabilitation: a wearable assistive and training system for prosopagnosia. In: Proceedings of the International Conference on Systems, Man, and Cybernetics (SMC).: IEEE; 2020 Presented at: International Conference on Systems, Man, and Cybernetics (SMC); 2020; Toronto, Canada. [doi: 10.1109/smc42975.2020.9283058]
73. Matthews M, Abdullah S, Gay G, Choudhury T. Tracking mental well-being: balancing rich sensing and patient needs. Computer 2014;47(4):36-43. [doi: 10.1109/MC.2014.107]
74. Doherty K, Barry M, Belisario JM, Morrison C, Car J, Doherty G. Personal information and public health: design tensions in sharing and monitoring wellbeing in pregnancy. Int J Hum Comput Stud 2020 Mar;135:102373 [FREE Full text] [doi: 10.1016/j.ijhcs.2019.102373] [Medline: 32127731]
75. Doherty K, Barry M, Marcano-Belisario J, Arnaud B, Morrison C, Car J, et al. A mobile app for the self-report of psychological well-being during pregnancy (BrightSelf): qualitative design study. JMIR Ment Health 2018 Nov 27;5(4):e10007 [FREE Full text] [doi: 10.2196/10007] [Medline: 30482742]
76. O'Connor S. Co-designing technology with people with dementia and their carers: exploring user perspectives when co-creating a mobile health application. Int J Older People Nurs 2020 Sep 14;15(3):e12288. [doi: 10.1111/opn.12288] [Medline: 31837096]
77. Ospina-Pinillos L, Davenport T, Diaz AM, Navarro-Mancilla A, Scott EM, Hickie IB. Using participatory design methodologies to co-design and culturally adapt the Spanish version of the mental health eClinic: qualitative study. J Med Internet Res 2019 Aug 02;21(8):e14127. [doi: 10.2196/14127]
78. Ospina-Pinillos L, Davenport TA, Navarro-Mancilla AA, Cheng VW, Cardozo Alarcón AC, Rangel AM, et al. Involving end users in adapting a Spanish version of a web-based mental health clinic for young people in Colombia: exploratory study using participatory design methodologies. JMIR Ment Health 2020 Feb 08;7(2):e15914 [FREE Full text] [doi: 10.2196/15914] [Medline: 32027313]
79. Mehrotra A, Musolesi M. Designing effective movement digital biomarkers for unobtrusive emotional state mobile monitoring. In: Proceedings of the 1st Workshop on Digital Biomarkers. 2017 Presented at: Digital BioMarkers; 2017; Niagara falls, NY, United States of America p. 3-8. [doi: 10.1145/3089341.3089342]
80. Chen Y, Chen Y, Zheng K, Dodakian L, See J, Zhou R, et al. A qualitative study on user acceptance of a home-based stroke telerehabilitation system. Top Stroke Rehabil 2020 Mar 04;27(2):81-92 [FREE Full text] [doi: 10.1080/10749357.2019.1683792] [Medline: 31682789]
81. Budrionis A, Wynn R, Marco-Ruiz L, Yigzaw KY, Bergvik S, Oyeyemi SO, et al. Impact of the use of electronic health tools on the psychological and emotional well-being of electronic health service users (The Seventh Tromsø Study - Part 3): population-based questionnaire study. J Med Internet Res 2020 Mar 05;22(3):e13118 [FREE Full text] [doi: 10.2196/13118] [Medline: 32134396]
82. Epstein YM, Rosenberg HS, Grant TV, Hemenway BA. Use of the internet as the only outlet for talking about infertility. Fertil Steril 2002 Sep;78(3):507-514. [doi: 10.1016/s0015-0282(02)03270-3]
83. Kabir M, Rice JL, Bush SH, Lawlor PG, Webber C, Grassau PA, et al. A mixed-methods pilot study of 'LIFEView' audiovisual technology: virtual travel to support well-being and quality of life in palliative and end-of-life care patients. Palliat Med 2020 Jul 08;34(7):954-965 [FREE Full text] [doi: 10.1177/0269216320918514] [Medline: 32383629]
84. Välimäki M, Mishina K, Kaakinen JK, Holm SK, Vahlo J, Kirjonen M, et al. Digital gaming for improving the functioning of people with traumatic brain injury: randomized clinical feasibility study. J Med Internet Res 2018 Mar 19;20(3):e77 [FREE Full text] [doi: 10.2196/jmir.7618] [Medline: 29555622]
85. Chan JC, Sui Y, Oldenburg B, Zhang Y, Chung HH, Goggins W, JADEPEARL Project Team. Effects of telephone-based peer support in patients with type 2 diabetes mellitus receiving integrated care: a randomized clinical trial. JAMA Intern Med 2014 Jun 01;174(6):972-981. [doi: 10.1001/jamainternmed.2014.655] [Medline: 24781960]
86. Malins S, Biswas S, Patel S, Levene J, Moghaddam N, Morriss R. Preventing relapse with personalized smart-messaging after cognitive behavioural therapy: a proof-of-concept evaluation. Br J Clin Psychol 2020 Jun 20;59(2):241-259 [FREE Full text] [doi: 10.1111/bjc.12244] [Medline: 31960467]
87. Mehta S, Hadjistavropoulos H, Nugent M, Karin E, Titov N, Dear BF. Guided internet-delivered cognitive-behaviour therapy for persons with spinal cord injury: a feasibility trial. Spinal Cord 2020 May 09;58(5):544-552. [doi: 10.1038/s41393-019-0398-6] [Medline: 31919447]
88. Espinoza M, Baños R, García-Palacios A, Cervera JM, Esquerdo G, Barrajón E, et al. Promotion of emotional wellbeing in oncology inpatients using VR. Stud Health Technol Inform 2012;181:53-57. [doi: 10.3233/978-1-61499-121-2-53]
89. Longobardi Y, Galli J, D'Alatri L, Savoia V, Mari G, Rigante M, et al. Patients with voice prosthesis rehabilitation during the covid-19 pandemic: analyzing the effectiveness of remote triage and management. Otolaryngol Head Neck Surg 2021 Feb 04;164(2):277-284 [FREE Full text] [doi: 10.1177/0194599820948043] [Medline: 32746738]
90. Sengpiel J, Fuehner T, Kugler C, Avsar M, Bodmann I, Boemke A, et al. Use of telehealth technology for home spirometry after lung transplantation: a randomized controlled trial. Prog Transpl 2010 Dec 01;20(4):310-317. [doi: 10.1177/152692481002000402]

91. Strøm J, Høybye MT, Laursen M, Jørgensen LB, Nielsen CV. Lumbar spine fusion patients' use of an internet support group: mixed methods study. J Med Internet Res 2019 Jul 04;21(7):e9805 [FREE Full text] [doi: 10.2196/jmir.9805] [Medline: 31274113]
92. Taylor A, Godden D, Aitken A, Colligan J, Wilson R. Delivering group-based services to the home via the Internet: Maximising clinical and social benefits. In: Proceedings of the Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth).: IEEE; 2011 Presented at: Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth); 2011; Dublin, Ireland. [doi: 10.4108/icst.pervasivehealth.2011.246004]
93. Quinn CC, Swasey KK, Crabbe JC, Shardell MD, Terrin ML, Barr EA, et al. The impact of a mobile diabetes health intervention on diabetes distress and depression among adults: secondary analysis of a cluster randomized controlled trial. JMIR Mhealth Uhealth 2017 Dec 07;5(12):e183 [FREE Full text] [doi: 10.2196/mhealth.8910] [Medline: 29217502]
94. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (WYSA) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR Mhealth Uhealth 2018 Nov 23;6(11):e12106 [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]
95. Espie CA, Emsley R, Kyle SD, Gordon C, Drake CL, Siriwardena AN, et al. Effect of digital cognitive behavioral therapy for insomnia on health, psychological well-being, and sleep-related quality of life: a randomized clinical trial. JAMA Psychiatry 2019 Jan 01;76(1):21-30 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.2745] [Medline: 30264137]
96. Hemmings NR, Kawadler JM, Whatmough R, Ponzo S, Rossi A, Morelli D, et al. Development and feasibility of a digital acceptance and commitment therapy-based intervention for generalized anxiety disorder: pilot acceptability study. JMIR Form Res 2021 Feb 09;5(2):e21737 [FREE Full text] [doi: 10.2196/21737] [Medline: 33560232]
97. Castro SC, Chiguluri V, Gumpina R, Abbott P, Madero EN, Payne M, et al. Outcomes of a digital health program with human coaching for diabetes risk reduction in a medicare population. J Aging Health 2017 Jan 01:898264316688791. [doi: 10.1177/0898264316688791] [Medline: 28553807]
98. Döbler A, Belnap B, Pollmann H, Farin E, Raspe H, Mittag O. Telephone-delivered lifestyle support with action planning and motivational interviewing techniques to improve rehabilitation outcomes. Rehabil Psychol 2018 May;63(2):170-181. [doi: 10.1037/rep0000224] [Medline: 29878825]
99. Zhang X, Xiao H, Chen Y. Evaluation of a WeChat-based life review programme for cancer patients: a quasi-experimental study. J Adv Nurs 2019 Jul 16;75(7):1563-1574. [doi: 10.1111/jan.14018] [Medline: 30950535]
100. Bianchi C. Exploring how internet services can enhance elderly well-being. J Serv Mark 2021 Feb 01;35(5):585-603. [doi: 10.1108/jsm-05-2020-0177]
101. Cahill J, McLoughlin S, Wetherall S. The design of new technology supporting wellbeing, independence and social participation, for older adults domiciled in residential homes and/or assisted living communities. Technologies 2018 Jan 26;6(1):18. [doi: 10.3390/technologies6010018]
102. Rickard N, Arjmand H, Bakker D, Seabrook E. Development of a mobile phone app to support self-monitoring of emotional well-being: a mental health digital innovation. JMIR Ment Health 2016 Nov 23;3(4):e49 [FREE Full text] [doi: 10.2196/mental.6202] [Medline: 27881358]
103. Siegmund D, Chiesa L, Hörr O, Gabler F, Braun A, Kuijper A. Talis - A design study for a wearable device to assist people with depression. In: Proceedings of the 41st Annual Computer Software and Applications Conference (COMPSAC).: IEEE; 2017 Presented at: 41st Annual Computer Software and Applications Conference (COMPSAC); 2017; Turin, Italy. [doi: 10.1109/compsac.2017.228]
104. Zulman DM, Schafenacker A, Barr KL, Moore IT, Fisher J, McCurdy K, et al. Adapting an in-person patient-caregiver communication intervention to a tailored web-based format. Psychooncology 2012 Mar;21(3):336-341 [FREE Full text] [doi: 10.1002/pon.1900] [Medline: 21830255]
105. Alvarado-Martel D, Cañas F, Velasco R, Alcubierre N, López-Ríos L, Rius F, et al. Design, construction, and implementation of an online platform for patients with type 1 diabetes: EncoDiab. Patient Prefer Adher 2015 Jun:767. [doi: 10.2147/ppa.s77730]
106. Williams V, Price J, Hardinge M, Tarassenko L, Farmer A. Using a mobile health application to support self-management in COPD: a qualitative study. Br J Gen Pract 2014 Jul;64(624):392-400 [FREE Full text] [doi: 10.3399/bjgp14X680473] [Medline: 24982491]
107. Nissen L, Lindhardt T. A qualitative study of COPD-patients' experience of a telemedicine intervention. Int J Med Inform 2017 Nov;107:11-17. [doi: 10.1016/j.ijmedinf.2017.08.004] [Medline: 29029687]
108. Coorey G, Peiris D, Usherwood T, Neubeck L, Mulley J, Redfern J. Persuasive design features within a consumer-focused eHealth intervention integrated with the electronic health record: a mixed methods study of effectiveness and acceptability. PLoS One 2019 Jun 20;14(6):e0218447 [FREE Full text] [doi: 10.1371/journal.pone.0218447] [Medline: 31220127]
109. Larson LR, Bock DE. Consumer search and satisfaction with mental health services. J Serv Mark 2016 Oct 10;30(7):736-748. [doi: 10.1108/jsm-09-2015-0281]
110. Guo Q, Cann B, McClement S, Thompson G, Chochinov HM. Keep in Touch (KIT): feasibility of using internet-based communication and information technology in palliative care. BMC Palliat Care 2017 May 06;16(1):29 [FREE Full text] [doi: 10.1186/s12904-017-0203-2] [Medline: 28477630]

111. Camberg L, Woods P, Ooi W, Hurley A, Volicer L, Ashley J, et al. Evaluation of Simulated Presence: a personalized approach to enhance wellbeing in persons with Alzheimer's disease. J Am Geriatr Soc 1999;47(4):446-452. [doi: 10.1111/j.1532-5415.1999.tb07237.x]
112. Weymann N, Dirmaier J, von Wolff A, Kriston L, Härter M. Effectiveness of a web-based tailored interactive health communication application for patients with type 2 diabetes or chronic low back pain: randomized controlled trial. J Med Internet Res 2015;17(3):e53 [FREE Full text] [doi: 10.2196/jmir.3904] [Medline: 25736340]
113. Gazit T, Amichai-Hamburger Y. Factors underlying engagement in Facebook support groups of female infertility patients. Psychol Rep 2021 Jun 27;124(3):1150-1173. [doi: 10.1177/0033294120934703] [Medline: 32597374]
114. Anderson LM, Chacko TP. Bariatric outcomes: self-management for sustained surgical success. J Addict Nurs 2018;29(1):32-42. [doi: 10.1097/jan.0000000000000210]
115. Cantor G. Designing technological interventions for patients with discordant chronic comorbiditiestype-2 diabetes. In: Proceedings of the CHI Conference on Human Factors in Computing Systems.: Association for Computing Machinery; 2018 Presented at: CHI Conference on Human Factors in Computing Systems; 2018; Montreal QC, Canada p. 1-6. [doi: 10.1145/3170427.3180304]
116. Cheek C, Fleming T, Lucassen MF, Bridgman H, Stasiak K, Shepherd M, et al. Integrating health behavior theory and design elements in serious games. JMIR Ment Health 2015 Apr 21;2(2):e11 [FREE Full text] [doi: 10.2196/mental.4133] [Medline: 26543916]
117. Derboven J, Voorend R, Slegers K. Design trade-offs in self-management technology: the HeartMan case. Behav Inform Technol 2019 Jul 03;39(1):72-87. [doi: 10.1080/0144929x.2019.1634152]
118. Patrício L, de Pinho NF, Teixeira JG, Fisk RP. Service design for value networks: enabling value cocreation interactions in healthcare. Serv Sci 2018 Mar;10(1):76-97. [doi: 10.1287/serv.2017.0201]
119. Rodrigues K, Conrado D, Neris V. Lessons learned in designing a digital therapeutic game to support the treatment and wellbeing of children with cancer. In: Proceedings of the International Conference of Design, User Experience, and Usability.: Springer; 2018 Presented at: International Conference of Design, User Experience, and Usability; 2018; Las Vegas, United States of America. [doi: 10.1007/978-3-319-91806-8_18]
120. Albright G, Adam C, Serri D, Bleeker S, Goldman R. Harnessing the power of conversations with virtual humans to change health behaviors. Mhealth 2016 Nov 28;2:44 [FREE Full text] [doi: 10.21037/mhealth.2016.11.02] [Medline: 28293614]
121. Simonÿ C, Andersen IC, Bodtger U, Nyberg M, Birkelund R. Raised illness mastering - a phenomenological hermeneutic study of chronic obstructive pulmonary disease patients' experiences while participating in a long-term telerehabilitation programme. Disabil Rehabil Assist Technol 2020 Aug 26:1-8. [doi: 10.1080/17483107.2020.1804630] [Medline: 32845801]
122. Högberg KM, Sandman L, Nyström M, Stockelberg D, Broström A. Caring through web-based communication: a qualitative evaluation of a nursing intervention to create holistic well-being among patients with hematological disease. J Holist Nurs 2018 Sep 22;36(3):218-227. [doi: 10.1177/0898010116667343] [Medline: 27659277]
123. Qudah I, Leijdekkers P, Gay V. Using mobile phones to improve medication compliance and awareness for cardiac patients. In: Proceedings of the 3rd International Conference on Pervasive Technologies Related to Assistive Environments. 2010 Presented at: 3rd International Conference on Pervasive Technologies Related to Assistive Environments; 2010; Samos, Greece p. 1-7. [doi: 10.1145/1839294.1839337]
124. Scholten MR, Kelders SM, Van Gemert-Pijnen JE. An empirical study of a pedagogical agent as an adjunct to an ehealth self-management intervention: what modalities does it need to successfully support and motivate users? Front Psychol 2019 May 9;10:1063 [FREE Full text] [doi: 10.3389/fpsyg.2019.01063] [Medline: 31143152]
125. Berrios-Thomas S, Engel R, McCall J, Rosen D. Assessing knowledge regarding managing congestive health failure symptoms: differences in patient and professional scores. J Patient Exp 2018 Dec 04;5(4):289-295 [FREE Full text] [doi: 10.1177/2374373518765793] [Medline: 30574550]
126. Kelly L, Jenkinson C, Morley D. Experiences of using web-based and mobile technologies to support self-management of type 2 diabetes: qualitative study. JMIR Diabetes 2018 May 11;3(2):e9 [FREE Full text] [doi: 10.2196/diabetes.9743] [Medline: 30291098]
127. Kujala S, Hörhammer I, Kaipio J, Heponiemi T. Health professionals' expectations of a national patient portal for self-management. Int J Med Inform 2018 Sep;117:82-87 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.06.005] [Medline: 30032968]
128. Mano R. Online health information, situational effects and health changes among e-patients in Israel: a 'push/pull' perspective. Health Expect 2015 Dec 17;18(6):2489-2500 [FREE Full text] [doi: 10.1111/hex.12218] [Medline: 24938381]
129. Nyatuka DR, De La Harpe R. Design considerations for patient‐centered interventions in an underserved context: a case of health and wellbeing services within Nairobi's informal settlements in Kenya. Electron J Inf Syst Dev Ctries 2021 Jan 12:e12164. [doi: 10.1002/isd2.12164]
130. Bengtsson U, Kasperowski D, Ring L, Kjellgren K. Developing an interactive mobile phone self-report system for self-management of hypertension. Part 1: patient and professional perspectives. Blood Press 2014 Oct;23(5):288-295 [FREE Full text] [doi: 10.3109/08037051.2014.883203] [Medline: 24564289]

131. Bradley SE, Haun J, Powell-Cope G, Haire S, Belanger HG. Qualitative assessment of the use of a smart phone application to manage post-concussion symptoms in Veterans with traumatic brain injury. Brain Inj 2020 Jul 02;34(8):1031-1038. [doi: 10.1080/02699052.2020.1771770] [Medline: 32493066]
132. Hallberg I, Ranerup A, Kjellgren K. Supporting the self-management of hypertension: patients' experiences of using a mobile phone-based system. J Hum Hypertens 2016 Feb;30(2):141-146 [FREE Full text] [doi: 10.1038/jhh.2015.37] [Medline: 25903164]
133. Wulfovich S, Fiordelli M, Rivas H, Concepcion W, Wac K. "I Must Try Harder": design implications for mobile apps and wearables contributing to self-efficacy of patients with chronic conditions. Front Psychol 2019 Oct 23;10:2388 [FREE Full text] [doi: 10.3389/fpsyg.2019.02388] [Medline: 31749733]
134. ICD-11 - International Classification of Diseases for mortality and morbidity statistics - eleventh revision. World Health Organization. 2021. URL: https://icd.who.int/browse11/l-m/en [accessed 2021-06-01]
135. Manders-Huits N. What values in design? The challenge of incorporating moral values into design. Sci Eng Ethics 2011 Jun 12;17(2):271-287 [FREE Full text] [doi: 10.1007/s11948-010-9198-2] [Medline: 20224927]
136. Health-related quality of life and wellbeing. HealthyPeople. 2010. URL: https://www.healthypeople.gov/2020/topics-objectives/topic/health-related-quality-of-life-well-being [accessed 2021-06-01]
137. Bowling A. Measuring Health: A Review of Quality of Life Measurement Scales. Milton Keynes, United States of America: Open University Press; 2005.
138. Huber M, Knottnerus JA, Green L, van der Horst H, Jadad AR, Kromhout D, et al. How should we define health? Br Med J 2011 Jul 26;343:d4163. [doi: 10.1136/bmj.d4163] [Medline: 21791490]
139. Seeman J. Toward a model of positive health. Am Psychol 1989;44(8):1099-1109. [doi: 10.1037/0003-066x.44.8.1099]
140. Ryff CD, Singer B. The contours of positive human health. Psychol Inq 2009 Nov 19;9(1):1-28. [doi: 10.1207/s15327965pli0901_1]
141. Seligman ME. Positive health. Appl Psychol 2008 Jul;57(s1):3-18. [doi: 10.1111/j.1464-0597.2008.00351.x]
142. Huber M. Towards a New, Dynamic Concept of Health. Its Operationalisation and Use in Public Health and Healthcare, and in Evaluating Health Effects of Food. Maastricht, The Netherlands: Maastricht University; 2014.
143. Locker D, Gibson B. The concept of positive health: a review and commentary on its application in oral health research. Community Dent Oral Epidemiol 2006 Jun;34(3):161-173. [doi: 10.1111/j.1600-0528.2006.00263.x] [Medline: 16674748]
144. 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]
145. Sackett DL. Evidence-based medicine. Seminars Perinatol 1997 Feb;21(1):3-5. [doi: 10.1016/s0146-0005(97)80013-4]
146. Medical devices: managing the mismatch. World Health Organization. 2010. URL: https://apps.who.int/iris/bitstream/handle/10665/44407/9789241564045_eng.pdf [accessed 2021-06-01]
147. Blandford A, Gibbs J, Newhouse N, Perski O, Singh A, Murray E. Seven lessons for interdisciplinary research on interactive digital health interventions. Digit Health 2018;4:2055207618770325 [FREE Full text] [doi: 10.1177/2055207618770325] [Medline: 29942629]
148. Orlowski S, Matthews B, Bidargaddi N, Jones G, Lawn S, Venning A, et al. Mental health technologies: designing with consumers. JMIR Hum Factors 2016 Jan 28;3(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.4336] [Medline: 27026210]
149. Fischer G. Exploring design trade-offs for quality of life in human-centered design. Interactions 2017 Dec 21;25(1):26-33. [doi: 10.1145/3170706]
150. Friedman B. Value-sensitive design. Interactions 1996 Dec;3(6):16-23. [doi: 10.1145/242485.242493]
151. Smits M, Ludden GD, Peters R, Bredie S, Van Goor H, Verbeek PP. Values that matter: a new method to design and assess moral mediation of technology. Design Issues 2022;38(1):39-54. [doi: 10.1162/desi_a_00669]
152. Norman D. Emotional Design - Why We Love (or Hate) Everyday Things. New York, United States of America: Basic Books; 2005.
153. Leikas J, Saariluoma P, Heinilä J, Ylikauppila M. A methodological model for life-based design. Int Rev Soc Sci Humanit 2013;4(2):118-136. [doi: 10.1057/978-1-137-53047-9_6]
154. Jacobs N. Capability sensitive design for health and wellbeing technologies. Sci Eng Ethics 2020 Dec 18;26(6):3363-3391 [FREE Full text] [doi: 10.1007/s11948-020-00275-5] [Medline: 33206283]
155. Desmet P, Pohlmeyer A. Positive design: an introduction to design for subjective wellbeing. Int J Design 2013;7(3):1-15 [FREE Full text]
156. Peters D, Calvo RA, Ryan RM. Designing for motivation, engagement and wellbeing in digital experience. Front Psychol 2018 May;9:797. [doi: 10.3389/fpsyg.2018.00797] [Medline: 29892246]
157. Riva G, Baños RM, Botella C, Wiederhold BK, Gaggioli A. Positive technology: using interactive technologies to promote positive functioning. Cyberpsychol Behav Soc Netw 2012 Feb;15(2):69-77. [doi: 10.1089/cyber.2011.0139] [Medline: 22149077]
158. Hassenzahl M. Experience design: technology for all the right reasons. Synth Lectures Hum-Cent Informat 2010 Jan;3(1):1-95. [doi: 10.2200/s00261ed1v01y201003hci008]
159. Calvo R, Peters D. Positive Computing: Technology for Wellbeing and Human Potential. Cambridge, Massachusetts, United States of America: MIT Press; 2014.

160. Albrechtslund A. Ethics and technology design. Ethics Inf Technol 2006 Dec 2;9(1):63-72. [doi: 10.1007/s10676-006-9129-8]
161. 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. [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
162. Mohr DC, Schueller SM, Riley WT, Brown CH, Cuijpers P, Duan N, et al. Trials of intervention principles: evaluation methods for evolving behavioral intervention technologies. J Med Internet Res 2015;17(7):e166 [FREE Full text] [doi: 10.2196/jmir.4391] [Medline: 26155878]
163. Cresswell KM, Blandford A, Sheikh A. Drawing on human factors engineering to evaluate the effectiveness of health information technology. J R Soc Med 2017 Aug;110(8):309-315 [FREE Full text] [doi: 10.1177/0141076817712252] [Medline: 28537105]
164. Collins LM, Murphy SA, Nair VN, Strecher VJ. A strategy for optimizing and evaluating behavioral interventions. Ann Behav Med 2005 Aug;30(1):65-73. [doi: 10.1207/s15324796abm3001_8] [Medline: 16097907]
165. Kiran AH. Mediating patienthood-from an ethics of to an ethics with technology. Nurs Philos 2017 Jan 29;18(1):e12153. [doi: 10.1111/nup.12153] [Medline: 27682857]
166. Lampe K, Mäkelä M, Garrido MV, Anttila H, Autti-Rämö I, Hicks NJ, et al. The HTA Core Model: a novel method for producing and reporting health technology assessments. Int J Technol Assess Health Care 2009 Dec 23;25(S2):9-20. [doi: 10.1017/s0266462309990638]
167. Kidholm K, Ekeland AG, Jensen LK, Rasmussen J, Pedersen CD, Bowes A, et al. A model for assessment of telemedicine applications: mast. Int J Technol Assess Health Care 2012 Jan;28(1):44-51. [doi: 10.1017/S0266462311000638] [Medline: 22617736]
168. Vlaeyen JW, Wicksell RK, Simons LE, Gentili C, De TK, Tate RL, et al. From boulder to Stockholm in 70 years: single case experimental designs in clinical research. Psychol Rec 2020 Jun 03;70(4):659-670. [doi: 10.1007/s40732-020-00402-5]
169. Enam A, Torres-Bonilla J, Eriksson H. Evidence-based evaluation of ehealth interventions: systematic literature review. J Med Internet Res 2018 Nov 23;20(11):e10971 [FREE Full text] [doi: 10.2196/10971] [Medline: 30470678]
170. Vis C, Bührmann L, Riper H, Ossebaard HC. Health technology assessment frameworks for eHealth: a systematic review. Int J Technol Assess Health Care 2020 Apr 16;36(3):204-216. [doi: 10.1017/s026646232000015x]
171. Guyatt G, Feeny D, Patrick D. Issues in quality-of-life measurement in clinical trials. Control Clin Trials 1991 Aug;12(4):81-90. [doi: 10.1016/s0197-2456(05)80014-5]
172. Taylor C. Ethics in health care and medical technologies. Theor Med Bioeth 1990 Jun;11(2):111-124. [doi: 10.1007/bf00489455]

Abbreviations

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
QoL: Quality of Life
RCT: Randomized Controlled Trial

Edited by A Mavragani; submitted 23.09.21; peer-reviewed by L McCann, E Farin; comments to author 21.12.21; revised version received 27.01.22; accepted 20.02.22; published 04.04.22.

Please cite as:
Smits M, Kim CM, van Goor H, Ludden GDS
From Digital Health to Digital Well-being: Systematic Scoping Review
J Med Internet Res 2022;24(4):e33787
URL: https://www.jmir.org/2022/4/e33787
doi:10.2196/33787
PMID:35377328

Review

The Potential of Current Noninvasive Wearable Technology for the Monitoring of Physiological Signals in the Management of Type 1 Diabetes: Literature Survey

Elena Daskalaki[1], MSc, PhD; Anne Parkinson[2], BA (Hons), PhD; Nicola Brew-Sam[2], BA, MA, PhD; Md Zakir Hossain[1,3,4], MSc, PhD; David O'Neal[5,6], MBBS, MD; Christopher J Nolan[7,8], BMedSci, MBBS, PhD; Hanna Suominen[1,9,10], MSc, PhD, Docent, MEdL

[1]School of Computing, College of Engineering and Computer Science, The Australian National University, Canberra, Australia

[2]Department of Health Services Research and Policy, Research School of Population Health, College of Health and Medicine, The Australian National University, Canberra, Australia

[3]School of Biology, College of Science, The Australian National University, Canberra, Australia

[4]Bioprediction Activity, Commonwealth Industrial and Scientific Research Organisation, Canberra, Australia

[5]Department of Medicine, University of Melbourne, Melbourne, Australia

[6]Department of Endocrinology and Diabetes, St Vincent's Hospital Melbourne, Melbourne, Australia

[7]Australian National University Medical School and John Curtin School of Medical Research, College of Health and Medicine, The Autralian National University, Canberra, Australia

[8]Department of Diabetes and Endocrinology, The Canberra Hospital, Canberra, Australia

[9]Data61, Commonwealth Industrial and Scientific Research Organisation, Canberra, Australia

[10]Department of Computing, University of Turku, Turku, Finland

Corresponding Author:
Elena Daskalaki, MSc, PhD
School of Computing
College of Engineering and Computer Science
The Australian National University
145 Science Road
Canberra, 2600
Australia
Phone: 61 0261256188
Email: eleni.daskalaki@anu.edu.au

Abstract

Background: Monitoring glucose and other parameters in persons with type 1 diabetes (T1D) can enhance acute glycemic management and the diagnosis of long-term complications of the disease. For most persons living with T1D, the determination of insulin delivery is based on a single measured parameter—glucose. To date, wearable sensors exist that enable the seamless, noninvasive, and low-cost monitoring of multiple physiological parameters.

Objective: The objective of this literature survey is to explore whether some of the physiological parameters that can be monitored with noninvasive, wearable sensors may be used to enhance T1D management.

Methods: A list of physiological parameters, which can be monitored by using wearable sensors available in 2020, was compiled by a thorough review of the devices available in the market. A literature survey was performed using search terms related to T1D combined with the identified physiological parameters. The selected publications were restricted to human studies, which had at least their abstracts available. The PubMed and Scopus databases were interrogated. In total, 77 articles were retained and analyzed based on the following two axes: the reported relations between these parameters and T1D, which were found by comparing persons with T1D and healthy control participants, and the potential areas for T1D enhancement via the further analysis of the found relationships in studies working within T1D cohorts.

Results: On the basis of our search methodology, 626 articles were returned, and after applying our exclusion criteria, 77 (12.3%) articles were retained. Physiological parameters with potential for monitoring by using noninvasive wearable devices in persons with T1D included those related to cardiac autonomic function, cardiorespiratory control balance and fitness, sudomotor function,

and skin temperature. Cardiac autonomic function measures, particularly the indices of heart rate and heart rate variability, have been shown to be valuable in diagnosing and monitoring cardiac autonomic neuropathy and, potentially, predicting and detecting hypoglycemia. All identified physiological parameters were shown to be associated with some aspects of diabetes complications, such as retinopathy, neuropathy, and nephropathy, as well as macrovascular disease, with capacity for early risk prediction. However, although they can be monitored by available wearable sensors, most studies have yet to adopt them, as opposed to using more conventional devices.

Conclusions: Wearable sensors have the potential to augment T1D sensing with additional, informative biomarkers, which can be monitored noninvasively, seamlessly, and continuously. However, significant challenges associated with measurement accuracy, removal of noise and motion artifacts, and smart decision-making exist. Consequently, research should focus on harvesting the information hidden in the complex data generated by wearable sensors and on developing models and smart decision strategies to optimize the incorporation of these novel inputs into T1D interventions.

(J Med Internet Res 2022;24(4):e28901) doi:10.2196/28901

KEYWORDS

type 1 diabetes; wearable sensors; big data; consumer health informatics; mobile health; survey

Introduction

Type 1 Diabetes

In healthy individuals, glucose levels are maintained within tight upper and lower bounds because of a complex physiological closed-loop regulatory process based on the accurate and timely secretion of insulin and glucagon into the portal vein by pancreatic islet cells [1]. To achieve this, the human body contains numerous sensory mechanisms that track and even anticipate fluctuations in glucose owing to meal intake, exercise, and other factors. This information is used to estimate the optimal secretion of insulin as well as other hormones, such as incretins, glucagon, and adrenaline, which play important modulating roles [2].

In persons with type 1 diabetes (T1D), insulin secretion by the pancreas is absent because of the autoimmune destruction of the pancreatic beta cells, breaking the normal closed-loop regulation process [3]. The absence of insulin results in the inability to metabolize glucose and an unregulated catabolic state leading to hyperglycemia and ketoacidosis, a recognized complication of diabetes [4]. Persons living with T1D are dependent on exogenous insulin (usually by subcutaneous administration) to regulate their glucose levels [5].

Poor glucose control leads to both acute and chronic complications, which may be life-threatening [6]. One common acute complication is hypoglycemia, a state of low blood glucose (BG) concentration that results from excessive insulin administration. Hypoglycemia can lead to loss of consciousness, coma, and even death [7]. Those who develop impaired awareness to hypoglycemia (IAH), a condition in which the individual does not experience the usual early warning symptoms of low BG, are at a much higher risk of severe hypoglycemic events [8].

On the other hand, chronic hyperglycemia leads to long-term complications, one of the most common being cardiac autonomic neuropathy (CAN). The reported prevalence of CAN in persons with T1D spans a very wide range, indicatively 17%-90%, depending on the criteria used for its diagnosis and the population studied [9]. CAN results in impaired cardiovascular autonomic control as a consequence of autonomic nerve neuronal metabolic and ischemic damage. Apart from CAN, T1D is linked to other long-term complications such as peripheral neuropathy, retinopathy, nephropathy, and atherosclerotic vascular disease [6].

Maintaining glucose levels within a healthy range in T1D represents a therapeutic challenge, and the accurate estimation of an individual's insulin requirements is key to achieving this goal [10,11]. Daily glucose management places a significant physical and cognitive burden on the persons living with T1D and their families. As part of ambulatory care, long and rapid acting insulins are administered using insulin injections (insulin pens) or rapid acting insulin is administered continuously and subcutaneously by insulin infusion pumps. Both insulin pumps and injections may be used in combination with continuous glucose monitors (CGMs). Insulin pumps integrated with CGMs are referred to as sensor-augmented pumps and their use has been shown to improve glycemic control [12,13]. More recently, control algorithms have been incorporated to automate basal insulin delivery; however, the user is still required to administer a bolus before meals. This treatment scheme is referred to as a hybrid closed-loop system. Hybrid closed-loop schemes have been shown to improve glucose control in comparison with insulin dosing entirely determined by the user [14] and have paved the way for the holy grail of T1D management, which is the development of fully closed-loop functionality or an artificial pancreas (AP) [15,16].

For optimal T1D management, accurate real-time sensing of key physiological parameters (ie, biomarkers) is essential. Sensing is crucial for daily acute glycemic management to ensure the correct estimation of insulin dose, as well as for the early recognition of long-term complications. To date, T1D sensing for acute glycemic management is focused on glucose monitoring, measured either by finger pricking or CGMs, which are associated with painful and sparse measurements or significant sensing lags, respectively [17,18]. However, monitoring of glucose alone has significant limitations in informing optimal insulin delivery, whether in open or closed-loop systems. In addition, monitoring of diabetes complications such as CAN is predominantly performed through infrequent assessments during clinic visits. Please refer to Multimedia Appendix 1 [9,17-23] for a more detailed discussion

on the current status and challenges of sensing biomarkers of T1D. This survey, which reviewed 77 publications, investigated novel physiological parameters with the potential to be incorporated into T1D management systems and reduce the impact of T1D on quality of life, specifically on parameters that can be monitored with the wearable sensors available today.

Wearable Sensors

Wearable sensors or wearables are becoming increasingly popular because they can provide seamless and continuous monitoring at low cost. Wearable sensors are available in various forms and shapes and can be worn at different body sites (Figure 1). Currently, wearables monitor physiological parameters, such as heart rate (HR), respiratory rate, oxygen saturation (SpO_2), skin temperature (ST), electrochemical skin conductance (ESC), and galvanic skin response (GSR). The most recent devices provide electrocardiogram (ECG) monitoring, which carries rich information about features of the cardiac state, such as HR variability (HRV) and QT interval (time interval from the start of the Q-wave to the end of the T-wave in an ECG).

This offers new potential for wearable devices to move beyond their original purpose of fitness and wellness monitoring to that of continuous health care monitoring. Although not yet integrated into clinical practice, research community is currently investing in the development of medical applications using wearable sensors that can assist or complement routine medical procedures and disease monitoring practices. These efforts usually combine wearable sensors with advanced data processing methods such as machine learning (ML) algorithms to address the volume and complexity of the produced data (noise, motion artifacts, and gaps) and to build smart decision support or diagnostic systems (Multimedia Appendix 2 [24-37]).

Despite this recent move toward health care, the potential of wearables in T1D management has not been investigated much to date. The purpose of this survey was to explore the potential of monitoring physiological parameters with wearable sensors to assist in acute glycemic management and diagnosis and monitoring of complications in T1D. We devised a search and analysis framework to investigate the potential of *wearable, noninvasive sensors* (hereafter referred to as wearable sensors without explicitly repeating their noninvasive property). To the authors' knowledge, a similar survey has not been conducted thus far. We hypothesized that the demonstration of strong links between physiological parameters measurable with wearable sensors and T1D, combined with their intrinsic advantages for use, will support and inspire developments and research focused on the enhancement of real-life T1D applications.

Figure 1. Different types of wearable technology by ForbesOste (license: CC BY-NC-ND 2.0).

Methods

Research Question

The survey aimed to answer the following research question: *Do the existing wearable, noninvasive sensors have the potential to improve how T1D is monitored and managed?* We focused only on noninvasive wearable sensors. Minimally invasive wearables (eg, CGMs) or noninvasive portable sensors (eg, breath sensors) were outside the scope of the survey. We did not limit our search to studies that used wearable sensors but rather formulated our research question toward exploring the existence of clinical relationships between T1D and physiological parameters that could be (but not necessarily) monitored with the wearable, noninvasive technology available in the market today.

Search Methodology

As a first step, a list of physiological parameters that could be monitored with wearable, noninvasive sensors available in 2020 was compiled, based on a review of the wearable devices available in the market today. These parameters were used in the search query combined with keywords related to T1D (Textboxes 1 and 2).

Figure 2 illustrates the methodological steps we took to select the final set of articles, following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [38]. The details of the search and selection methodology are described in Multimedia Appendix 3 [38]. Following these steps, 77 articles were retained for analysis.

Textbox 1. Search terms used for wearable-enabled physiological parameters.

Heart rate

- Heart rate

Heart rate variability

- Heart rate variability

Breath rate

- Breath* rate
- Respirat* rate

Breath rate variability

- Respirat* variability
- Breath* rate variability

Oxygen saturation

- Oxygen saturation
- SpO_2

Body motion

- Accelerometer*
- Gyroscope

Skin properties

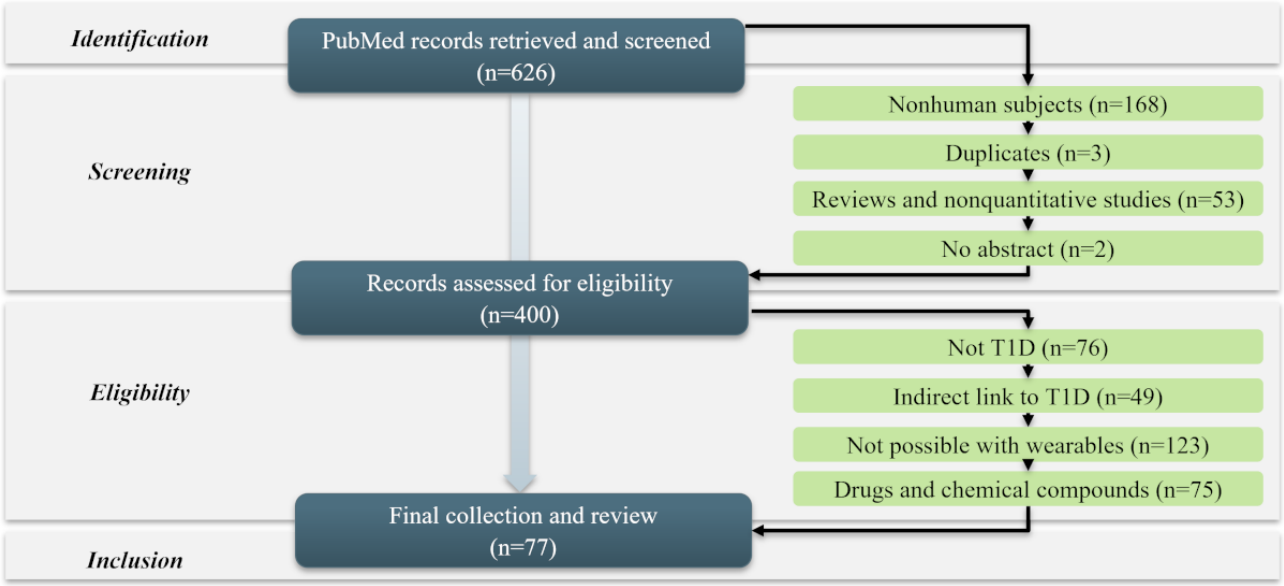
- Galvanic skin response
- Skin conductance
- Skin impedance
- Skin temperature
- Sweat

Textbox 2. The final search query used.

Search query

- Diabetes AND ("Type 1" OR "Type one" OR juvenile) AND ("heart rate" OR "heart rate variability" OR "respiration rate" OR "respiratory rate" OR "breath rate" OR "breathing rate" OR "respiration variability" OR "galvanic skin response" OR "skin conductance" OR "skin impedance" OR sweat OR accelerometer* OR gyroscope* OR "oxygen saturation" OR SpO_2)

Figure 2. Article selection methodology according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. T1D: type 1 diabetes.



Results

Article Structure and Analysis

The 77 retained articles shared similar structures. Data were collected through a clinical protocol whereby each study collected two types of data: (1) phenotypic characteristics of the participants with T1D (eg, age, sex, duration of diabetes, and glucose and hemoglobin A_{1c} [HbA_{1c}] levels), with or without the presence of nondiabetic control participants and (2) physiological function of interest (eg, cardiac autonomic function), with one or more physiological parameters measured (eg, HRV for autonomic function and ESC for sudomotor function). Through the extraction of features and data analysis, every study explored the relationship between T1D phenotypic characteristics and the physiological function of interest. All articles referred to clinical studies involving human participants. Although the articles did not necessarily use wearable devices in their methods, they all measured physiological parameters monitorable with commercially available wearable sensors, as guaranteed by our exclusion criteria (please refer to Multimedia Appendix 3).

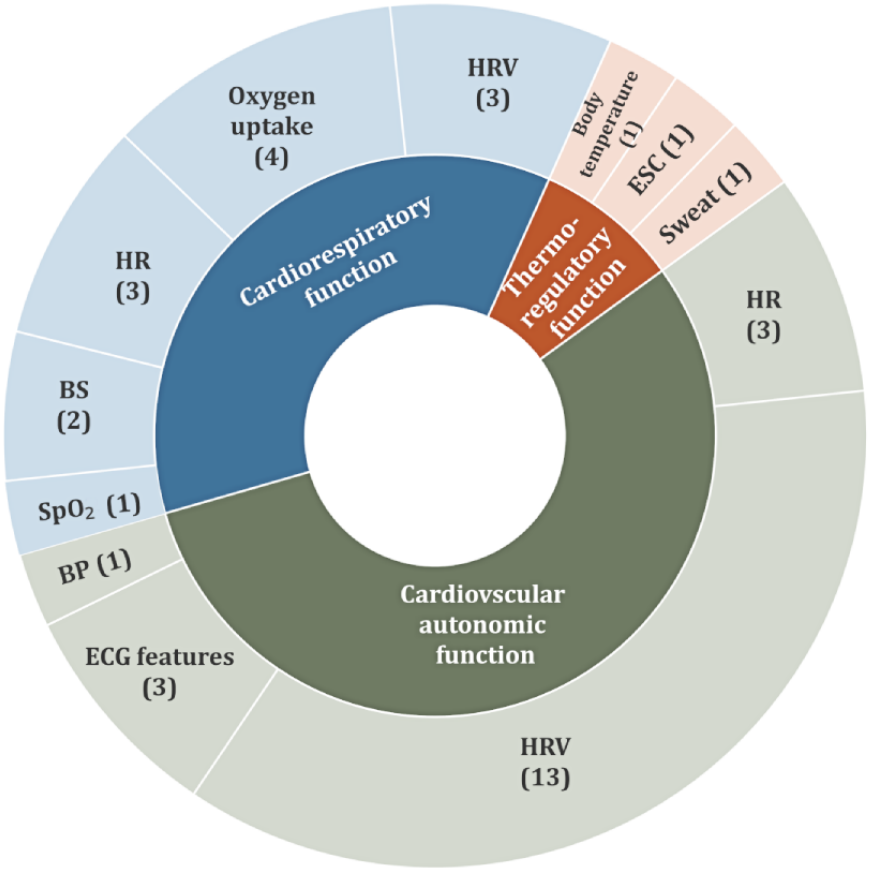
Our analysis of the retained articles evolved along the axis of this survey's research question: *Do the existing wearable, noninvasive sensors have the potential to improve how T1D is monitored and managed?* The surveyed studies were clustered into the following two broad categories: (1) those that performed a comparison between a cohort of persons with T1D and a matched group of healthy control participants and (2) those that explored relations within a cohort of persons with T1D. The first category informed about the physiological functions and parameters that were altered in persons living with T1D. The second category explored whether the monitoring of these parameters could benefit the management of T1D and its complications. All articles included in this survey are tabulated in Multimedia Appendix 4 [39-114].

Comparison Between Persons With T1D and Healthy Cohorts

Overview

The articles that compared people who were healthy and those with T1D revealed a wide spectrum of physiological functions or conditions that were affected by T1D (Figure 3). These predominantly included aspects of cardiovascular autonomic function, cardiorespiratory control balance, and thermal homeostasis.

Figure 3. Physiological functions found that are affected by type 1 diabetes (T1D) and the number of relevant studies. This graph refers only to the studies that compared T1D with healthy cohorts. BP: blood pressure; BS: Baroreflex Sensitivity; ECG: electrocardiogram; ESC: electrochemical skin conductance; HR: heart rate; HRV: heart rate variability; SpO_2: oxygen saturation.



Cardiovascular Autonomic Function

Many studies demonstrated that T1D impacted cardiac autonomic function by *reducing parasympathetic activity* through increased HR [39,40], reduced average HRV, and modified HRV features at rest or during cardiac autonomic reflex tests. The most explored HRV features were the following time- and frequency-domain features: SD of the normal-to-normal intervals, root mean square of the difference of successive intervals, high frequency (HF), low frequency (LF), and their ratio (LF/HF) [41-43]. Nonlinear features related to the complexity, dynamics, and chaotic components of HRV have also shown to be altered in persons with T1D. These include HRV randomness [44], symbolic indices [45], Katz fractal dimension [46], and geometric indices, such as the Poincare plot [47]. The effect of T1D on HRV was also demonstrated during exercise [48]. The overall HRV entropy during vigorous exercise was shown to be reduced in both healthy and T1D cohorts, although the attenuation in the T1D group was greater [49]. The impact on parasympathetic function was additionally demonstrated through the *reduction of the cardiac vagal tone*, which was shown to be associated with the presence of neuropathy [50]. *Cardiac depolarization and repolarization time intervals* were shown to increase in T1D demonstrated through modified ECG features such as increased HR-corrected QT (QT corrected [QTc]) interval [51,52] and more asymmetrical and flatter T-wave [53]. Table S1 in Multimedia Appendix 4 summarizes the methods and main findings of the above articles.

Cardiorespiratory Control Balance and Fitness

The impact of T1D on the *cardiorespiratory control balance and fitness* was assessed during cardiopulmonary exercise, as well as by cardiovascular (baroreflex) and respiratory (chemoreflex) response testing (Table S2 in Multimedia Appendix 4). Moser et al [54] demonstrated that there were clear differences in the HR response during cardiopulmonary exercise between persons with T1D and those in the control group. Cardiorespiratory control imbalance, with impaired *sensitivity to hypoxia* as evidenced by lower *resting* SpO_2 and *baroreflex sensitivity* and increased *chemoreflex sensitivity to hypercapnia*, was shown in persons with recent onset of T1D compared with control participants [55]. With respect to cardiorespiratory fitness, lower *peak oxygen uptake* (VO_2) and *respiratory exchange ratios* were reported in persons with T1D compared with control participants [56]; however, Rohling et al [57] showed that cardiorespiratory function was preserved in T1D, whereas the maximum VO_2 correlated with HRV time indices. Several studies demonstrated that persons with T1D who were well trained or well controlled could have similar cardiorespiratory fitness as the control participants [58-61].

Thermal Homeostasis

The *sudomotor function* is an activation of the sympathetic nervous system, controls the perspiration through sweat glands

and heat loss via skin, and is associated with increased blood flow. This function, which is usually assessed by measuring the ESC or GSR, was shown to be lower in persons with T1D compared with healthy control participants [62]. Other aspects of thermal homeostasis affected in persons with T1D included the *sweat profile*, *body temperature*, and *ST during exercise* [49,63,64]. The methods and findings of the aforementioned studies are presented in Table S3 in Multimedia Appendix 4.

Summary

The aforementioned studies demonstrated that T1D affected several physiological functions by comparing T1D cohorts with healthy control participants under resting conditions or during physical activity. Most of the studies focused on aspects of cardiac autonomic function, which seems to be one of the first nonmetabolic physiological functions affected by T1D.

Correlations Within a T1D Cohort

Overview

The potential to improve the life of persons living with T1D using wearable sensors relies on not only the capability to detect differences between persons with T1D and healthy individuals but also on the intra- and interindividual differences within populations with T1D. In this section, we present the studies found in this survey that assessed physiological functions within T1D cohorts and discuss the potential for monitoring these functions with wearables to incorporate into T1D acute glycemic management (Figure 4) and early recognition of long-term complications (Figure 5).

Figure 4. Identified areas of potential impact of wearable devices on the acute glycemic management of type 1 diabetes. The numbers indicate the number of relevant studies found for each physiological signal. ECG: electrocardiogram; ESC: electrochemical skin conductance; HR: heart rate; HRV: heart rate variability; NIR: near-infrared; SpO_2: oxygen saturation; ST: skin temperature.

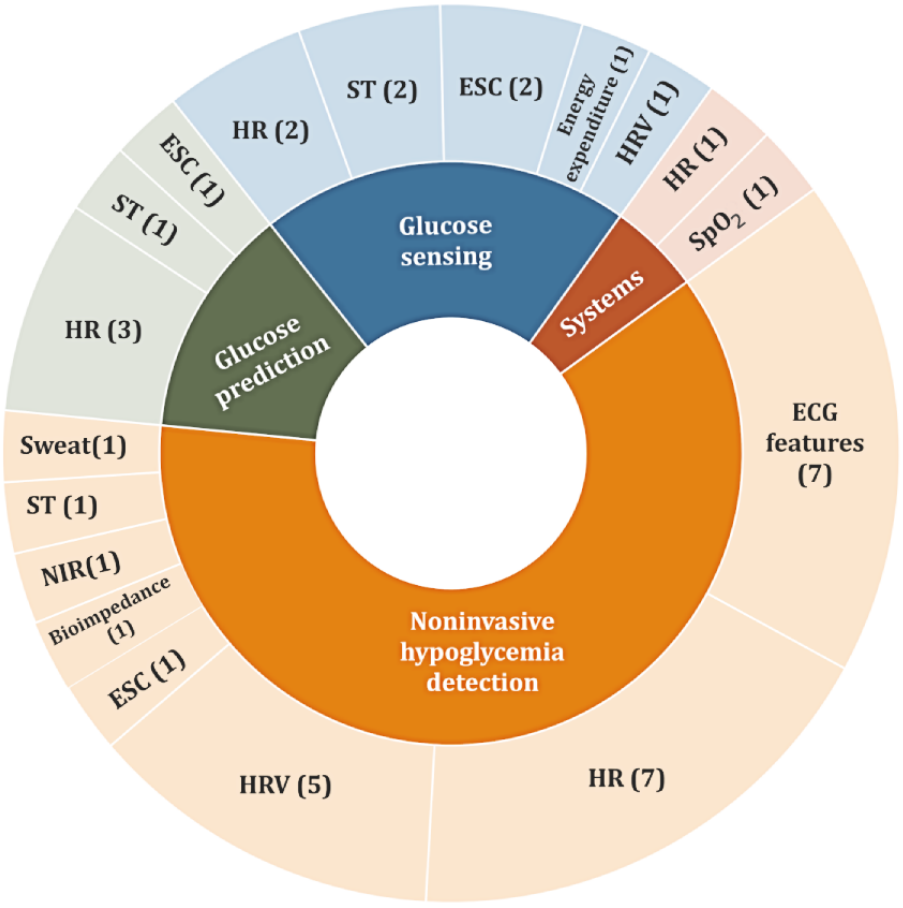
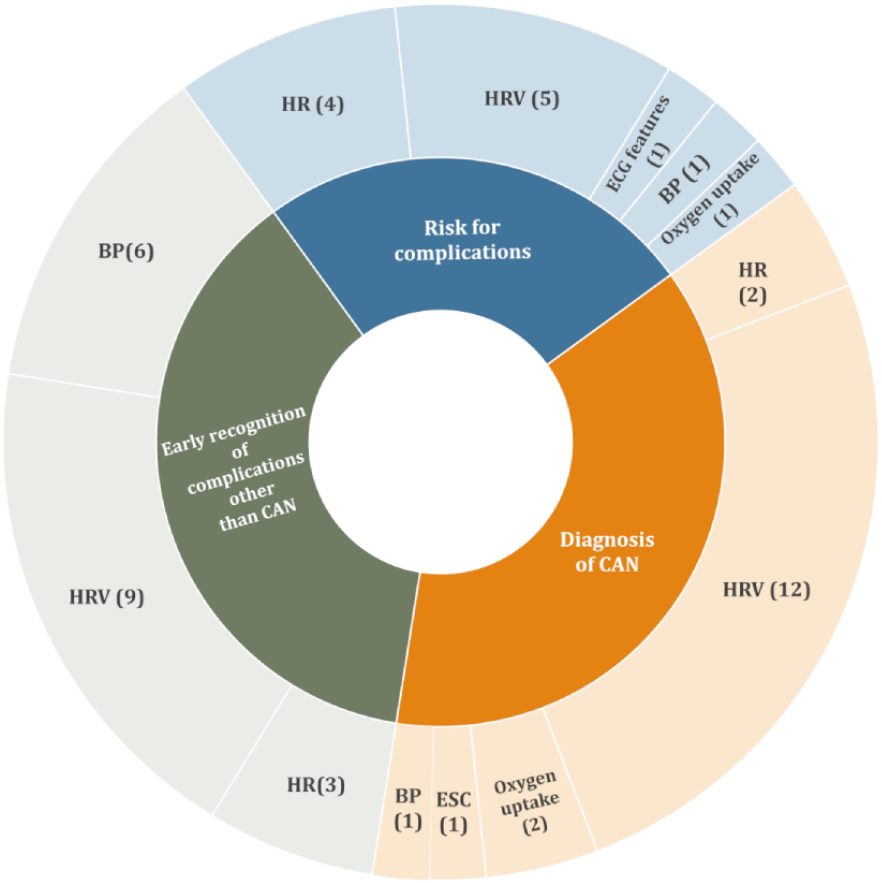


Figure 5. Identified areas of potential impact for wearable devices related to long-term type 1 diabetes complications. The numbers indicate the number of relevant studies found for each physiological signal. BP: blood pressure; CAN: cardiac autonomic neuropathy; ECG: electrocardiogram; ESC: electrochemical skin conductance; HR: heart rate; HRV: heart rate variability.



Acute T1D Glycemic Management

Noninvasive Hypoglycemia Detection

As discussed above, the presence of T1D influences cardiac function and the presence of hypoglycemia exaggerates this with stronger activation of sympathetic activity and inhibition of parasympathetic activity as shown in modified time and frequency indices of HRV [65-67] and QTc prolongation [68-70] (Table S4 in Multimedia Appendix 4). The effect of hypoglycemia on HRV was shown to be the same for people with and without CAN [71]. The signature of hypoglycemia on physiological features was leveraged in 5 studies to investigate methods by which hypoglycemia could be detected noninvasively. Of the 5 studies, 3 (60%) used cardiac features from an ECG (HR and QTc interval) as inputs [72-74] and ML-based techniques for developing relevant models. Elvebakk et al [75] explored noninvasive hypoglycemia detection for persons with and without IAH using cardiac and sudomotor features. The study found that it was only possible to detect hypoglycemia in the group of people without IAH, but it was difficult to reliably detect hypoglycemia in persons with IAH. However, from a second study focused only on persons with IAH, the same group proposed the use of a probabilistic model and the combined use of more cardiac and thermoregulatory features, ECG-derived HR, QTc interval, sudomotor activity, near-infrared and bioimpedance spectroscopy, and achieved a detection *F* score of 88% [76]. Finally, Reddy et al [77], using a wearable sensor and an ML algorithm, showed that the levels of HR and BG before exercise were predictors of exercise-induced hypoglycemia with 80% accuracy.

Near-Future Glucose Prediction

Of the 77 studies, 3 (4%) explored the potential of using physiological features monitored with wearable sensors to enhance the accuracy of near-future glucose prediction (Table S5 in Multimedia Appendix 4). The study by Rodriguez et al [78] presented a feature selection method for the estimation of feature importance related to near-future BG prediction and showed that HR and sleep were significant for this task (P=.03 and P=.04 respectively). This result was confirmed by Hobbs et al [79], who explored the impact on the 30-minute ahead glucose prediction of HR data collected from a commercial wearable device in adults during physical activity. The study showed that HR information led to better model prediction accuracy, assessed over several metrics, such as root mean square error and the Akaike information criterion, and reduced the prediction delay by 2-3 minutes compared with the model that did not consider HR. Mirshekarian et al [80] showed that when ESC and HR were added as features to an ML model for glucose prediction, the performance in terms of root mean square error improved for horizons of 30 and 60 minutes, although the improvement was only small.

Glucose Sensing

Of the 77 studies, 3 (4%) explored methods to improve the accuracy of glucose sensing (Table S5 in Multimedia Appendix 4). Here, we identified the work of Laguna et al [81], which aimed to improve the known increase in CGM error during exercise by coupling the CGM with wearable sensors. The study found that when CGMs were coupled with energy expenditure (metabolic equivalent of tasks) and ST information from the wearable sensors, the glucose measurement error dropped from 17.5% to 13.6%. A study by Turksoy et al [82] followed a similar approach whereby the objective was to model the glucose changes during exercise as a function of biometric inputs including HR, GSR, ST, and energy expenditure. The study involved 26 participants who wore wearable sensors for 6 days. The participants performed a series of exercises during the regular clinic visits. The study results showed that ST was the most significant feature describing glucose fluctuation during exercise, followed by HR and energy expenditure. Finally, Rothberg et al [83] explored links between HRV features and absolute BG levels, which were confirmed for persons with type 2 diabetes but not for those with T1D.

Integrated Systems Solutions

In view of the broad application range of wearable devices in T1D management, one study identified in this survey focused on the development of an integrated system solution for home-based monitoring and explored the engagement of persons living with T1D using this technology [115].

Summary

The aforementioned studies illustrate the potential of using noninvasive biomarkers, which can be monitored with wearable sensors, to improve many aspects in the daily management of T1D (Figure 4). Noninvasive hypoglycemia detection could enhance the recognition of hypoglycemic events in people with IAH and for those who do not use CGMs and rely only on sporadic BG measurements. Near-future glucose prediction is crucial to compensate for the CGM and insulin action delays for more accurate prediction of upcoming hypo- and hyperglycemic events. All the above functionalities are also indispensable for the enhancement of the current AP systems, both for the development of more efficient control algorithms and for the support of the general safety of APs. Although the improvements demonstrated so far have limited clinical significance, these studies are promising first steps. More studies need to be conducted, including a thorough exploration of prediction modeling techniques, as well as the collection of comprehensive data sets that span representative real-life scenarios.

Long-term Complications

Diagnosis and Monitoring of CAN

As demonstrated by many studies, among the cardiac features, HRV plays a central role in the assessment of autonomic function in general and of CAN in particular [84-88]. Silva et al [89] showed that a high resting HR was associated with reduced parasympathetic activity and lower HRV. Finally, insulin resistance was shown to be associated with lower cardiac output [90] and cardiovascular suppression [91] during and after exercise or cardiac autonomic reflex tests. Anxiety [92] and psychosocial stress [93] were shown to reduce HRV and induce further parasympathetic suppression in persons with T1D, indicating the need to consider confounding factors in assessing CAN. Gender and race were also found to be associated with autonomic function [94]. Finally, a study by Ang et al [95] explored the potential of using ESC as a measure of sudomotor function for the early recognition of CAN but found no significant correlations. However, Riguetto et al [96] showed that postprandial sweating, as well as hypertension, diastolic blood pressure, retinopathy, and nephropathy were independent predictors of CAN. The methods and findings of the above studies are presented in Table S6 in Multimedia Appendix 4.

Early Recognition of Complications Other Than CAN

Of the 77 studies, 10 (13%) studies explored other long-term complications of T1D and their relation to cardiovascular autonomic function, including the presence of CAN (Table S7 in Multimedia Appendix 4). In a study by Duvnjak et al [97], the coefficient of variation and spectral indices of HRV correlated with *diabetic retinopathy* development and progression in persons with T1D. In another study of persons with T1D, correlations were demonstrated between HRV indices and neuroretinal layer thickness [98]. A study by Mala et al [99] demonstrated that CAN had a positive association with carotid intima media thickness in T1D and suggested its potential role in the pathogenesis of *atherosclerosis*. HRV is associated with *diabetic kidney disease* as discussed by Sekercioglu et al [100], who showed that, for adults with prolonged T1D duration, the factors of older age at diagnosis and lower HRV might indicate a risk for this complication. CAN in persons with T1D has also been shown to be associated with reduced *bone density* [101], *renal function and albuminuria* [102,103], and *female sexual dysfunction and urinary incontinence* [104]. Finally, in a study by da Silva et al [105], young persons with T1D and increased risk for cardiovascular disease presented greater parasympathetic autonomic dysfunction, whereas a study by Christensen et al [106] suggested that sudomotor function was associated with diabetic *peripheral neuropathy* and could be used as a diagnostic tool for these complications.

Risk for Long-term Complications

The quality of glucose control is an uncontroversial marker of the risk of long-term complications. Of the 77 studies, 8 (10%) studies explored the relationship between features of the cardiac and cardiorespiratory function and the quality of glucose control such that their findings could be useful in developing tools for complication risk assessment (Table S8 in Multimedia Appendix 4). Higher levels of HbA_{1c} were shown to correlate with lower HRV levels and higher CAN prevalence [107,108]. A study by Guan et al [109] showed that people with higher HbA_{1c} levels displayed an impaired autonomic response to stress, including a greater change in the HF component of their HRV, whereas Stern et al [110] showed that the QTc interval in persons with T1D correlated with HbA_{1c} and autonomic function. Glucose variability has been shown to correlate with CAN [111] but not with the loss of nocturnal lowering (ie, nondipping) of blood pressure [112]. Similarly, an association between HbA_{1c} and HR dynamics, including the HR-to-performance curve, was

demonstrated during cardiopulmonary exercise tests [113], as well as an association between HbA_{1c} and cardiorespiratory fitness as assessed by the time to exhaustion and peak VO_2 [114]. Finally, the study by Zabeen et al [13] showed that glucose variability, independent of HbA_{1c}, plays a role in the development of risk for long-term complications, such as retinopathy and neuropathy.

Summary

The evidence presented in the aforementioned studies indicates that autonomic function and CAN could be continuously assessed through wearable devices that provide HRV monitoring. In such a case, the effects of factors such as time of day, meals, exercise, sleep, and glycemic level, as well as therapies for CAN, can be assessed. Moreover, it was shown that autonomic and sudomotor functions could flag risks or onset for a range of T1D long-term complications. Monitoring autonomic function has the potential to complement measures of overall glycemic control and glucose variability in optimizing management to mitigate the risk of long-term complications. Figure 5 summarizes these findings. The relationship found between physiological parameters and quality of glucose control, such as HbA_{1c} or glucose variability, could also be used in the formulation of AP cost functions. In all cases, further studies are needed to investigate how to use the information harvested by wearable sensors toward supporting the early recognition of chronic T1D complications.

Discussion

Survey Findings

The data from many studies in this survey showed that a variety of physiological parameters (1) can differentiate persons with T1D from healthy control participants and (2) are associated with aspects of glycemic control, in particular hypoglycemia, and the presence of diabetes complications within cohorts of persons with T1D.

The most explored physiological functions were cardiac autonomic, cardiorespiratory control balance, and thermal homeostasis. The survey identified that monitoring of physiological parameters, such as HR, HRV, QTc, ESC, baroreflex sensitivity, and VO_2, as well as the autonomic, cardiorespiratory, and thermal homeostasis functions, can be used to identify differences between persons with T1D and healthy control participants. The most pronounced differences between the 2 populations were shown in aspects of cardiac autonomic function, including parasympathetic activity, vagal tone, and cardiac repolarization and depolarization.

With respect to noninvasive hypoglycemia prediction and detection, physiological parameters of the ECG, including HR, HRV, and QTc, have considerable potential to be leveraged. Apart from hypoglycemia detection, these physiological parameters, as well as ESC and ST, can be monitored in conjunction with glucose to compensate for the deficiencies in CGM, such as signal lag, which is vital to the enhancement of AP development.

The screening of long-term T1D complications can also be enhanced. The demonstrated relationship between CAN and HRV paves the way for its continuous, and at-home, assessment. Other complications such as retinopathy and diabetic peripheral neuropathy were shown to relate to cardiac autonomic function and CAN and their onset could be predicted through monitoring of these functions.

Most of the studies discussed in this survey followed conventional statistical analysis methods to assess the existence of correlations in their measured data. ML methods have been used in some studies for the recognition of hypoglycemia and near-future glucose prediction [72-74,77-80].

Most of the surveyed studies used conventional devices; however, the measurement of the above physiological parameters can be performed with the 2020 commercially available wearable, noninvasive sensors. Table 1 lists the identified physiological functions and parameters, together with examples of commercially available wearable sensors that can monitor them today. Only a few studies found in this survey used wearable sensors [64,66,77-82]. These studies mostly focused on the detection of hypoglycemia, glucose prediction, or improvement of CGM glucose measurement accuracy.

The existence of wearable technology that can perform this type of physiological parameter monitoring is a crucial first step in the confirmation of our survey hypothesis that wearables have the potential to enhance T1D sensing with richer information seamlessly and continuously toward improved daily management decisions, and mitigation of complications. In view of this potential, the challenges and perspectives of this endeavor are discussed further below.

Table 1. Identified physiological functions and parameters and examples of corresponding commercially available, wearable, noninvasive sensors.

Physiological functions	Physiological parameters	Existing wearable devices
Cardiac autonomic function	HR[a], HRV[b] (ECG[c])	QardioCore and Apple watch
Cardiac repolarization	QT[d], QTc[e], T-wave (ECG)	QardioCore and Apple watch
Cardiac output	Bioimpedance	BIOPAC
Energy expenditure	VO_2[f]	Garmin Forerunner 935, Fitbit Charge 2
Baroreflex sensitivity	ECG, BP[g]	QardioCore, Apple watch
Sweat rate	Sweat rate	KuduSmart monitor
Oxygen saturation	SpO_2[h]	Withings Pulse Ox; Garmin Fenix 6x
Sudomotor function	ESC[i]	Shimmer3 GSR[j]+ unit
Skin temperature	ST[k]	Tempatilumi CEBrazil; TIDA-00824 Texas Instruments (prototype)

[a]HR: heart rate.

[b]HRV: heart rate variability.

[c]ECG: electrocardiogram.

[d]QT: time interval from the start of the Q-wave to the end of the T-wave in an electrocardiogram.

[e]QTc: QT corrected.

[f]VO_2: oxygen uptake.

[g]BP: blood pressure.

[h]SpO_2: oxygen saturation.

[i]ESC: electrochemical skin conductance.

[j]GSR: galvanic skin response.

[k]ST: skin temperature.

Wearable Sensors Versus Medical Grade Devices

An important parameter when considering the potential of wearables is the quality of the generated data and the ability to extract the required information from them. A clinical ECG setup offers much higher accuracy than a wearable bracelet. Wearable sensors must compromise accuracy for small size, low cost, and high autonomy. Moreover, their default use involves people undertaking daily activities, which introduce motion artifacts and data corruption. Although in most of the studies reported in this survey, the involved participants followed a specified protocol under the supervision of a clinical staff member, this condition cannot be guaranteed or controlled in a daily life setting. To this end, a one-to-one comparison between a medical grade device and its wearable counterpart would always be an uneven battle. However, the claim of wearable sensors is not to substitute medical grade devices but rather to take up the space where the latter cannot be used; that is, the space of at-home, daily life routine. This different use case offers the following critical advantage over medical grade devices: the massive generation of data [24].

Compared with a clinical study, a wearable monitoring scenario produces vastly larger volumes of data over much longer periods and during various conditions, such as sleep, physical activity, eating, resting, stress, and working. Although wearables compromise accuracy, they can offer a significantly better and more representative coverage of a human life's spectrum. The *big data* return can compensate for data noise by leveraging redundancies and information fusion coming from different wearable sources. Long-term information, such as seasonality, might be revealed, which would not be feasible for short-duration clinical protocols. Finally, the possibility of monitoring a large number of participants opens the potential for population studies that are currently very difficult or very costly to perform. At the same time, efforts for disease management personalization can be significantly boosted by the increase in data availability per user. To this end, the usability of wearable sensors in health care should be considered in light of their space of function, the added value that this space can encompass, and the additional knowledge extraction possibilities that follow from their much larger volume of generated data.

In view of their high data quantity and complexity, wearable devices usually require advanced processing techniques for the harvesting of their data. Although classical signal processing techniques may be sufficient for the extraction of the HRV or QT interval from a medical grade ECG signal, ML strategies may need to be used to perform the same task on ECG data collected from a wearable device. At the same time, there is a requirement for increasingly complex solutions to support clinical judgment and decision-making to meet the current medical and user demands. To this end, wearable-based applications require a postsensing stage of complex data processing to drive usable and viable solutions. The development of processing techniques for information extraction and decision-making based on data generated by wearable sensors is a field that currently receives intensive attention and research. This manifests itself in most of the studies that propose the use

of wearables in health care, as discussed in the Introduction and Results sections.

Impact of Existing Wearable, Noninvasive Sensors on T1D

Despite the depth of research dedicated to the exploration of relations between T1D and other physiological functions, only a few studies found in this survey used wearable sensors. These studies, although still small in number, support the contention that off-the-shelf wearable devices could be readily used in T1D interventions. However, the adoption of this technology in the field is very slow. In addition to the data quality–related challenges, the complex relationships between these biomarkers and glucose regulation provide an additional hurdle. Further research is required for the development of models and simulations, design of treatment strategies for the inclusion of new inputs, and finally, conduction of further clinical trials to demonstrate the added clinical impact of the methods. Finally, the strict safety constraints that need to be guaranteed during the daily management of T1D render the validation of new methods very demanding. In view of these challenges, publicly available data sets have been released to support the research output toward the development of data analytics tools to incorporate information received by wearable devices into T1D interventions [116,117].

In summary, the major challenges in the adoption of wearable technology in the management of T1D and its complications are as follows:

1. Data coming from wearables tend to be noisy, corrupted by motion artifacts, and have lower accuracy than those originating from medical devices.
2. The relationship between the parameters monitored with wearables and glucose regulation is complex.
3. The strict safety constraints in the management of T1D impose hard boundaries on the testing and validation of new decision tools.

Benefits and Future Directions

Despite the hurdles, this survey makes the case that wearable sensors have a significant potential to enhance the life of those living with T1D. The identified links between physiological parameters that wearables can monitor and T1D can be used to augment the T1D sensing space and develop better management tools. The continuous monitoring potential and the abundance of generated data per person can assist in the personalization of interventions. At the same time, wearable sensors that provide seamless and noninvasive monitoring are expected to add a minimum sensing burden compared with other types of sensors, whereas the automation of management processes that today require the cognitive effort of the persons living with T1D (eg, insulin bolus calculation) or induce stress (eg, fear of hypoglycemia) is expected to lead to better quality of life and lower daily burden.

To harvest this potential, 2 main directions for future research can be identified. First, advanced data processing strategies need to be developed to extract the information obtained from the data collected through wearable sensors. This research direction is not specific to T1D, and T1D can profit from the research outcomes of every field (health care or other), which opts to use wearable sensors. Second, further simulations, models, and clinical studies need to be conducted to support the development of decision tools for T1D based on the data collected through wearables. This direction is T1D-specific and involves multidisciplinary collaboration among data scientists, engineers, clinicians in the field, and persons with lived experience of T1D.

Comparison With Previous Work

This survey aims to bridge on one side the large volume of research dedicated to identifying correlations between T1D glycemic control and complications with measurable physiological functions and on the other side, the novel potential of wearable technology in medical applications. The bulk of existing work dedicated to wearable technology and T1D is related to glucose sensing and omits other biomarkers that can be readily and noninvasively monitored with the available wearable sensors. To the authors' knowledge, a survey that explores the potential of wearable sensors in T1D has not been conducted till date. This is the first attempt to bring the fields of T1D and wearables together to highlight the potential of these sensors in the daily management of this disease and the mitigation of its long-term complications.

Survey Limitations

The main limitation of the survey was that the search was conducted based on a list of wearable-enabled biomarkers (Textbox 1), which may not be exhaustive. Additional knowledge models related to wearables and T1D may exist that were not identified. Moreover, the focus was only on noninvasive sensors and did not consider minimally invasive wearable technology, which can be another pathway to further enhance T1D diagnosis and management. Similarly, the survey focused only on wearable sensors and did not discuss noninvasive, nonwearable devices, such as breath or saliva sensors, which can also be highly advantageous in T1D (eg, breath acetone sensing).

Conclusions

Considering the wearable sensor boom and its gradual adoption in the health care domain, this survey aimed to investigate their potential impact on T1D, a chronic disease that affects millions of people worldwide and requires daily and costly management and care. The survey search strategy targeted the discovery of studies that examined the relationship between physiological functions or conditions measurable by wearable sensors and T1D. Our analysis showed that T1D greatly affects cardiac, cardiorespiratory, and thermoregulatory functions, and its impact can be readily observed through features of the ECG, such as HRV, QT interval, and T-wave, as well as skin properties such as ESC, temperature, and sweat profile. The effects of T1D on these functions manifest themselves at rest, overnight, during and after exercise, and during daily life activities. Importantly, they can be leveraged to improve the prompt detection of hypoglycemia, the efficiency of the AP, and the diagnosis of CAN and other complications.

Commercially available wearable technology exists for continuous, noninvasive monitoring of the above parameters.

For the successful adoption of this technology in health care in general, and T1D in particular, several challenges still need to be resolved, such as issues related to motion artifacts and noise removal, accurate extraction of the features of interest, and development of decision algorithms for improved and safe disease management. Current research efforts are working toward advanced algorithmic solutions for the efficient processing of massive amounts of data produced by wearable sensors. Their promising results can pave the way for similar endeavors for T1D.

Acknowledgments

This research was funded by and has been delivered in partnership with *Our Health in Our Hands*, a strategic initiative of the Australian National University, which aims to transform health care by developing new personalized health technologies and solutions in collaboration with patients, clinicians, and health care providers. The authors would like to thank Professor Matthew Cook, Professor Antonio Tricoli, and Mr Chris Pugmire for their helpful comments and suggestions. We also thank the *Our Health in Our Hands* Type 1 Diabetes Health Experience Team members with the lived experience of T1D for their insights and suggestions.

Authors' Contributions

ED and HS were responsible for conceptualizing and conducting the study, as well as writing the original manuscript. Subsequently, CJN, DON, NBS, MZH, and AP critically reviewed, commented, and revised the manuscript until its submission for a peer review by JMIR. HS and CJN contributed to the oversight and leadership responsibilities for research activity planning, resourcing, and execution.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sensing in type 1 diabetes.
[DOCX File , 35 KB - jmir_v24i4e28901_app1.docx]

Multimedia Appendix 2
Noninvasive wearable sensors in health care.
[DOCX File , 37 KB - jmir_v24i4e28901_app2.docx]

Multimedia Appendix 3
Survey methodology.
[DOCX File , 34 KB - jmir_v24i4e28901_app3.docx]

Multimedia Appendix 4
Summary of the studies included in the survey.
[DOCX File , 56 KB - jmir_v24i4e28901_app4.docx]

References

1. Röder PV, Wu B, Liu Y, Han W. Pancreatic regulation of glucose homeostasis. Exp Mol Med 2016 Mar 11;48:e219. [doi: 10.1038/emm.2016.6] [Medline: 26964835]
2. Rorsman P, Braun M. Regulation of insulin secretion in human pancreatic islets. Annu Rev Physiol 2013;75:155-179. [doi: 10.1146/annurev-physiol-030212-183754] [Medline: 22974438]
3. Atkinson MA, Eisenbarth GS, Michels AW. Type 1 diabetes. Lancet 2014 Jan 04;383(9911):69-82 [FREE Full text] [doi: 10.1016/S0140-6736(13)60591-7] [Medline: 23890997]
4. Dhatariya KK, Glaser NS, Codner E, Umpierrez GE. Diabetic ketoacidosis. Nat Rev Dis Primers 2020 May 14;6(1):40. [doi: 10.1038/s41572-020-0165-1] [Medline: 32409703]
5. Mortensen HB, Robertson K, Aanstoot H, Danne T, Holl R, Hougaard P, et al. Insulin management and metabolic control of Type 1 diabetes mellitus in childhood and adolescence in 18 countries. Diabet Med 1998 Sep;15(9):752-759 [FREE Full text] [doi: 10.1002/(sici)1096-9136(199809)15:9<752::aid-dia678>3.0.co;2-w]
6. Deshpande A, Harris-Hayes M, Schootman M. Epidemiology of diabetes and diabetes-related complications. Phys Ther 2008 Nov;88(11):1254-1264 [FREE Full text] [doi: 10.2522/ptj.20080020] [Medline: 18801858]
7. Cryer P, Davis S, Shamoon H. Hypoglycemia in diabetes. Diabet Care 2003 Jun;26(6):1902-1912 [FREE Full text] [doi: 10.2337/diacare.26.6.1902] [Medline: 12766131]

8. Gerich J, Mokan M, Veneman T, Korytkowski M, Mitrakou A. Hypoglycemia unawareness. Endocr Rev 1991 Nov;12(4):356-371 [FREE Full text] [doi: 10.1210/edrv-12-4-356] [Medline: 1760993]
9. Bissinger A. Cardiac autonomic neuropathy: why should cardiologists care about that? J Diabetes Res 2017;2017:5374176 [FREE Full text] [doi: 10.1155/2017/5374176] [Medline: 29214181]
10. Juvenile Diabetes Research Foundation. URL: https://jdrf.org.au/ [accessed 2022-02-15]
11. American Diabetes Association. URL: https://www.diabetes.org/ [accessed 2022-02-15]
12. Bergenstal RM, Tamborlane WV, Ahmann A, Buse JB, Dailey G, Davis SN, STAR 3 Study Group. Effectiveness of sensor-augmented insulin-pump therapy in type 1 diabetes. N Engl J Med 2010 Jul 22;363(4):311-320. [doi: 10.1056/NEJMoa1002853] [Medline: 20587585]
13. Zabeen B, Craig M, Virk S, Pryke A, Chan A, Cho Y, et al. Insulin pump therapy is associated with lower rates of retinopathy and peripheral nerve abnormality. PLoS One 2016;11(4):e0153033 [FREE Full text] [doi: 10.1371/journal.pone.0153033] [Medline: 27050468]
14. McAuley SA, Lee MH, Paldus B, Vogrin S, de Bock MI, Abraham MB, Australian JDRF Closed-Loop Research Group. Six months of hybrid closed-loop versus manual insulin delivery with fingerprick blood glucose monitoring in adults with type 1 diabetes: a randomized, controlled trial. Diabet Care 2020 Dec;43(12):3024-3033. [doi: 10.2337/dc20-1447] [Medline: 33055139]
15. Boughton C, Hovorka R. The artificial pancreas. Curr Opin Organ Transplant 2020 Aug;25(4):336-342 [FREE Full text] [doi: 10.1097/MOT.0000000000000786] [Medline: 32618719]
16. Kovatchev B. A century of diabetes technology: signals, models, and artificial pancreas control. Trends Endocrinol Metab 2019 Jul;30(7):432-444 [FREE Full text] [doi: 10.1016/j.tem.2019.04.008] [Medline: 31151733]
17. Schmelzeisen-Redeker G, Schoemaker M, Kirchsteiger H, Freckmann G, Heinemann L, Del Re L. Time delay of CGM sensors: relevance, causes, and countermeasures. J Diabetes Sci Technol 2015 Aug 04;9(5):1006-1015 [FREE Full text] [doi: 10.1177/1932296815590154] [Medline: 26243773]
18. Zaharieva DP, Turksoy K, McGaugh SM, Pooni R, Vienneau T, Ly T, et al. Lag time remains with newer real-time continuous glucose monitoring technology during aerobic exercise in adults living with type 1 diabetes. Diabetes Technol Ther 2019 Jun;21(6):313-321 [FREE Full text] [doi: 10.1089/dia.2018.0364] [Medline: 31059282]
19. Serhiyenko VA, Serhiyenko AA. Cardiac autonomic neuropathy: Risk factors, diagnosis and treatment. World J Diabet 2018 Jan 15;9(1):1-24 [FREE Full text] [doi: 10.4239/wjd.v9.i1.1] [Medline: 29359025]
20. Pop-Busui R. Cardiac autonomic neuropathy in diabetes: a clinical perspective. Diabet Care 2010 Feb;33(2):434-441 [FREE Full text] [doi: 10.2337/dc09-1294] [Medline: 20103559]
21. Donsa K, Spat S, Beck P, Pieber T, Holzinger A. Towards personalization of diabetes therapy using computerized decision support and machine learning: some open problems and challenges. In: Holzinger A, Röcker C, Ziefle M, editors. Smart Health: Open Problems and Future Challenges. Cham: Springer; 2015:237-260.
22. Vinik AI, Maser RE, Mitchell BD, Freeman R. Diabetic autonomic neuropathy. Diabet Care 2003 May;26(5):1553-1579. [doi: 10.2337/diacare.26.5.1553] [Medline: 12716821]
23. Brew-Sam N, Chhabra M, Parkinson A, Hannan K, Brown E, Pedley L, et al. Experiences of young people and their caregivers of using technology to manage type 1 diabetes mellitus: systematic literature review and narrative synthesis. JMIR Diabetes 2021 Feb 02;6(1):e20973 [FREE Full text] [doi: 10.2196/20973] [Medline: 33528374]
24. Gao W, Emaminejad S, Nyein H, Challa S, Chen K, Peck A, et al. Fully integrated wearable sensor arrays for multiplexed in situ perspiration analysis. Nature 2016 Jan 28;529(7587):509-514 [FREE Full text] [doi: 10.1038/nature16521] [Medline: 26819044]
25. Industrial wearables market by device type (AR glasses, VR headsets, smartwatches, smart bands), industry (automotive, aerospace, manufacturing, oil and gas, power and energy), component, and region - Global forecast to 2024. Research and Markets. 2019. URL: https://tinyurl.com/bdedmmzd [accessed 2022-02-15]
26. Fernández-Caramés T, Fraga-Lamas P. Towards the internet-of-smart-clothing: a review on IoT wearables and garments for creating intelligent connected e-textiles. Electronics 2018 Dec 07;7(12):405 [FREE Full text] [doi: 10.3390/electronics7120405]
27. Mehrang S, Pietilä J, Korhonen I. An activity recognition framework deploying the random forest classifier and a single optical heart rate monitoring and triaxial accelerometerwrist-band. Sensors (Basel) 2018 Feb 22;18(2):613 [FREE Full text] [doi: 10.3390/s18020613] [Medline: 29470385]
28. Rachim V, Chung W. Multimodal wrist biosensor for wearable cuff-less blood pressure monitoring system. Sci Rep 2019 May 28;9(1):7947 [FREE Full text] [doi: 10.1038/s41598-019-44348-3] [Medline: 31138845]
29. Witt D, Kellogg R, Snyder M, Dunn J. Windows into human health through wearables data analytics. Curr Opin Biomed Eng 2019 Mar;9:28-46 [FREE Full text] [doi: 10.1016/j.cobme.2019.01.001] [Medline: 31832566]
30. Haghi M, Thurow K, Stoll R. Wearable devices in medical internet of things: scientific research and commercially available devices. Healthc Inform Res 2017 Jan;23(1):4-15 [FREE Full text] [doi: 10.4258/hir.2017.23.1.4] [Medline: 28261526]
31. Song Y, Tan Y, Song Y, Wu P, Cheng J, Kim M, et al. Spatial and temporal variations of spatial population accessibility to public hospitals: a case study of rural–urban comparison. GIScience Remote Sens 2018 Mar 15;55(5):718-744 [FREE Full text] [doi: 10.1080/15481603.2018.1446713]

32. Airaksinen M, Räsänen O, Ilén E, Häyrinen T, Kivi A, Marchi V, et al. Automatic posture and movement tracking of infants with wearable movement sensors. Sci Rep 2020 Jan 13;10(1):169-113 [FREE Full text] [doi: 10.1038/s41598-019-56862-5] [Medline: 31932616]
33. Inan OT, Baran Pouyan M, Javaid AQ, Dowling S, Etemadi M, Dorier A, et al. Novel wearable seismocardiography and machine learning algorithms can assess clinical status of heart failure patients. Circ Heart Fail 2018 Jan;11(1):e004313 [FREE Full text] [doi: 10.1161/circheartfailure.117.004313]
34. Yin H, Jha N. A health decision support system for disease diagnosis based on wearable medical sensors and machine learning ensembles. IEEE Trans Multi-Scale Comp Syst 2017 Oct 1;3(4):228-241 [FREE Full text] [doi: 10.1109/tmscs.2017.2710194]
35. Izmailova E, Wagner J, Perakslis E. Wearable devices in clinical trials: hype and hypothesis. Clin Pharmacol Ther 2018 Jul;104(1):42-52 [FREE Full text] [doi: 10.1002/cpt.966] [Medline: 29205294]
36. Yavelberg L, Zaharieva D, Cinar A, Riddell M, Jamnik V. A pilot study validating select research-grade and consumer-based wearables throughout a range of dynamic exercise intensities in persons with and without type 1 diabetes: a novel approach. J Diabetes Sci Technol 2018 May;12(3):569-576 [FREE Full text] [doi: 10.1177/1932296817750401] [Medline: 29320885]
37. Castle JR, El Youssef J, Wilson LM, Reddy R, Resalat N, Branigan D, et al. Randomized outpatient trial of single- and dual-hormone closed-loop systems that adapt to exercise using wearable sensors. Diabet Care 2018 Jul;41(7):1471-1477 [FREE Full text] [doi: 10.2337/dc18-0228] [Medline: 29752345]
38. Liberati A, Altman D, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009 Jul 21;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
39. Wilson LC, Peebles KC, Hoye NA, Manning P, Sheat C, Williams MJA, et al. Resting heart rate variability and exercise capacity in Type 1 diabetes. Physiol Rep 2017 Apr;5(8) [FREE Full text] [doi: 10.14814/phy2.13248] [Medline: 28420762]
40. Metwalley K, Hamed S, Farghaly H. Cardiac autonomic function in children with type 1 diabetes. Eur J Pediatr 2018 Jun;177(6):805-813 [FREE Full text] [doi: 10.1007/s00431-018-3122-1] [Medline: 29500542]
41. Silva AK, Christofaro DG, Bernardo AF, Vanderlei FM, Vanderlei LC. Sensitivity, specificity and predictive value of heart rate variability indices in type 1 diabetes mellitus. Arq Bras Cardiol 2017 Mar;108(3):255-262 [FREE Full text] [doi: 10.5935/abc.20170024] [Medline: 28443958]
42. Ribeiro IJ, Pereira R, Neto PF, Freire I, Casotti C, Reis MG. Relationship between diabetes mellitus and heart rate variability in community-dwelling elders. Medicina (Kaunas) 2017;53(6):375-379 [FREE Full text] [doi: 10.1016/j.medici.2017.12.001] [Medline: 29428337]
43. Giacon TR, Vanderlei FM, Christofaro DG, Vanderlei LC. Impact of diabetes type 1 in children on autonomic modulation at rest and in response to the active orthostatic test. PLoS One 2016;11(10):e0164375 [FREE Full text] [doi: 10.1371/journal.pone.0164375] [Medline: 27788152]
44. Souza NM, Giacon TR, Pacagnelli FL, Barbosa MP, Valenti VE, Vanderlei LC. Dynamics of heart rate variability analysed through nonlinear and linear dynamics is already impaired in young type 1 diabetic subjects. Cardiol Young 2016 Oct;26(7):1383-1390. [doi: 10.1017/S104795111500270X] [Medline: 26838682]
45. Oliveira EA, Silva AK, Christofaro DG, Vanzella LM, Gomes RL, Vanderlei FM, et al. Influence of type 1 diabetes on the symbolic analysis and complexity of heart rate variability in young adults. Arq Bras Cardiol 2018 Jul;111(1):94-101 [FREE Full text] [doi: 10.5935/abc.20180117] [Medline: 30020326]
46. Garner D, Souza ND, Vanderlei L. Heart rate variability analysis: higuchi and katz's fractal dimensions in subjects with type 1 diabetes mellitus. Romanian J Diabetes Nutr Metab Dis 2018;25(3):289-295. [doi: 10.2478/rjdnmd-2018-0034]
47. Giacon TR, Vanderlei FM, Silva AK, Silva NT, Valenti VE, Vanderlei LC. Influence of diabetes on autonomic function in children: analysis through the geometric indices. J Hum Growth Dev 2016 Apr 28;26(1):81 [FREE Full text] [doi: 10.7322/jhgd.113728]
48. Leite D, Reis J, De BL, Pires W, Deresz L, Guimarães J. Physical inactivity is associated with reduced heart rate variability in exercising eutrophic, type 1 diabetic and obese children. J Phys Educ Sport 2019:350 [FREE Full text]
49. Macartney M, Notley S, Herry C, Seely A, Sigal R, Kenny G. Cardiac autonomic modulation in type 1 diabetes during exercise-heat stress. Acta Diabetol 2020 Aug;57(8):959-963 [FREE Full text] [doi: 10.1007/s00592-020-01505-9] [Medline: 32144491]
50. Brock C, Jessen N, Brock B, Jakobsen P, Hansen T, Rantanen J, et al. Cardiac vagal tone, a non-invasive measure of parasympathetic tone, is a clinically relevant tool in Type 1 diabetes mellitus. Diabet Med 2017 Oct;34(10):1428-1434 [FREE Full text] [doi: 10.1111/dme.13421] [Medline: 28703868]
51. Isaksen J, Graff C, Ellervik C, Jensen J, Rossing P, Kanters J, et al. Cardiac repolarization and depolarization in people with Type 1 diabetes with normal ejection fraction and without known heart disease: a case-control study. Diabet Med 2018 Oct;35(10):1337-1344 [FREE Full text] [doi: 10.1111/dme.13689] [Medline: 29797352]
52. Inanır M, Gunes Y, Sincer I, Erdal E. Evaluation of electrocardiographic ventricular depolarization and repolarization variables in type 1 diabetes mellitus. Arq Bras Cardiol 2020 Feb;114(2):275-280 [FREE Full text] [doi: 10.36660/abc.20180343] [Medline: 32215498]

53. Isaksen JL, Graff C, Ellervik C, Jensen JS, Andersen HU, Rossing P, et al. Type 1 diabetes is associated with T-wave morphology changes. The Thousand and 1 study. J Electrocardiol 2018 Dec;51(6S):72-77. [doi: 10.1016/j.jelectrocard.2018.05.015] [Medline: 30007776]
54. Moser O, Tschakert G, Mueller A, Groeschl W, Eckstein M, Koehler G, et al. Different heart rate patterns during cardio-pulmonary exercise (CPX) testing in individuals with type 1 diabetes. Front Endocrinol (Lausanne) 2018;9:585 [FREE Full text] [doi: 10.3389/fendo.2018.00585] [Medline: 30333794]
55. Bianchi L, Porta C, Rinaldi A, Gazzaruso C, Fratino P, DeCata P, et al. Integrated cardiovascular/respiratory control in type 1 diabetes evidences functional imbalance: possible role of hypoxia. Int J Cardiol 2017 Oct 01;244:254-259 [FREE Full text] [doi: 10.1016/j.ijcard.2017.06.047] [Medline: 28666602]
56. Turinese I, Marinelli P, Bonini M, Rossetti M, Statuto G, Filardi T, et al. "Metabolic and cardiovascular response to exercise in patients with type 1 diabetes". J Endocrinol Invest 2017 Sep;40(9):999-1005. [doi: 10.1007/s40618-017-0670-6] [Medline: 28386795]
57. Röhling M, Strom A, Bönhof G, Püttgen S, Bódis K, Müssig K, German Diabetes Study Group. Differential patterns of impaired cardiorespiratory fitness and cardiac autonomic dysfunction in recently diagnosed type 1 and type 2 diabetes. Diabet Care 2017 Feb;40(2):246-252. [doi: 10.2337/dc16-1898] [Medline: 27899499]
58. Mourot L, Fornasiero A, Rakobowchuk M, Skafidas S, Brighenti A, Stella F, et al. Similar cardiovascular and autonomic responses in trained type 1 diabetes mellitus and healthy participants in response to half marathon. Diabetes Res Clin Pract 2020 Feb;160:107995 [FREE Full text] [doi: 10.1016/j.diabres.2019.107995] [Medline: 31901470]
59. Francescato M, Ajčević M, Stella A, Accardo A. Modeling of carbohydrates oxidation rate during exercise in type 1 highly-trained diabetic patients. In: Proceedings of the XV Mediterranean Conference on Medical and Biological Engineering and Computing – MEDICON 2019. 2019 Presented at: XV Mediterranean Conference on Medical and Biological Engineering and Computing – MEDICON 2019; September 26-28, 2019; Coimbra, Portugal p. 559-566 URL: https://doi.org/10.1007/978-3-030-31635-8_67 [doi: 10.1007/978-3-030-31635-8_67]
60. Nascimento M, Espindola C, do Prado C, Amarins M, Potenza A, Pacheco L, et al. Type 1 diabetes does not impair the physical capacity of non-sedentary adolescents. Diabetol Metab Syndr 2017;9:100 [FREE Full text] [doi: 10.1186/s13098-017-0300-7] [Medline: 29259658]
61. Ansell S, Jester M, Tryggestad J, Short K. A pilot study of the effects of a high-intensity aerobic exercise session on heart rate variability and arterial compliance in adolescents with or without type 1 diabetes. Pediatr Diabetes 2020 May;21(3):486-495 [FREE Full text] [doi: 10.1111/pedi.12983] [Medline: 31951305]
62. Gandecka A, Araszkiewicz A, Piłaciński S, Wierusz-Wysocka B, Zozulińska-Ziółkiewicz D. Evaluation of sudomotor function in adult patients with long-lasting type 1 diabetes. Pol Arch Intern Med 2017 Jan 10;127(1):16-24 [FREE Full text] [doi: 10.20452/pamw.3884] [Medline: 28075427]
63. Fujii N, Dervis S, Sigal RJ, Kenny GP. Type 1 diabetes modulates cyclooxygenase- and nitric oxide-dependent mechanisms governing sweating but not cutaneous vasodilation during exercise in the heat. Am J Physiol Regul Integr Comp Physiol 2016 Dec 01;311(6):1076-1084 [FREE Full text] [doi: 10.1152/ajpregu.00376.2016] [Medline: 27733388]
64. Barone MT, Gonçalves B, Menna-Barreto L. Peripheral body temperature impairment in individuals with type 1 diabetes mellitus. Sleep Sci 2018;11(3):137-140 [FREE Full text] [doi: 10.5935/1984-0063.20180026] [Medline: 30455844]
65. Silva TP, Rolim LC, Filho CS, Zimmermann LM, Malerbi F, Dib SA. Association between severity of hypoglycemia and loss of heart rate variability in patients with type 1 diabetes mellitus. Diabetes Metab Res Rev 2017 Feb;33(2). [doi: 10.1002/dmrr.2830] [Medline: 27239809]
66. Olde Bekkink M, Koeneman M, de Galan BE, Bredie SJ. Early detection of hypoglycemia in type 1 diabetes using heart rate variability measured by a wearable device. Diabet Care 2019 Apr;42(4):689-692. [doi: 10.2337/dc18-1843] [Medline: 30877089]
67. Pertseva N, Tyshchenko I, Moshenets K. The influence of glycemic control on heart rate variability and blood pressure in adult patients with the type 1 diabetes. Romanian J Diabetes Nutr Metab Dis 2018:67-75 [FREE Full text] [doi: 10.2478/rjdnmd-2018-0007]
68. Koivikko ML, Kenttä T, Salmela PI, Huikuri HV, Perkiömäki JS. Changes in cardiac repolarisation during spontaneous nocturnal hypoglycaemia in subjects with type 1 diabetes: a preliminary report. Acta Diabetol 2017 Mar;54(3):251-256. [doi: 10.1007/s00592-016-0941-2] [Medline: 27933514]
69. Novodvorsky P, Bernjak A, Chow E, Iqbal A, Sellors L, Williams S, et al. Diurnal differences in risk of cardiac arrhythmias during spontaneous hypoglycemia in young people with type 1 diabetes. Diabet Care 2017 May;40(5):655-662. [doi: 10.2337/dc16-2177] [Medline: 28213374]
70. Alimova I, Demyanenko A. Functional condition of the cardiovascular system in children with type 1 diabetes depending on the level of glycemia. Ross Vestn Perinatol Pediatr 2018 Nov 20;63(5):69-74 [FREE Full text] [doi: 10.21508/1027-4065-2018-63-5-69-74]
71. Cichosz S, Frystyk J, Tarnow L, Fleischer J. Are changes in heart rate variability during hypoglycemia confounded by the presence of cardiovascular autonomic neuropathy in patients with diabetes? Diabetes Technol Ther 2017 Feb 1;19(2):91-95 [FREE Full text] [doi: 10.1089/dia.2016.0342] [Medline: 28118049]

72. Ling S, San P, Nguyen H. Non-invasive hypoglycemia monitoring system using extreme learning machine for Type 1 diabetes. ISA Trans 2016 Sep;64:440-446 [FREE Full text] [doi: 10.1016/j.isatra.2016.05.008] [Medline: 27311357]
73. San PP, Ling SH, Nguyen HT. Deep learning framework for detection of hypoglycemic episodes in children with type 1 diabetes. In: Proceedings of the 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2016 Presented at: 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Aug. 16-20, 2016; Orlando, FL, USA p. 3503-3506 URL: https://doi.org/10.1109/embc.2016.7591483 [doi: 10.1109/embc.2016.7591483]
74. Ling S, San P, Lam H, Nguyen H. Hypoglycemia detection: multiple regression-based combinational neural logic approach. Soft Comput 2015 Jul 25;21(2):543-553 [FREE Full text] [doi: 10.1007/s00500-015-1809-z]
75. Elvebakk O, Tronstad C, Birkeland K, Jenssen T, Bjørgaas MR, Frøslie KF, et al. Evaluation of hypoglycaemia with non-invasive sensors in people with type 1 diabetes and impaired awareness of hypoglycaemia. Sci Rep 2018 Oct 03;8(1):14722 [FREE Full text] [doi: 10.1038/s41598-018-33189-1] [Medline: 30283093]
76. Elvebakk O, Tronstad C, Birkeland KI, Jenssen TG, Bjørgaas MR, Gulseth HL, Martinsen. A multiparameter model for non-invasive detection of hypoglycemia. Physiol Meas 2019 Sep 03;40(8):085004 [FREE Full text] [doi: 10.1088/1361-6579/ab3676] [Medline: 31357185]
77. Reddy R, Resalat N, Wilson LM, Castle JR, El Youssef J, Jacobs PG. Prediction of hypoglycemia during aerobic exercise in adults with type 1 diabetes. J Diabetes Sci Technol 2019 Sep;13(5):919-927 [FREE Full text] [doi: 10.1177/1932296818823792] [Medline: 30650997]
78. Rodríguez-Rodríguez I, Rodríguez J, González-Vidal A, Zamora M. Feature selection for blood glucose level prediction in type 1 diabetes mellitus by using the sequential input selection algorithm (SISAL). Symmetry 2019 Sep 14;11(9):1164 [FREE Full text] [doi: 10.3390/sym11091164]
79. Hobbs N, Hajizadeh I, Rashid M, Turksoy K, Breton M, Cinar A. Improving glucose prediction accuracy in physically active adolescents with type 1 diabetes. J Diabetes Sci Technol 2019 Jul;13(4):718-727 [FREE Full text] [doi: 10.1177/1932296818820550] [Medline: 30654648]
80. Mirshekarian S, Shen H, Bunescu R, Marling C. LSTMs and neural attention models for blood glucose prediction: comparative experiments on real and synthetic data. In: Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2019 Presented at: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); July 23-27, 2019; Berlin, Germany p. 706-712 URL: https://doi.org/10.1109/embc.2019.8856940 [doi: 10.1109/embc.2019.8856940]
81. Laguna Sanz AJ, Díez JL, Giménez M, Bondia J. Enhanced Accuracy of Continuous Glucose Monitoring during Exercise through Physical Activity Tracking Integration. Sensors (Basel) 2019 Aug 30;19(17) [FREE Full text] [doi: 10.3390/s19173757] [Medline: 31480343]
82. Turksoy K, Monforti C, Park M, Griffith G, Quinn L, Cinar A. Use of wearable sensors and biometric variables in an artificial pancreas system. Sensors (Basel) 2017 Mar 07;17(3) [FREE Full text] [doi: 10.3390/s17030532] [Medline: 28272368]
83. Rothberg L, Lees T, Clifton-Bligh R, Lal S. Association between heart rate variability measures and blood glucose levels: implications for noninvasive glucose monitoring for diabetes. Diabetes Technol Ther 2016:366-376 [FREE Full text] [doi: 10.1089/dia.2016.0010]
84. Carricarte Naranjo C, Sanchez-Rodriguez LM, Martínez MB, Báez ME, García AM. Permutation entropy analysis of heart rate variability for the assessment of cardiovascular autonomic neuropathy in type 1 diabetes mellitus. Comput Biol Med 2017 Jul 01;86:90-97. [doi: 10.1016/j.compbiomed.2017.05.003] [Medline: 28527351]
85. Razanskaite-Virbickiene D, Danyte E, Mockeviciene G, Dobrovolskiene R, Verkauskiene R, Zalinkevicius R. Can coefficient of variation of time-domain analysis be valuable for detecting cardiovascular autonomic neuropathy in young patients with type 1 diabetes: a case control study. BMC Cardiovasc Disord 2017 Jan 19;17(1):34 [FREE Full text] [doi: 10.1186/s12872-016-0467-0] [Medline: 28103812]
86. Pan Q, Li Q, Deng W, Zhao D, Qi L, Huang W, et al. Prevalence and diagnosis of diabetic cardiovascular autonomic neuropathy in Beijing, China: a retrospective multicenter clinical study. Front Neurosci 2019;13:1144 [FREE Full text] [doi: 10.3389/fnins.2019.01144] [Medline: 31708736]
87. Carricarte-Naranjo C, Cornforth D, Sanchez-Rodriguez L, Brown M, Estévez M, Machado A, et al. Rényi and permutation entropy analysis for assessment of cardiac autonomic neuropathy. In: Proceedings of the European Medical and Biological Engineering Conference. 2020 Presented at: European Medical and Biological Engineering Conference; November 29 - December 3, 2020; Portorož, Slovenia p. 755-758 URL: https://doi.org/10.1007/978-981-10-5122-7_189 [doi: 10.1007/978-981-10-5122-7_189]
88. Kane P, Larsen P, Wiltshire E. Early identification of cardiac autonomic neuropathy using complexity analysis in children with type 1 diabetes. J Paediatr Child Health 2020 May;56(5):786-790. [doi: 10.1111/jpc.14745] [Medline: 31883287]
89. Silva A, Christofaro D, Vanderlei F, Barbosa M, Garner D, Vanderlei L. Association of cardiac autonomic modulation with physical and clinical features of young people with type 1 diabetes. Cardiol Young 2016 Mar 16;27(1):37-45 [FREE Full text] [doi: 10.1017/s1047951116000044]

90. Niedzwiecki P, Naskret D, Pilacinski S, Pempera M, Uruska A, Adamska A, et al. The higher the insulin resistance the lower the cardiac output in men with type 1 diabetes during the maximal exercise test. Metab Syndr Relat Disord 2017 Jun;15(5):252-257. [doi: 10.1089/met.2017.0007] [Medline: 28394194]
91. Ziegler D, Strom A, Bönhof G, Püttgen S, Bódis K, Burkart V, GDS group. Differential associations of lower cardiac vagal tone with insulin resistance and insulin secretion in recently diagnosed type 1 and type 2 diabetes. Metabolism 2018 Feb;79:1-9 [FREE Full text] [doi: 10.1016/j.metabol.2017.10.013] [Medline: 29113812]
92. Bashir M, Ahluwalia H, Sayeed S, Kapoor R. A comparative study of anxiety levels and its relation with heart rate variability (HRV) indices in adolescents with type 1 diabetes mellitus. Medeniyet Med J 2018;33 [FREE Full text] [doi: 10.5222/mmj.2018.13334]
93. Kristiansen E, Wanby P, Åkesson K, Blomstrand P, Brudin L, Thegerström J. Assessing heart rate variability in type 1 diabetes mellitus-psychosocial stress a possible confounder. Ann Noninvasive Electrocardiol 2020 Sep;25(5):e12760 [FREE Full text] [doi: 10.1111/anec.12760] [Medline: 32353221]
94. Faulkner MS, Quinn L, Fritschi C, Tripp N, Hayat MJ. Heart rate variability and cardiorespiratory fitness in non-hispanic black versus non-hispanic white adolescents with type 1 diabetes. J Cardiovasc Nurs 2019;34(5):372-379 [FREE Full text] [doi: 10.1097/JCN.0000000000000590] [Medline: 31343621]
95. Ang L, Jaiswal M, Callaghan B, Raffel D, B Brown M, Pop-Busui R. Sudomotor dysfunction as a measure of small fiber neuropathy in type 1 diabetes. Auton Neurosci 2017 Jul;205:87-92 [FREE Full text] [doi: 10.1016/j.autneu.2017.03.001] [Medline: 28325598]
96. Riguetto C, Takano C, Admoni S, Parisi M, Giannella M, Pavin E, et al. Identification and performance of multiple clinical and laboratorial risk factors for diagnosis of cardiac autonomic neuropathy in type 1 diabetes patients. J Diabetes Metab Disord 2019 Dec;18(2):565-573 [FREE Full text] [doi: 10.1007/s40200-019-00467-2] [Medline: 31890683]
97. Duvnjak L, Tomić M, Blaslov K, Rebrina S. Autonomic nervous system function assessed by conventional and spectral analysis might be useful in terms of predicting retinal deterioration in persons with type 1 diabetes mellitus. Diabetes Res Clin Pract 2016 Jun;116:111-116. [doi: 10.1016/j.diabres.2016.04.042] [Medline: 27321325]
98. Pemp B, Palkovits S, Howorka K, Pumprla J, Sacu S, Garhöfer G, et al. Correlation of retinal neurodegeneration with measures of peripheral autonomic neuropathy in type 1 diabetes. Acta Ophthalmol 2018 Nov;96(7):804-810 [FREE Full text] [doi: 10.1111/aos.13733] [Medline: 29504257]
99. Mala S, Potockova V, Hoskovcova L, Pithova P, Brabec M, Kulhankova J, et al. Cardiac autonomic neuropathy may play a role in pathogenesis of atherosclerosis in type 1 diabetes mellitus. Diabetes Res Clin Pract 2017 Dec;134:139-144 [FREE Full text] [doi: 10.1016/j.diabres.2017.10.002] [Medline: 28993155]
100. Sekercioglu N, Lovblom LE, Bjornstad P, Lovshin JA, Lytvyn Y, Boulet G, et al. Risk factors for diabetic kidney disease in adults with longstanding type 1 diabetes: results from the Canadian Study of Longevity in Diabetes. Ren Fail 2019 Nov;41(1):427-433 [FREE Full text] [doi: 10.1080/0886022X.2019.1614057] [Medline: 31162987]
101. Hansen CS, Theilade S, Lajer M, Hansen TW, Rossing P. Cardiovascular autonomic neuropathy and bone metabolism in Type 1 diabetes. Diabet Med 2018 Nov;35(11):1596-1604. [doi: 10.1111/dme.13777] [Medline: 29999549]
102. Bulum T, Hrvoje P, Duvnjak L. Heart rate and blood pressure is associated with renal function in patients with type 1 diabetes in the absence of nephropathy and therapeutical interventions. Arter Hypertens 2019:30-34 [FREE Full text] [doi: 10.5603/ah.a2019.0001]
103. Lu L, Marcovecchio ML, Dalton RN, Dunger D. Cardiovascular autonomic dysfunction predicts increasing albumin excretion in type 1 diabetes. Pediatr Diabetes 2018 May;19(3):464-469. [doi: 10.1111/pedi.12614] [Medline: 29171134]
104. Hotaling JM, Sarma AV, Patel DP, Braffett BH, Cleary PA, Feldman E, Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group. Cardiovascular autonomic neuropathy, sexual dysfunction, and urinary incontinence in women with type 1 diabetes. Diabet Care 2016 Sep;39(9):1587-1593 [FREE Full text] [doi: 10.2337/dc16-0059] [Medline: 27352953]
105. da Silva AK, Christofaro DG, Vanzella LM, Vanderlei FM, Laurino MJ, Vanderlei LC. Relationship of the aggregation of cardiovascular risk factors in the parasympathetic modulation of young people with type 1 diabetes. Medicina (Kaunas) 2019 Aug 26;55(9) [FREE Full text] [doi: 10.3390/medicina55090534] [Medline: 31454959]
106. Christensen M, Hommel E, Jørgensen ME, von Scholten BJ, Fleischer J, Hansen C. Prevalence of diabetic neuropathy in young adults with type 1 diabetes and the association with insulin pump therapy. Diabetes Technol Ther 2018 Nov 21:787-796 [FREE Full text] [doi: 10.1089/dia.2018.0249] [Medline: 30461290]
107. Kim HY, Jung HW, Lee YA, Shin CH, Yang SW. Cardiac autonomic neuropathy in nonobese young adults with type 1 diabetes. Ann Pediatr Endocrinol Metab 2019 Sep;24(3):180-186 [FREE Full text] [doi: 10.6065/apem.2019.24.3.180] [Medline: 31607111]
108. Jaiswal M, Divers J, Urbina EM, Dabelea D, Bell RA, Pettitt DJ, SEARCH for Diabetes in Youth Study Group. Cardiovascular autonomic neuropathy in adolescents and young adults with type 1 and type 2 diabetes: the SEARCH for Diabetes in Youth Cohort study. Pediatr Diabet 2018 Jun;19(4):680-689 [FREE Full text] [doi: 10.1111/pedi.12633] [Medline: 29292558]
109. Guan L, Metzger D, Lavoie P, Collet J. Glucose control and autonomic response during acute stress in youth with type 1 diabetes: a pilot study. Pediatr Diabet 2018 Aug;19(5):1020-1024 [FREE Full text] [doi: 10.1111/pedi.12680] [Medline: 29654713]

110. Stern K, Cho YH, Benitez-Aguirre P, Jenkins AJ, McGill M, Mitchell P, et al. QT interval, corrected for heart rate, is associated with HbA1c concentration and autonomic function in diabetes. Diabet Med 2016 Oct;33(10):1415-1421. [doi: 10.1111/dme.13085] [Medline: 26823095]
111. Nyiraty S, Pesei F, Orosz A, Coluzzi S, Vági OE, Lengyel C, et al. Cardiovascular autonomic neuropathy and glucose variability in patients with type 1 diabetes: is there an association? Front Endocrinol (Lausanne) 2018;9:174 [FREE Full text] [doi: 10.3389/fendo.2018.00174] [Medline: 29725320]
112. Jaiswal M, Ang L, Mizokami-Stout K, Pop-Busui R. Is there an association between non-dipping blood pressure and measures of glucose variability in type 1 diabetes? J Diabet Complicat 2018 Oct;32(10):947-950 [FREE Full text] [doi: 10.1016/j.jdiacomp.2018.07.008] [Medline: 30087063]
113. Moser O, Eckstein ML, McCarthy O, Deere R, Bain SC, Haahr HL, et al. Heart rate dynamics during cardio-pulmonary exercise testing are associated with glycemic control in individuals with type 1 diabetes. PLoS One 2018;13(4):e0194750 [FREE Full text] [doi: 10.1371/journal.pone.0194750] [Medline: 29608593]
114. Moser O, Eckstein ML, McCarthy O, Deere R, Bain SC, Haahr HL, et al. Poor glycaemic control is associated with reduced exercise performance and oxygen economy during cardio-pulmonary exercise testing in people with type 1 diabetes. Diabetol Metab Syndr 2017;9:93 [FREE Full text] [doi: 10.1186/s13098-017-0294-1] [Medline: 29201153]
115. Krishna C, Sasikala T. Home based healthcare monitoring system for diabetes patients using IoT. In: Proceedings of the International Conference on Intelligent Data Communication Technologies and Internet of Things. 2018 Presented at: International Conference on Intelligent Data Communication Technologies and Internet of Things; August 7-8, 2018; Coimbatore, India p. 676-686 URL: https://doi.org/10.1007/978-3-030-03146-6_76 [doi: 10.1007/978-3-030-03146-6_76]
116. Dubosson F, Ranvier J, Bromuri S, Calbimonte J, Ruiz J, Schumacher M. The open D1NAMO dataset: a multi-modal dataset for research on non-invasive type 1 diabetes management. Inform Med Unlocked 2018;13:92-100 [FREE Full text] [doi: 10.1016/j.imu.2018.09.003]
117. Marling C, Bunescu R. The OhioT1DM dataset for blood glucose level prediction: update 2020. CEUR Workshop Proc 2020 Sep;2675:71-74 [FREE Full text] [Medline: 33584164]

Abbreviations

AP: artificial pancreas
BG: blood glucose
CAN: cardiac autonomic neuropathy
CGM: continuous glucose monitor
ECG: electrocardiogram
ESC: electrochemical skin conductance
GSR: galvanic skin response
HbA_{1c}: hemoglobin A_{1c}
HF: high frequency
HR: heart rate
HRV: heart rate variability
IAH: impaired awareness to hypoglycemia
LF: low frequency
ML: machine learning
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QTc: QT corrected
SpO_2: oxygen saturation
ST: skin temperature
T1D: type 1 diabetes
VO_2: peak oxygen uptake

Edited by A Mavragani; submitted 05.04.21; peer-reviewed by Y Song, H Mehdizadeh; comments to author 30.07.21; revised version received 06.12.21; accepted 23.12.21; published 08.04.22.

Please cite as:
Daskalaki E, Parkinson A, Brew-Sam N, Hossain MZ, O'Neal D, Nolan CJ, Suominen H
The Potential of Current Noninvasive Wearable Technology for the Monitoring of Physiological Signals in the Management of Type 1 Diabetes: Literature Survey
J Med Internet Res 2022;24(4):e28901
URL: https://www.jmir.org/2022/4/e28901
doi:10.2196/28901
PMID:35394448

Review

Human Support in App-Based Cognitive Behavioral Therapies for Emotional Disorders: Scoping Review

Emily E Bernstein[1,2], PhD; Hilary Weingarden[1,2], PhD; Emma C Wolfe[1], BA; Margaret D Hall[1], BA; Ivar Snorrason[1,2], PhD; Sabine Wilhelm[1,2], PhD

[1]Massachusetts General Hospital, Boston, MA, United States

[2]Department of Psychiatry, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Emily E Bernstein, PhD
Massachusetts General Hospital
185 Cambridge Street
Suite 2000
Boston, MA, 02114
United States
Phone: 1 6177246032
Email: eebernstein@mgh.harvard.edu

Abstract

Background: Smartphone app–based therapies offer clear promise for reducing the gap in available mental health care for people at risk for or people with mental illness. To this end, as smartphone ownership has become widespread, app-based therapies have become increasingly common. However, the research on app-based therapies is lagging behind. In particular, although experts suggest that human support may be critical for increasing engagement and effectiveness, we have little systematic knowledge about the role that human support plays in app-based therapy. It is critical to address these open questions to optimally design and scale these interventions.

Objective: The purpose of this study is to provide a scoping review of the use of human support or coaching in app-based cognitive behavioral therapy for emotional disorders, identify critical knowledge gaps, and offer recommendations for future research. Cognitive behavioral therapy is the most well-researched treatment for a wide range of concerns and is understood to be particularly well suited to digital implementations, given its structured, skill-based approach.

Methods: We conducted systematic searches of 3 databases (PubMed, PsycINFO, and Embase). Broadly, eligible articles described a cognitive behavioral intervention delivered via smartphone app whose primary target was an emotional disorder or problem and included some level of human involvement or support (*coaching*). All records were reviewed by 2 authors. Information regarding the qualifications and training of coaches, stated purpose and content of the coaching, method and frequency of communication with users, and relationship between coaching and outcomes was recorded.

Results: Of the 2940 titles returned by the searches, 64 (2.18%) were eligible for inclusion. This review found significant heterogeneity across all of the dimensions of coaching considered as well as considerable missing information in the published articles. Moreover, few studies had qualitatively or quantitatively evaluated how the level of coaching impacts treatment engagement or outcomes. Although users tend to self-report that coaching improves their engagement and outcomes, there is limited and mixed supporting quantitative evidence at present.

Conclusions: Digital mental health is a young but rapidly expanding field with great potential to improve the reach of evidence-based care. Researchers across the reviewed articles offered numerous approaches to encouraging and guiding users. However, with the relative infancy of these treatment approaches, this review found that the field has yet to develop standards or consensus for implementing coaching protocols, let alone those for measuring and reporting on the impact. We conclude that coaching remains a significant hole in the growing digital mental health literature and lay out recommendations for future data collection, reporting, experimentation, and analysis.

(J Med Internet Res 2022;24(4):e33307) doi:10.2196/33307

KEYWORDS

digital health; mental health; cognitive behavioral therapy; coaching; guided; mobile app; emotional disorder; mobile phone

Introduction

Background

Smartphones are presently owned by 85% of the US population—a larger proportion than people who have access to computers or broadband subscriptions at home [1]. This high ownership rate represents substantial growth over the past 20 years, with rates in 2011 at just 35% [2]. Alongside this growth in smartphone ownership, there has been a corresponding proliferation in the recent development and deployment of app-based mental health treatments. In fact, in 2019, there were over 10,000 mental health apps available for download in the app market [3] with higher numbers likely available today.

There is good reason for the enthusiasm over app-based mental health treatments and skills-based approaches such as cognitive behavioral therapy (CBT) in particular that ostensibly lend themselves well to structured, standardized, self-paced platforms [4]. Smartphone-delivered therapies offer clear potential for addressing some of the most critical barriers to accessing mental health care, including prohibitive costs of treatment [5,6], patient-level logistical barriers (eg, need for time off work, transportation, and childcare) [7], and lack of access to providers who offer frontline evidence-based interventions such as CBT [8]. Indeed, in many parts of the United States there are fewer than 10 licensed psychologists per 100,000 people, with even fewer presumably offering evidence-based treatments [9]. In 2019—before the COVID-19 pandemic began—an estimated 1 in 5 adults were experiencing mental illness and even more subthreshold symptoms [10]. Prevalence rates are only increasing for younger cohorts and age groups [11,12]. This further underscores the enormous structural gap in available mental health care. Smartphones offer an opportunity to deliver impactful therapies that are readily accessible and widely scalable [13].

Although face-to-face CBT is the best studied psychosocial intervention for depression, anxiety, and related disorders, there is encouraging data showing that app-based CBT can be similarly effective [14]. Importantly, although app-based CBT offers substantial promise for addressing gaps in access to evidence-based mental health care, there are also key challenges compared with face-to-face treatment [13]. Most notably, app-based treatments often suffer from poor rates of sustained engagement, and efficacy and effectiveness studies are lagging behind app development. For example, within the IntelliCare suite of CBT skills apps, a study showed that the modal number of uses per app was once per user [15]. In an examination of engagement with the top 50 publicly available apps for depression and anxiety, more than half (63% for depression and 56% for anxiety) of the apps had no active monthly users [16]. Moreover, some engagement is likely a minimum requirement for an app to be effective. Regarding our understanding of efficacy and effectiveness, the large majority of mental health apps lack data altogether [17]. For example, in a review of available apps for anxiety and worry, 96.2% lacked efficacy data [18]. Thus, more systematic review and testing of app-based treatments is needed.

A frequent strategy advised and used by experts in digital mental health is to include human support [19,20]. However, how this has been implemented varies widely from light-touch reminders from lay support persons to in-depth, regular clinical attention and guidance from a specialized clinician. Note that in the literature, numerous titles are used to describe individuals who support patients in their use of app-based treatments, such as coach, therapist, specialist, and mentor. For clarity, the terms *coaches* and *guided* are used herein as umbrella terms to refer to human involvement in the delivery of app-based treatments. Coaches may enhance users' accountability and motivation, potentially boosting engagement with otherwise impersonal app-based treatments. In support, a recent review of engagement in digital mental health interventions found that guided interventions had higher overall engagement compared with unguided interventions [21]. Coaches may also enhance the potency of app-based CBT, by delivering some of the treatment content, helping to personalize content for individual users, correcting how users implement skills, or answering questions. A meta-analysis of app-based mental health treatment efficacy showed that apps that offered coached guidance had larger effect sizes across several efficacy outcomes [4]. Altogether, both expert opinion [22] and initial, early evidence underscore the potentially critical role that coaches may play in enhancing the value of app-based CBT.

Despite the proposed benefits of incorporating coaches within smartphone CBT, we know surprisingly little at a systematic or empirical level about coaching. For instance, there are vastly different models of coach support being implemented across app-based CBT programs. We have little knowledge about whether professional-level support is necessary or if lay person or paraprofessional support may be equally beneficial (and more cost-effective and scalable). We do not know how much support (ie, dosing) is necessary or what type of support (eg, phone calls vs messaging and user- vs clinician-initiated communication) works best. In addition, we do not know how these recommendations would vary for different age groups (eg, adolescents and older adults). Ultimately, beyond a small number of initial studies, we know little about whether coaching reliably plays a role in enhancing engagement as intended or positively impacts the effectiveness of smartphone interventions. In fact, some studies have found that external supports are associated with worse treatment outcomes, as self-contained apps may be more comprehensive in their design or users may feel compelled to be more independently responsible for working through materials [23].

Objectives

Each of these questions has direct implications for the scalability, effectiveness, and cost-effectiveness of app-based CBT. At a time when app-based CBT programs are being developed and deployed rapidly, we require a systematic, comprehensive evaluation of how coaching is currently implemented within interventions to guide the optimal design of future interventions and their scientific reporting. Given that the digital health field is increasingly moving toward app-based tools [24,25], this study centers specifically on app formats rather than collapse across internet and app-based approaches or focus on differences between them. To this end, the purpose

of this study is to provide a scoping review of available data regarding the use of coaching in app-based CBT for emotional disorders, identify critical knowledge gaps, and offer recommendations for future directions [26].

Methods

Eligibility Criteria

Given the relative novelty of this topic, we opted to survey how coaching has been defined, implemented, and evaluated to date. Studies were included in this review if they met the inclusion criteria (Textbox 1).

Textbox 1. Inclusion criteria.

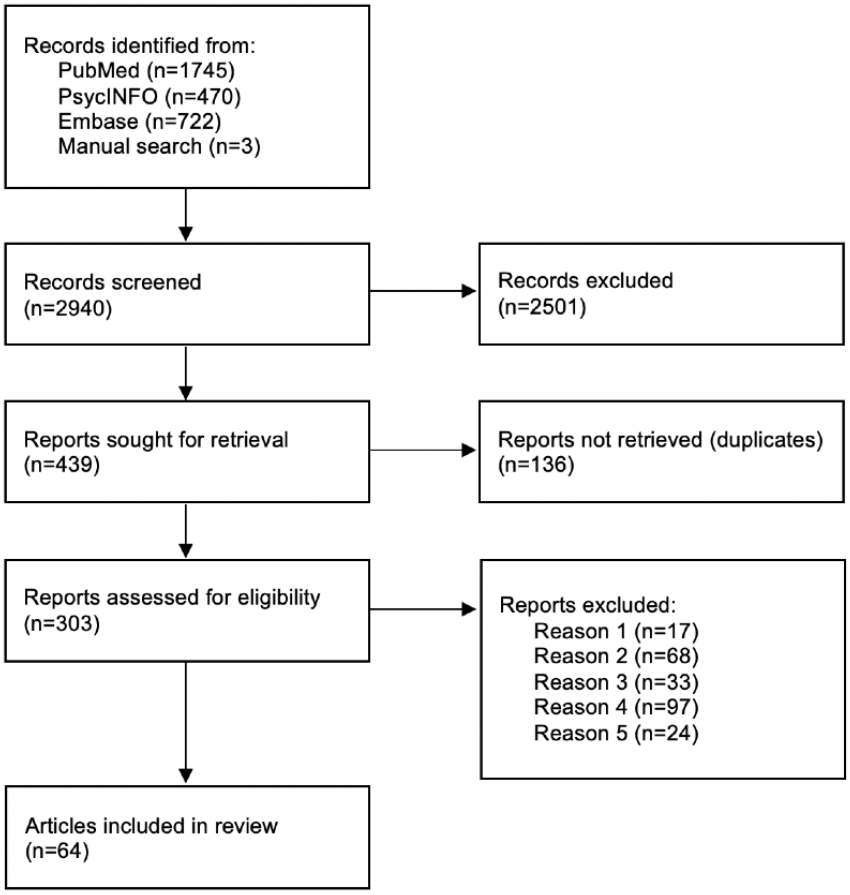
Inclusion criteria

1. Describe or report on an intervention intended to improve an emotional disorder or concern as the primary aim. Targets include depression, anxiety, stress, psychological well-being, obsessive compulsive and related disorders, posttraumatic stress disorder or posttraumatic stress symptoms, and mood. Other targets necessitating meaningfully different interventions and thus potentially more different supports, such as serious mental illness (eg, substance use, bipolar disorder, and psychosis), a primary medical condition (eg, pain and sleep), and autism spectrum disorders, were excluded for this review.
2. Describe or report on an intervention based on cognitive behavioral therapy or skills, including cognitive (eg, restructuring and core beliefs) and behavioral approaches (eg, behavioral activation, exposure, and ritual prevention) Interventions based mostly or entirely on other approaches were excluded.
3. Describe or report on a treatment delivered entirely or in part via smartphone app outside of an in-person session. Interventions delivered exclusively in person or via the internet or that used smartphone apps only for scheduling, monitoring, reminders, or ecologic momentary assessment were excluded.
4. Describe or report on some aspect of human involvement or support (eg, coach or clinician) during app-based treatment.
5. Report was published in English and as a peer-reviewed journal article (eg, dissertations or conference abstracts were excluded). Published protocols were only included if a corresponding outcomes paper had not been published. Secondary analyses were included only if new analyses regarding human involvement were reported therein.
6. Report was published before April 1, 2021.

Literature Search

To identify eligible articles, the authors conducted systematic searches in PubMed, PsycINFO, and Embase web-based databases using the following search terms in the title or abstract: *smartphone*, *mobile application*, *mobile app*, *app-based*, or *app-assisted*, in combination with *therapy*, *treatment*, *CBT*, or *iCBT*, and in combination with *depression*, *dysthymia*, *mood*, *MDD*, *anxiety*, *phobia*, *GAD*, *trauma*, *post-traumatic stress*, *posttraumatic stress*, *PTSD*, *obsessive compulsive disorder*, *obsessive-compulsive disorder*, *OCD*, *affective disorder*, *emotional disorder*, *emotional problem*, *stress*, *well-being*, or *wellness*. The database search and additional manual search (eg, searching reference sections of articles identified through database searches) occurred through April 1, 2021. In total, 2 researchers (EEB and ECW) read each title and abstract independently to screen for eligibility. In the event of disagreement, the article was included for the subsequent round of full-text review. In addition, two researchers (EEB and ECW) read the remaining articles in full, and exclusion required agreement. Uncertainties regarding inclusion at this level were discussed until consensus was reached. A flowchart summarizing this process is included (Figure 1). This report was informed by the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist [27] (Table S1 in Multimedia Appendix 1).

Figure 1. Flowchart of literature search. Reasons for exclusion: (1) primary aim or target is not an emotional disorder or concern, (2) intervention is not based on cognitive behavioral therapy or skills, (3) treatment is not delivered at least in part via smartphone app outside of an in-person session, (4) treatment does not include human involvement, and (5) published protocol or preliminary report is excluded given subsequent publication of outcomes paper.



Data Review

Given the heterogeneity of resultant data and consistent with a scoping review, we provide a narrative rather than empirical synthesis of present evidence. Eligible articles were surveyed to determine how coaching is included in studies of app-based CBT. We examined the types of personnel serving as coaches (eg, bachelor's level staff vs doctoral-level clinicians); the training, supervision, and standardization included with coaching; and the nature of the coaching itself. The latter includes the stated purpose or rationale for including a coach, the frequency and duration of coach contact, the method of contact (eg, phone call vs messaging), and the content of coaching (eg, encouragement vs teaching therapeutic content). Finally, we looked for evidence of whether coaching impacted users' experience of treatment, engagement with the app, or clinical outcomes. In this effort, we also considered the consistency and absence of reporting in these domains.

To summarize some of these data, the following new categorical variables were created. *Coach qualifications* was operationalized as the minimum allowable degree or qualification for a coach; for example, trials including bachelor's or master's level coaches were labeled *bachelor's level or above*. *Coach training* required of coaches was grouped as follows: (1) app- or study-specific training (ie, coaches underwent a seminar, workshop, or other formal training exercise to prepare them for the trial or coaches received detailed protocols or manuals to follow in their role), (2) app- or study-specific training plus ongoing supervision, (3) reliance on previous experience (ie, coaches are individuals with previous training, experience, or expertise in the therapy being delivered, such as behavioral activation or CBT for body dysmorphic disorder; no app- or study-specific training referenced), (4) previous experience plus ongoing supervision during the trial, or (5) ongoing supervision during the trial (ie, no other training or previous experience referenced). The *frequency* of coach communications was defined as daily, more than weekly (ie, 2-6 times per week), weekly, biweekly (ie, every other week), or less than biweekly (ie, less than twice per month). We summed across methods of communication; for example, if a coach made 1 phone call per week and sent 2 SMS text messages per week, the trial would be categorized as more than weekly frequency. *Method* of communication was defined as in-person (individual or group), phone (individual or group), or messaging (texting, chat, or email). We also sought to characterize *whether users were able to select the method*. If users had any autonomy in this regard (eg, users could select whether questions could be addressed via messaging or telephone), the trial would be labeled *yes*. We then labeled *who could initiate contact*: users only (eg, users reach out if questions arise or to confirm they have completed a module), coaches only (eg, communication only occurs during weekly planned phone calls), or both. We aim to characterize the *content* of

coach communications with the following four categories: (1) encouragement (ie, reminders, motivational messages, technical support, or other attempts to increase sustained engagement and adherence), (2) encouragement plus questions (ie, coaches are available to respond with clinical advice when users reach out with questions about the treatment), (3) clinical intervention (ie, coaches initiate contact with all participants with the purpose guiding treatment, such as giving feedback on skills practice, recommending or prescribing specific skills or activities, and teaching or reviewing therapeutic concepts), or (4) a full course of treatment (ie, coaches administer in-person or telephone-based treatment as usual, such as behavioral activation, exposure with ritual prevention, or school counseling). Finally, *trigger for communication* was defined as whether certain conditions would necessarily prompt coaches to message participants (ie, score on a self-report measure, indication of suicidal ideation or self-harm risk, and not using the app for a set period).

In the event of missing data, we contacted corresponding authors to request clarification. Such inquiries were sent to 66% (42/64) of corresponding authors. We received responses from 55% (23/42). In the *Results* section, we report both the data included in the published articles and supplementary data provided by the authors who responded to our inquiries. However, patterns of missing data in the published literature are noted. Given these high rates of missing information, we examined patterns by publication year to explore whether reporting has changed over time.

Results

Overview

Our scoping review yielded 64 eligible articles (listed in Table 1). Of these 64 articles, 12 (19%) are published protocols without outcome data available as of the time of the review, 24 (38%) reported on randomized controlled trials, and 19 (30%) reported on open pilot or feasibility trials. A small number of articles described case studies (5/64, 8%), quasiexperimental designs (2/64, 3%), or field trials or real-world tests (2/64, 3%). Primary treatment targets included anxiety disorder or symptoms (12/64, 19%), depressive disorder or symptoms (24/64, 38%), transdiagnostic anxiety and depression symptoms (11/64, 17%), an obsessive compulsive or related disorder (5/64, 8%), posttraumatic stress disorder (PTSD) or posttraumatic stress symptoms (PTSS; 4/64, 6%), suicidality (1/64, 2%), and general mental health (eg, stress, well-being, and quality of life; 7/54, 11%). Treatment durations ranged from 3 to 24 (mean 8.83, SD 3.88; median and mode 8) weeks. Of the 64 studies, 54 (84%) were designed for adults, and 10 (16%) included children or adolescents.

A majority of articles reported on apps that were explicitly intended to be the primary mode of treatment delivery (52/64, 81%). An additional 13% (8/64) described apps that were designed to complement in-person treatment in also providing substantive content and skills implementation. The remaining projects varied based on provider preference and patient need. Notably, only 6% (4/64) of studies include experimentally varying the inclusion of a coach [53,69,79,89], one of which was a published protocol [89].

Table 1. Apps and their availability.

App (or suite of apps) name[a]	Trials, n[b]	Citation	Still active[c]	Commercially available[c]
¡Aptívate! (Behavioral Activation Tech, LLC)	1	Dahne et al, 2019 [28]	Yes	Yes
Agoraphobia Free (Health eLiving Partnership Ltd)	1	Christoforou et al, 2017 [29]	No	No
Anxiety Coach (Mayo Clinic)	2	Whiteside et al, 2014 [30]; Whiteside et al, 2019 [31]	No (on the web only)	No
Ascend (Meru Health)	3	Goldin et al, 2019 [32]; Economides et al, 2019 [33]; Economides et al, 2020 [34]	No	No
AWAKE	1	Berg et al, 2020 [35]	No	No
Behavioral Apptivation (Behavioral Activation Tech, LLC)	1	Dahne et al, 2018 [36]	Yes	No
BiP OCD (Stockholms läns landsting)	1	Lenhard et al, 2017 [37]	No	No
Boost Me (Voyage42)	1	Stiles-Shield et al, 2019 [38]	Yes	Yes
CBT Mobile-Work	1	Callan et al, 2021 [39]	No	No
CONEMO (The Latin America Treatment & Innovation Network in Mental Health)	1	Menezes et al, 2019 [40]	No	No
DCombat	1	Giosan et al, 2017 [41]	No	No
eQuoo (PsycApps Ltd)	1	Litvin et al, 2020 [42]	Yes	Yes
EVO (Akili Interactive Labs)	1	Arean et al, 2016 [43]	Yes	Yes
Get Happy Program (Developers of the Sadness Program)	1	Watts et al, 2013 [44]	Yes	No
GET.ON	2	Ebenfeld et al, 2020 [45]; Ebenfeld et al, 2021 [46]	No	No
Happy (independent programmers; specific developers not stated in paper)	1	Otero et al, 2020 [47]	Unclear	Unclear
HARUToday (Inha Intelligent Mobile Computing Lab)	1	Ham et al, 2019 [48]	Yes	Yes (as Haru: ASD)
Helpath (CICESE-UT3)	1	Martínez-Miranda et al, 2019 [49]	Yes	Yes (Google Play only)
iCanThrive (UVA Apps, LLC)	1	Chow et al, 2020 [50]	Yes	Yes (Google Play only)
IntelliCare (suite; Adaptive Health)	5	Chen et al, 2019 [51]; Graham et al, 2020 [52]; Mohr et al, 2019 [53]; Mohr et al, 2017 [54]; Orr et al, 2020 [55]	Yes	Yes
iPST	1	Arean et al, 2016 [43]	No	No
Journey to the West (The App Happy Project)	1	Lee et al, 2014 [56]	Yes	Yes
Kokoro (Flatt Steering Committee)	2	Mantani et al, 2017 [57]; Watanabe et al, 2015 [58]	No	No
Lantern (Thrive Network, Inc)	2	Newman et al, 2021 [59]; Oser et al, 2019 [60]	No	No
Meru Health Program (Meru Health)	1	Raevuori et al, 2021 [61]	Yes	Yes
MindClimb (Optio Publishing Inc)	1	Newton et al, 2020 [62]	Yes	Yes (Google Play only)
mindLAMP (Division of Digital Psychiatry)	1	Rauseo-Ricupero and Torous, 2021 [63]	Yes	Yes
Monsenso (Monsenso A/S CVR 35517391)	1	Tønning et al, 2021 [64]	Yes	Yes
nOCD (nOCD Inc)	1	Gershkovich et al, 2021 [65]	Yes	Yes
Pacifica (now Sanvello; Sanvello Health)	1	Broglia et al, 2019 [66]	Yes	Yes

App (or suite of apps) name[a]	Trials, n[b]	Citation	Still active[c]	Commercially available[c]
Perspectives BDD (Koa Health)	1	Wilhelm et al, 2020 [13]	Yes	No
PsychAssist	1	Clough et al, 2015 [67]	No	No
PTSD Coach (US Department of Veterans Affairs)	3	Pacella-LaBarbara et al, 2020 [68]; Possemato et al, 2016 [69]; Tiet et al, 2019 [70]	Yes	Yes
RAW HAND	1	Hong et al, 2018 [71]	No	No
Run4Love (WeChat-based)	1	Guo et al, 2020 [72]	Yes	No
SmartCAT (University of Pittsburgh)	2	Silk et al, 2020 [73]; Pramana et al, 2014 [74]	Yes	Yes
SPARX (University of Auckland)	1	Werner-Seidler et al, 2020 [75]	Yes	Yes
Step-by-Step (World Health Organization)	1	Liem et al, 2020 [76]	Unclear	No
Stress Free (Thrive Therapeutic Software)	1	Christoforou et al, 2017 [29]	No	No
StressProffen (Oslo Universitetssykehus HF)	1	Børøsund et al, 2018 [77]	Yes	Yes
StudiCare Stress (Clinical Psychology and Psychotherapy Work Unit)	1	Harrer et al, 2018 [78]	No	No
Thought Challenger (part of the IntelliCare suite; Adaptive Health, Inc)	1	Stiles-Shields et al, 2019 [38]	Yes	Yes
VA apps (suite; US Department of Veterans Affairs)	1	Roy et al, 2017 [79]	Yes	Yes
Vida Health	1	Venkatesan et al, 2020 [80]	Yes	Yes
No name provided	10	Dagöö et al, 2014 [81]; Imamura et al, 2019 [82]; Ly et al, 2012 [83]; Ly et al, 2015 [84]; Ly et al, 2014 [85]; Springgate et al, 2018 [86]; Stolz et al, 2018 [87]; Uwatoko et al, 2018 [88]; Vázquez et al, 2018 [89]; Wilanksy et al, 2016 [90]	—[d]	—

[a]Developer included in parentheses where available.

[b]The numbers of trials do not add up to 64, as some articles reported on multiple apps.

[c]Data as of March 23, 2022.

[d]Not available.

Who Is Providing Coaching?

Table 2 presents a summary of personnel details and training. Coach qualifications were not initially specified in more than one-third of the trials (24/64, 38%); however, authors of 50% (12/24) of these articles provided these data by email. Coaches ranged from upper-level undergraduate students, to bachelor's level, graduate students, master's level, and doctoral level. App- or study-specific training was initially described in only one-third of articles, most of which also included ongoing supervision for coaches. Note that details of these training materials were generally low; consequently, this group likely comprises wide variability in the time and resources devoted to coach preparation and ongoing quality control, including supervision or review of coaching transcripts or tapes. A number of studies using coaches with advanced degrees relied on relevant prior training or experience with the target population or in the target treatment. More than one-third of trials (23/64, 36%) did not describe whether there was any required training, supervision, or required previous experience for coaches; authors of 48% (11/23) of these articles provided data by email. Trials using bachelor's level coaches consistently reported that coaches received explicit training and ongoing supervision. Graduate student coaches also frequently received this level of support, although it was not explicitly stated across all studies. About one-third of studies (23/64, 36%) described using coaching manuals, detailed protocols, scripts, or message or email templates to standardize at least some procedures, and authors of an additional 5% (3/64) of studies described using similar materials in email correspondence. This practice was not more or less common given the qualifications or study-specific training of coaches. Of the 64 studies, only 1 (2%) included formal fidelity checks for coaching [69].

Table 2. Coach characteristics (N=64).[a]

Characteristics	Value, n (%)
Coach qualifications	
Undergraduate student or above	1 (2)
Bachelor's degree or above	17 (27)
Graduate student or above	7 (11)
Master's level or above	17 (27)
Doctoral level	10 (16)
Not specified	12 (19)
Coach training	
Prior professional experience required (ie, no study-specific training)	13 (20)
Prior professional experience required with ongoing supervision (ie, no study-specific training)	2 (3)
Study-specific training (ie, coaches received a manual or underwent a seminar, workshop, or other formal instruction)	18 (28)
Study-specific training with ongoing supervision	16 (25)
Supervision (ie, no study-specific training or prior experience specified)	2 (3)
Not specified	13 (20)

[a]Counts reflect data included in the published articles and provided via email correspondence. Regarding coach qualifications, articles are grouped by the minimum training required for eligible coaches. For example, a trial using bachelor's level coaches as well as first-year graduate students would be classified as bachelor's degree or above.

What Is the Stated Purpose of Coaching?

Although a number of studies provided no explicit rationale for including human support, the most common themes were for bolstering user motivation, engagement, and treatment adherence. Secondarily, availability for technical and clinical support or questions was often cited. When references were included, authors frequently drew on the broad internet-based CBT literature showing that technology-based interventions often fare better with some human support, including for user engagement or clinical outcomes [91-95]. When surveying articles for the content of coach communications, coaches were focused only on bolstering motivation and engagement through reminders and general encouragement in 23% (15/64) of studies. Coaches provided this encouragement and were available as needed for answering clinical questions in an additional 19% (12/64) of studies. It should be noted that encouragement comprises a wide range of approaches, including a simple reminder or motivational messages to which users were unable to respond as well as phone calls in which coaches worked to more actively engage and motivate participants. In 36% (23/64) of studies, coaches actively provided clinical intervention, including initiating review of therapeutic content with participants, giving feedback, or assigning specific activities or homework. Again, reports generally provided few details about the clinical content, such as the types of questions or strategies users requested, the extent of feedback or recommendations, or whether coaches would go beyond the content encompassed in the app (eg, offering non-CBT strategies). Coaches in 16% (10/64) of studies were tasked with delivering full courses of treatment to complement the app content (eg, weekly group therapy, individual CBT, or other therapeutic sessions). The role of coaches in 5% (3/64) of studies was unclear (eg, described coaches *providing support*), varied significantly by coach, or was not reported.

Few studies reported using triggers for coach communications, and most descriptions were vague (eg, *signs of deterioration* without further definition). A total of 5% (3/64) of studies noted that coaches would message participants had they not logged in to the app for a certain number of days. Of the 64 studies, 6 (9%) noted coaches monitoring or receiving alerts for indication of suicidality or self-harm risk (eg, Patient Health Questionnaire–9, item 9), and 5 (8%) noted that coaches would respond to signs of deterioration (eg, increase in score of depression scale). However, additional studies (33/64, 52%) did include at least some descriptions of information coaches had access to beyond their messages, calls, or sessions with participants to support or guide their communications. For example, coaches in 23% (15/64) of studies were able to view what participants had completed within the apps (eg, content viewed and activities logged); note that the level of detail available to coaches was frequently unclear (eg, the number of activities recorded vs what the specific activities were). Coaches in 3% (2/64) of the projects were able to see metrics of app use (eg, number of log-ins); of the 64 studies, 1 (2%) was unclear, describing users' *progress*. In 13% (8/64) of studies, coaches could view results of self-report or other clinical measures. Finally, of the 64 projects, 3 (5%) allowed for access to both completed activities and clinical measures, and 3 (5%) allowed for access to both app use and clinical measures.

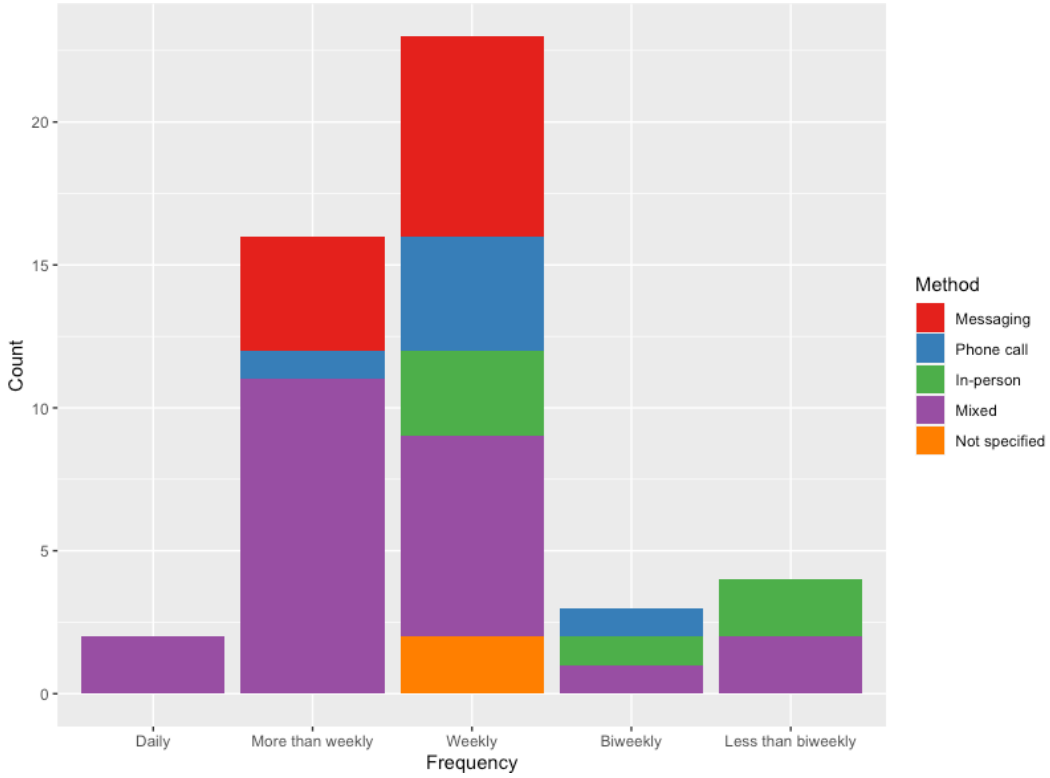
How Do Coaches Communicate With Users?

Almost one-third (42/64, 66%) of included studies used messaging (ie, texting, app-based chat or messages, or email) as a method of communication between coaches and users, and 22% (14/64) of studies used messaging as the sole method.

Phone or video calls were included in 50% (32/64) of studies and used as the sole method of communication in just 11% (7/64). In-person individual or group meetings were used in 25% (16/64) of studies and used as the sole method of communication in 11% (7/64). The most common approach was combining messaging with phone calls (22/64, 34%). Communication methods were unclear in 8% (5/64) of eligible studies. Of the 64 studies, only 9 (14%) allowed users to decide on the method of contact at least some of the time. Whereas many studies allowed patients to initiate contact (28/64, 44%), contact points were either preset or initiated only by the coach in just as many (29/64, 45%). Remaining projects did not provide sufficient or any information (7/64, 11%). A total of 11% (7/64) of projects reported on the number of messages sent by and to coaches, consistently finding that coaches typically initiate contact more than users [32,51,52,54,59,61,78].

Frequency of communication varied widely from only twice across treatment to daily messages. Given the lack of data and clarity, we collapse across intended and actual reported frequencies. Notably, a large number of studies did not report on the intended or actual frequency (19/64, 30%); however, authors of 13% (8/64) of studies clarified their procedures via email. Coaches were most often in contact at least weekly (41/64, 64%), with more frequent communication largely driven by use of messaging as the sole or most used modality. The frequency of contact broken down by method of communication is visualized (Figure 2).

Figure 2. Coaching communication patterns. Counts reflect data included in the published studies and provided via email correspondence.



Discerning patterns related to coaches' time was challenging. Given the lack of data and clarity, we collapsed across intended and actual durations. Among studies that offered at least some relevant data, weekly time per participant ranged from <10 to 60 minutes. In-person components tended to be equally distributed between standard session lengths (50-60 minutes) and 20- to 30-minute interactions. Phone calls were roughly evenly distributed among <10, <30, and 30 to 45 minutes. Longer phone calls were typically introductory contacts, and shorter calls tended to be for follow-up. Time devoted to messaging was rarely quantified; the few estimates (6/64, 9%) ranged from an average of 2.2 minutes per participant per week to upward of 30 minutes per participant per week. Unsurprisingly, coaches providing some level of clinical intervention typically spent more time per participant than coaches providing only encouragement and reminders. Notably, these patterns should not be overinterpreted, as more than half of the studies did not include any information regarding the amount of time coaches spent in total or per contact (37/64, 58%; in total, 10 authors were able to provide at least partial, additional estimates when contacted) and many more did not break down time commitments by method of communication. In addition, 27% (17/64) of studies offered no information regarding frequency or duration of coach communications; however, 11% (7/64) of studies were able to provide more information regarding one or both in follow-up communications.

What Is the Impact of Coaching?

Overview

Ultimately, only 22% (14/64) of studies considered whether the presence of a coach or the level of coaching received contributed to their intervention (Table 3). In addition, of the 64 studies, 6 (9%) included qualitative feedback from users [32,35,40,46,70,84], 5 (8%) examined the impact of coaching on engagement [38,51,53,79,80], and 7 (11%) analyzed the relationship between coaching and outcomes [35,52,53,59,69,79,80].

Table 3. Summary of studies analyzing the role of coaching in treatment engagement and outcomes.

Citation	Analysis	Finding	Sample	Treatment	App	Coach[a]	Communication[b]
Ebenfeld et al, 2021 [46]	Qualitative	Insufficient coaching as reason for dropout	92 adults, diagnosed panic disorder, majority women, White, mean age 38 (SD 10.4) years	CBT[c] for panic	GET.ON	Bachelor's level	Weekly messages
Goldin et al, 2019 [32]	Qualitative	Positive regard for coaching	2 studies, 22 and 95 adults, at least mild depressive symptoms, majority women, White, mean age 23.2 (SD 1.1) and 32 (SD 9.9) years	MBSR[d] and MCBT[e] exercises for depression	Ascend	Master's level	Messaging or phone calls 2-3 times per week
Ly et al, 2015 [84]	Qualitative	Positive regard for coaching	12 adult, diagnosed with MDD[f], 50% (6/12) women, mean age 38 (SD 14) years	BA[g] for depression	Not reported	Graduate students	At least weekly messaging
Menezes et al, 2019 [40]	Qualitative	Positive regard for coaching	66 adults, at least moderate depressive symptoms and comorbid hypertension or diabetes, majority women, aged 41-60 years	BA for depression	CONEMO	Nurse or nurse assistant	Weekly in-person meetings or phone calls
Tiet et al, 2019 [70]	Qualitative	Positive regard for coaching	29 adults, probable PTSD[h] diagnosis, majority men, White, median age 61 years	CBT skills for PTSD symptoms	PTSD Coach	Paraprofessional	Phone calls every other week
Berg et al, 2020 [35]	Qualitative and outcomes	Positive regard for coaching; mixed effects for outcomes	38 adults, cancer survivors, majority women, White, mean age 32 (SD 5.5) years	CBT for general mental health	AWAKE	Master's level	Weekly phone calls; twice-weekly texts
Mohr et al, 2019 [53]	Engagement and outcomes	Mixed effects for engagement; mixed effects for outcomes	301 adults, at least moderate depressive symptoms or mild to moderate general anxiety symptoms, majority women, White, mean age 37 (SD 12) years	CBT skills for transdiagnostic depression or anxiety	IntelliCare (suite)	Bachelor's level	Initial call; optional midtreatment call; 2-3 messages per week
Roy et al, 2017 [79]	Engagement and outcomes	Positive effects for engagement; slower symptom change	144 adults, subthreshold PTSD symptoms, majority men, White, mean age 33 (SD 11) years	CBT skills for PTSD symptoms	VA apps (suite)	Doctoral level	Introductory meeting; daily messages
Venkatesan et al, 2020 [80]	Engagement and outcomes	Positive effects for engagement; mixed effects for outcomes	323 adults, mild to moderate depressive or general anxiety symptoms, majority women, mean age 36 (SD 9) years	CBT skills for transdiagnostic depression or anxiety	Vida Health	Master's level	Weekly phone calls; messaging as needed
Chen et al, 2019 [51]	Engagement	No effects for engagement	98 adults	CBT skills for transdiagnostic depression and anxiety	IntelliCare (suite)	Bachelor's level	Initial call; 2 messages per week
Stiles-Shields et al, 2019 [38]	Engagement	No effects for engagement	30 adults, at least moderate depressive symptoms	BA or cognitive restructuring for depression	Boost Me; Thought Challenger	Master's level	Weekly phone calls or emails

Citation	Analysis	Finding	Sample	Treatment	App	Coach[a]	Communication[b]
Graham et al, 2020 [52]	Outcomes	No effects for outcomes	146 adults, at least moderate depressive or mild to moderate GAD[i] symptoms, majority women, White, mean age 42 (SD 13.8) years	CBT skills for transdiagnostic depression or anxiety	IntelliCare (suite)	Bachelor's level	Initial call; optional midtreatment call; 2 messages per week
Newman et al, 2021 [59]	Outcomes	No effects for outcomes	100 college students, self-reported GAD, majority women, White, mean age 21 years	CBT for anxiety	Lantern	Bachelor's level	Phone calls or messaging as needed
Possemato et al, 2016 [69]	Outcomes	Positive effects for outcomes	20 veterans, likely PTSD diagnosis, majority men, mean age 42 (SD 12) years	CBT skills for PTSD symptoms	PTSD Coach	Master's level	In-person meetings or phone calls every other week

[a]Coach: minimum required degree or qualification to be in the supportive human role.

[b]Frequency and method of coach contact.

[c]CBT: cognitive behavioral therapy.

[d]MBSR: mindfulness-based stress reduction.

[e]MCBT: mindfulness-based cognitive behavioral therapy.

[f]MDD: major depressive disorder.

[g]BA: behavioral activation.

[h]PTSD: posttraumatic stress disorder.

[i]GAD: generalized anxiety disorder.

Qualitative Feedback

Users largely shared positive impressions about coaching. Berg et al [35] reported on master's level coaches who engaged with users via weekly phone calls and twice-weekly SMS text messages over the course of 8 weeks. Treatment focused on promoting hope, positive mood, and behavioral goals among young adult cancer survivors. Note that emotional well-being, rather than managing physical health or related processes, was the primary treatment target. Coaches aimed to review content from the app, practice skills with users, assign homework, and offer encouragement. After the treatment, 94% of users recommended that coaching remain part of the program moving forward. Similarly, 90% of users in the study by Menezes et al [40] reported coach support to be an important treatment component. In this study, coaches were nurses or nursing assistants offering weekly phone or in-person meetings over 6 weeks to answer questions and offer encouragement for individuals with depression and comorbid hypertension or diabetes. Goldin et al [32] asked users (adults with depressive symptoms) to rate the value of coaching at weeks 1, 3, and 6 and follow-up time points. Coaches were master's level or above professionals who engaged with users via messaging or phone calls 2 to 3 times per week over 8 weeks. Coaches aimed to check in, answer questions, and facilitate a group chat among users. Users rated the value of coach interactions on average 4.13 out of 5. Tiet et al [70] used paraprofessionals (graduate students) as coaches, who offered users with likely PTSD six 5-10–minute phone calls for technical support and encouragement over 12 weeks. They found that 91% of users reported that coach support was at least somewhat helpful and 74% reported that coach's support positively impacted the frequency and consistency of their app use. In a follow-up, qualitative analysis of the primary report by Ly et al [85], researchers asked for feedback from patients with depression regarding how they perceived having up to 20 minutes of weekly messaging with a coach related to their treatment engagement and outcomes. Coaches in this study offered encouragement, general education, and weekly feedback on patients' written reflections [85]. Users shared that coaching was crucial to their app use and treatment effects, and most indicated that more frequent and more personalized contact would be preferable [84]. Both Ly et al [84] and Ebenfeld et al [46] described 1 participant each who withdrew from the trial citing insufficient direct contact with their coach. In the latter study, coaches were at least bachelor's level staff members and offered patients with panic disorder weekly feedback messages over 6 weeks [46].

Engagement

In total, 3 trials of the IntelliCare suite or its individual apps, designed to help patients with depressive or anxious symptoms, described the relationship between coaching and objective metrics of engagement [38,51,53]. Chen et al [51] examined how responsive patients were to coaches' messages. In their trial, coaches offered an initial call (30-45 minutes) and sent at least two messages per patient per week over 8 weeks to answer questions, provide recommendations and encouragement, and help patients problem solve. Responsiveness (ie, number of messages patients responded to) was ultimately unrelated to engagement with the apps, operationalized as the number of times apps were opened. In the study by Stiles-Shields et al [38], master's level coaches offered weekly encouraging phone calls or emails (<10 minutes per patient per week) over 6 weeks. Similarly, they found that neither the number nor duration of these contacts correlated with app use, including number of app launches and number of activities logged. Mohr et al [53]

manipulated whether patients had access to a coach at all. Over 8 weeks, coaches offered encouragement and answered patient questions via an initial call (30-45 minutes), optional midtreatment call (10 minutes), and 2 to 3 messages per patient per week. Patients in this coached condition did download more skills apps but did not engage in more consistent app sessions than patients who were not given coaches. When testing the VA suite of apps for patients with PTSS, Roy et al [79] found that patients who were given access to guidance from a doctoral-level coach (eg, directing users to particular skills) did self-report using the apps more frequently compared with the group that received nondirective contact, particularly for apps providing psychoeducation and tools for controlled breathing. Over 6 weeks, coaches conducted introductory in-person meetings to go over CBT skills followed by daily messages to guide app use. Finally, Venkatesan et al [80] tested the Vida Health app among adults with mild to moderate depression or anxiety. Master's level therapists provided weekly 30-minute consultations via video or phone call over 12 weeks and were available for additional in-app messaging as needed. The number of lessons or activities users completed in the app was strongly correlated with the number of consultations they completed and moderately correlated with the number of messages they sent to their therapists. Overall, across studies, there was no consistent pattern regarding the effect of coaching on app engagement. To build our understanding of the value of coaching on engagement, more studies that adopt a randomized (ie, varying presence or quantity of coaching) design are necessary. In addition, there was no consistent operationalization of engagement either, further challenging our ability to draw conclusions.

Outcomes

Outcomes data were similarly limited and mixed. A total of 3 papers reported some positive effects of coaching, 2 (67%) of which compared groups of users with and without access to a coach. First, Possemato et al [69] tested the PTSD Coach app for veterans that screened positive for a likely PTSD diagnosis. Participants in the coaching condition received biweekly in-person or phone sessions over 8 weeks in which at least master's level clinicians introduced and reviewed content and assigned homework. Compared with participants in the self-guided condition, participants with coaches demonstrated larger gains for PTSD symptoms, depression, and quality of life. In the aforementioned study of the IntelliCare suite of apps by Mohr et al [53], users with access to a coach exhibited larger declines in anxiety (General Anxiety Disorder scale–7 scores) than peers without coaches but exhibited no differences for depression (Patient Health Questionnaire–9 scores), the other primary outcome. Similarly, in the aforementioned study by Berg et al [35], the number of phone calls users completed with a coach was associated with greater reductions in days of alcohol use and improvements in pain-related functioning, but there were no differences for other primary outcomes including hope, depressive symptoms, or other domains of quality of life. In addition, in the aforementioned study by Venkatesan et al [80], therapist consultations but not messages predicted decline in depressive symptoms, and there were no reported effects for anxiety.

In contrast, 2 papers reported no effects for level of coaching; notably, however, all participants did receive a coach in these studies. Newman et al [59] offered 12 weeks of app-based CBT for anxiety to college students. Coaches were at least bachelor's level study staff and provided as-needed phone calls and messaging to support goal setting, provide encouragement and feedback, and answer questions throughout treatment. The number of messages between coaches and users was ultimately not associated with symptom change. Graham et al [52] also reported that the number of messages exchanged with a coach was not associated with symptom change. In their study, at least bachelor's level coaches provided initial phone calls, optional midtreatment phone calls, and ≥2 messages per week to users experiencing depressive or anxious symptoms over 8 weeks. Coaching focused on goal setting, making recommendations for skills to practice, and offering encouragement. Finally, the aforementioned trial by Roy et al [79] of the VA apps for PTSS found that access to a coach offering treatment guidance actually correlated with slower symptom change. Authors speculated that this effect could be due to the control condition (nondirective messages comprising nonspecific positive aphorisms) possibly being perceived as encouraging as well or the fact that the 2 groups engaged similarly with the more active skills apps (eg, social engagement and relaxation exercises). Similar to the aforementioned findings, reports yielded mixed results for coaching's effect on outcomes, and more research that is specifically designed to answer questions about the effect of coaching on outcomes is needed to clarify mixed findings.

It is notable that in research studies, participants may have contact with study staff outside of these intended supports (eg, communicating with research assistants about scheduling assessments). Such contact could be perceived as encouraging or could increase a participant's sense of accountability, thus potentially accounting for differences in observed engagement and clinical outcomes. Consequently, we also reviewed these studies for additional touch points. Studies did not report on this in depth and largely described pretreatment contact, if any. Studies that had study staff systematically reach out to participants—including for a pretreatment [35,38,69,79] or posttreatment [59] phone call or interview or to administer periodic web-based assessments during treatment [52]—did not have consistently better engagement or outcomes than those that did not report such additional contact [51,53,80].

Has Reporting Improved Over Time?

We present the proportion of papers with missing data or descriptions of various aspects of coaching by year (Table 4). Overall, time spent communicating with users (*duration*) was the most often omitted data point, followed by coach training, coach qualifications, and frequency of communications. It is difficult to determine statistically meaningful trends owing to the increasing number of publications in recent years; however, it does not appear that reporting has substantially changed over time.

Table 4. Proportions of missing data over time.[a]

Dimension	2012 (n=1), n (%)	2013 (n=1), n (%)	2014 (n=5), n (%)	2015 (n=3), n (%)	2016 (n=3), n (%)	2017 (n=6), n (%)	2018 (n=7), n (%)	2019 (n=16), n (%)	2020 (n=17), n (%)	2021 (n=5), n (%)	Total (N=64), n (%)
Qualifications	1 (100)	1 (100)	2 (40)	0 (0)	1 (33)	3 (50)	3 (43)	4 (25)	5 (29)	0 (0)	20 (31)
Type of training	1 (100)	1 (100)	2 (40)	1 (33)	1 (33)	4 (67)	2 (29)	4 (25)	4 (24)	1 (20)	21 (33)
Frequency	1 (100)	1 (100)	2 (40)	0 (0)	1 (33)	1 (17)	2 (29)	3 (19)	7 (41)	1 (20)	19 (30)
Duration	1 (100)	0 (0)	3 (60)	2 (67)	2 (67)	1 (17)	5 (71)	10 (63)	12 (71)	4 (80)	40 (63)
Method	1 (100)	0 (0)	2 (40)	0 (0)	0 (0)	0 (0)	0 (0)	1 (6)	2 (12)	0 (0)	6 (9)
Selection	1 (100)	0 (0)	2 (40)	0 (0)	0 (0)	0 (0)	0 (0)	1 (6)	1 (6)	0 (0)	5 (8)
Initiation	1 (100)	0 (0)	1 (20)	1 (33)	0 (0)	0 (0)	1 (14)	1 (6)	2 (12)	0 (0)	7 (11)
Content	1 (100)	0 (0)	1 (20)	0 (0)	0 (0)	0 (0)	1 (14)	1 (6)	2 (12)	2 (40)	8 (13)

[a]Values reflect the proportion of studies in a given year that did not report a given facet of their coaching protocol or results. For example, 40% (2/5) of the published studies from 2014 did not describe the qualifications of their coaches in the text.

What Is the Current Status of Included Apps?

Table 1 includes a summary of all apps reported on in the eligible studies as well as their current status. As of the time of review, only 26 of the 44 (59%) named apps were active (22/26, 85% available commercially).

Discussion

Principal Findings

The primary aim of this scoping review is to characterize patterns of coaching in guided app-based CBT for emotional disorders or concerns and to identify knowledge gaps. Digital mental health is a young but rapidly expanding field with enormous potential to improve the reach of evidence-based care. Researchers across the reviewed articles offered numerous approaches to encouraging and guiding users involved in these treatments (eg, prescheduled weekly phone calls to review content, daily encouraging messages, and communication as needed). Such efforts are foundational as researchers continue to improve the flexibility and accessibility of psychotherapy. However, with the relative infancy of app-based treatment approaches, this review found the field has yet to develop standards or consensus for measuring and reporting on coaching. For example, nearly half of the trials did not specify the level of training the coaches received, including their degrees, instruction, or experience with the treatment or the target population. A large number of studies also did not define the role of coaches in the intervention itself. Only a minority of studies reported on the actual number of phone calls, sessions, or messages users attended or received and even fewer the actual time commitments of coaches. In addition, little information was provided on the nature of supervision or ensuring that coaches adhered to their specified roles. Most immediately, these missing data limit our ability to identify meaningful trends in coaching for app-based treatments. From this review, it appears that no one type of coach, level of training, stated purpose of the coaching, method of coaching, frequency of communication, or duration of communication is currently the norm. This heterogeneity holds true when examining subsets of treatments, such as apps designed for specific populations (eg, children vs adults, and depression vs general mental health) or specific treatment lengths.

More consistent reporting moving forward will allow for replication of interventions across sites, a prerequisite for establishing treatment effectiveness, determining under what conditions and in what forms coaching adds clinical utility and thus amassing an evidence base for best practices. In this vein, few studies have evaluated the effects of coaching in their analyses. Qualitative data did suggest that users largely appreciate having access to a coach, describing this human element as helpful and even critical to their experience. This aligns with a number of recent reviews and meta-analyses finding that guided app-based treatments—or those with some human support—generally have better completion rates and treatment outcomes than self-guided ones [4,96,97]. However, when we then examined articles that present direct comparisons of guided versus unguided versions of the *same* app-based treatment or considered *level* of coaching within a trial, the quantitative data offered a murkier story. Although a number of studies found positive results, 1 project reported negative effects that guided treatment led to slower symptom change than unguided treatment, and half of studies reported no effects at all.

Looking again to the internet-based CBT literature, this inconsistency aligns with other reviews that have found that supported and unsupported digital interventions may not systematically differ [98,99] and that more coach time may not linearly lead to better outcomes [100]. Nevertheless, the prevailing wisdom is that the provision of human support can improve user engagement and outcomes in digital mental health [91-95]. This sentiment has been echoed in the development of app-based approaches. However, closer inspection shows that these highly cited reviews of guided versus unguided digital interventions similarly suffer from a lack of detail. *Guidance* is broadly and often unclearly defined [91], and decisions are often based on intuition or feasibility rather than evidence-based guidance or coaching protocols [101]. Heterogeneity in these reviews' conclusions about the importance of coaching could be accounted for by the heterogeneity in how authors choose to define *guidance*, CBT, and the populations studied [102].

Taken together, we cannot draw strong conclusions from this study's data, not only because of the small number of studies available but also as there are too many differences across them, including the target population, symptom severity, precise type and duration of treatment, method and frequency of coach communication, and type of coaches and coaching offered. How authors chose to quantify coaching, engagement, and outcomes was similarly variable. We conclude that coaching remains a significant hole in the rapidly growing digital mental health literature.

Future Directions

Increase Data Collection and Reporting

Thus, the first takeaway from this review is that more consistent, thorough, and standardized reporting is needed. Recommended dimensions of coaching are included as a checklist for future use in Table S2 in Multimedia Appendix 1. At minimum, the following data should be included in guided digital mental health trials moving forward (highlighted in Textbox 2): (1) Coach qualifications (if any). App-based treatments that are successfully guided by bachelor's level or paraprofessional coaches will be less expensive and more scalable than app-based treatments requiring guidance by doctoral-level coaches, underscoring the importance of these data. (2) Details regarding the materials, training, or other support provided to coaches. (3) Instructions provided to coaches regarding the purpose and boundaries of their roles and expectations set for participants in this regard. (4) The modalities and timing of coach communications—planned and actual. A coach's time is likely to be one of the most expensive parts of scaling an app-based treatment; consequently, studies should more consistently set expectations around coaches' time and measure actual time spent performing this role. (5) Adherence metric. Although feasibility may limit the thoroughness of adherence measures (eg, availability of independent raters to evaluate tapes or text), some effort should be made to assess whether coaching largely followed the intended parameters.

Textbox 2. Highlighted recommendations: key recommended dimensions to report regarding human support in app-based therapy.

Recommendations
1. Coach qualifications: what criteria were used to select eligible coaches?
2. Coach training: what written, live, or other training opportunities did coaches receive? Were coaches supervised in their work?
3. Coach instructions: what were coaches instructed to do in their role, including the purpose, nature, or boundaries of their communications with users?
4. Logistics: how, how often, and for how long were coaches in communication with users?
5. Adherence: to what extent did coaches' actual communications with users match the expected content, modalities, frequencies, and durations?

Beyond the aforementioned *must haves*, additional data, when feasible, would also be helpful: (1) differentiating communications initiated by users versus coaches; (2) user preferences for modality, frequency, timing, and content of coach communications; (3) whether the coaching that users received matched these preferences; and (4) audio recordings or transcripts of coach communications for adherence or quality ratings as well as more detailed future analysis. Notably, data from internet-based therapies reveal that coaches often deviate from guidelines, particularly for more complex interventions (eg, offering feedback, facilitating understanding, and reinforcing practice), with less engaged users, and when coaches have less specialized experience [103]. Such norms in reporting would allow for more accurate evaluation of treatments, identification of common challenges, greater comparison across trials, and dissemination into real-world settings, including cost-benefit analysis. With more data, guidelines for assessing the quality of coaching protocols would also be possible [104].

Experimentally Evaluate the Effects of Coaching

The second takeaway is that more explicit testing of coaching effects is needed. This includes additional randomized controlled trials comparing guided and unguided versions of the *same* treatment, as well as experimentally varying the type or amount of coaching users receive and exploring moderators of coaching effects. This should be done in a hypothesis-driven way as researchers consider the target population (eg, age, diagnosis, and illness severity) and the content of the app itself (eg, level of detail, personalization, and structure). Microrandomized trials could be particularly powerful for these types of questions, allowing investigators to systematically test multiple small but potentially impactful and expensive types of communication at various decision points [105]. Real-world testing will be a critical contribution to this literature as well. Human support is most frequently conceptualized as a tool for boosting adherence; individuals enrolled in clinical trials of unguided app-based treatments more often than not still receive at least some contact with study staff, which could be serving an unanticipated but similar purpose to reminders and encouragement from coaches [102]. In fact, in some of the reviewed articles, it was challenging to differentiate such clinical trial implementation processes from intended reminders or supports that were of extremely light touch [86]. Without these added contacts, outcomes could look quite different. Indeed, outside of controlled research settings, engagement data are less encouraging; for example, an analysis of the most popular apps addressing anxiety, depression, or emotional well-being—defined as the 93 apps with at least 10,000 downloads—revealed a median daily engagement rate of just 4% [106]. Interestingly, the few studies included in this review that examined the relationship between coaching and engagement or outcomes did not report significant study staff contact outside of care delivery (largely before treatment). However, scientific papers rarely describe this type of extra study contact, let alone in detail. Better tracking all forms of

intended and unintended support in the future would help address this question.

It is likely that some dimensions of coaching needs will also vary by individual user characteristics or preferences as well as other dynamic variables, such as time in treatment, status of treatment response, or how a user is using coaching (eg, reassurance seeking, problem-solving barriers to skills use, and accountability checks). For example, coaching may make the most difference for users with more severe symptoms [94] or early on in treatment [44], or the frequency or tone of communications may need to change over time to remain engaging [107]. Conversely, some dimensions may matter less. For example, research on internet-based interventions have shown that the qualifications of coaches are weak predictors of outcomes [101,108]; this review also did not identify any patterns in results based on coach qualifications.

In addition, in this review, many studies comprised small or homogeneous samples of largely White, female, and relatively young adult participants. Results could therefore be meaningfully different with larger, more diverse populations. First, age may be an important factor, although the included studies did not test this moderator. For example, quantity of coaching could be more important to older adults than to their younger counterparts. Studies have found that desire for human contact and technology literacy barriers can be deterrents for mental health app use, particularly among older adults; thus, more coaching may be beneficial for these users [55,109]. Moreover, others have found older adults to be more engaged with digital coaches than younger ones [110]. In contrast, rather than necessarily wanting more contact, adolescents may prefer different approaches; to illustrate, integrating peer support and strength-based messaging may be preferable [111]. Younger users may also be more open to virtual avatars providing guidance and support or mirroring their experience [112]. In general, more work is needed around age-appropriate messaging, for both coaches and in-app content [113,114]. Second, there is some evidence that digital interventions may reduce longstanding disparities in treatment access, response, and dropout between White and racial and ethnic minority patients [115]. Understanding how coaching could strengthen this trend as well as reach other underrepresented potential users, including sexual and gender minority individuals or older adults, should be a priority. For example, researchers may closely examine patterns of user preferences, use coaches to increase the credibility of technology-based options, support technological literacy, or actively work with individuals to adapt skills to their unique stressors or contexts [116].

Test Hypothesized Targets

Relatedly and third, consistent with broader efforts in psychotherapy research, more mechanism- or process-oriented evaluation is needed [117]. There should be testable hypotheses regarding a coach's function in app-based treatment, such as enhancing motivation, reducing a specific barrier, increasing comprehension, or supporting use of skills. In this way, the number of minutes or messages with a coach alone may not be the best metric of coaching quality, nor should end-of-study symptom severity or diagnoses be the sole outcome variables. For example, the supportive accountability model [19] and efficiency model of support [118] offer compelling frameworks for designing coaching protocols for digital interventions. Supportive accountability predicts that treatment adherence will increase if coaches are viewed as trustworthy, knowledgeable, helpful, and collaborative; in addition, coaches should increase salience and perceived utility or personal relevance of new behaviors for individual users to increase their motivation and thus engagement [19] The efficiency model argues that human support should directly address specific failure points of digital treatments (usability, engagement, fit, knowledge, and implementation) [118]. Multiple studies in this review reported implementing protocols based on these principles [38,51,53,54,78]. However, to the best of our knowledge, none tested the impact beyond the overall treatment effects. Fortunately, clear, testable targets are outlined in both models. These can be operationalized and researchers should test, for example, whether and how coaches effectively mitigate these failure points (and, perhaps, do so better than less time- and resource-intensive solutions), and whether doing so directly leads to better treatment outcomes as intended.

In this effort, researchers can capitalize on digital technologies to evaluate these variables and their interactions, including ecological momentary assessment, text analysis, machine learning, or other dynamic, idiographic approaches [119]. Leveraging these insights could lead to not only more targeted, streamlined coaching but also more effective automated support, a less expensive, more scalable alternative. Currently, although coaching ostensibly outperforms automated messages or reminders in enhancing engagement with digital therapies, the latter may still provide clinically meaningful benefits and are appreciated by users [21] and fully automated chatbots and other conversational agents are becoming more sophisticated [120]. Process-focused and individualized data of these kinds could support better algorithms dictating the timing and content of automated coaching. It remains possible that automated coaching that is based on high-quality data-driven learning and that is by nature available continuously could become as or even more effective than live coaches.

Limitations

Additional limitations of this scoping review merit noting. Because of the heterogeneity among trials and extent of missing data, a more systematic review including evaluation of the quality of studies or meta-analysis of results was not possible. The aim of this scoping review is to assess current practices and encourage new norms in reporting of guided app-based treatments to allow for this much needed next step. Relatedly, although search terms and eligibility criteria were intentionally broad and manual searches were conducted, it is also possible that some projects were missed. Present data were numerically skewed toward depression-focused treatments, a pattern consistent with other scoping and systematic reviews of mobile and app-based mental health treatments in general. It is possible that with more work on other patient populations, different or clearer trends could emerge.

Conclusions

The promise of digital treatments is scalability. Whether an app requires a bachelor's level versus doctoral-level coach or a coach at all, 2 versus 45 minutes of a coach's time per patient per week or per month, extensive training and supervision, or messaging versus face-to-face conversation to be engaging or effective would all substantially impact how scalable it can be. After reviewing the literature on guided app-based CBT, it is clear that guided interventions can be operationalized with startling diversity. Consequently, it is time to retire the notion that coaching of any kind will improve outcomes. Not only do data not consistently support this conclusion, but also the conclusion itself does not offer sufficient direction for thoughtful design and implementation of new treatments. What does seem clear is that users typically respond positively to the availability of a coach. How much ensuing interactions actually change their behavior or outcomes remains an open question. Fortunately, apps open up previously unfeasible, diverse, and creative ways to reach patients. In parallel, new technologies open up new possibilities for optimizing their impact. To take advantage of these opportunities, as a field, we must increase transparency around coaching and prioritize this treatment component as a focus in future research.

Acknowledgments

The research reported in this publication was funded by the Hope Fund (Bernstein) and National Institute of Mental Health of the National Institutes of Health under award K23MH119372 (Weingarden). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

EEB has received salary support from Koa Health. ECW has no financial conflicts of interests to disclose. MDH has no financial conflicts of interest to disclose. HW receives salary support from Koa Health (formerly Telefónica Alpha, Inc) and is a presenter for the Massachusetts General Hospital Psychiatry Academy in educational programs supported through independent medical education grants from pharmaceutical companies. IS has received salary support from Koa Health. SW is a presenter for the Massachusetts General Hospital Psychiatry Academy in educational programs supported through independent medical education grants from pharmaceutical companies; she has received royalties from Elsevier Publications, Guilford Publications, New Harbinger Publications, Springer, and Oxford University Press. SW has also received speaking honoraria from various academic institutions and foundations, including the International Obsessive Compulsive Disorder Foundation, the Tourette Association of America, and the Centers For Disease Control and Prevention. In addition, she received payment from the Association for Behavioral and Cognitive Therapies for her role as Associate Editor for the Behavior Therapy journal and John Wiley & Sons, Inc. for her role as Associate Editor of the journal Depression & Anxiety. SW has also received honoraria from One-Mind for her role on PsyberGuide Scientific Advisory Board. SW is also on the Scientific Advisory Board for Koa Health, Inc and Noom, Inc. SW has received research and salary support from Koa Health, Inc.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist and checklist for recommended reporting of human support in digital mental health treatment.
[DOCX File , 21 KB - jmir_v24i4e33307_app1.docx]

References

1. Perrin A. Mobile technology and home broadband 2021. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/2021/06/03/mobile-technology-and-home-broadband-2021/ [accessed 2007-01-21]
2. Mobile fact sheet. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2007-01-21]
3. Marshall JM, Dunstan DA, Bartik W. The digital psychiatrist: in search of evidence-based apps for anxiety and depression. Front Psychiatry 2019;10:831 [FREE Full text] [doi: 10.3389/fpsyt.2019.00831] [Medline: 31803083]
4. 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;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
5. Crow SJ, Agras WS, Halmi KA, Fairburn CG, Mitchell JE, Nyman JA. A cost effectiveness analysis of stepped care treatment for bulimia nervosa. Int J Eat Disord 2013;46(4):302-307 [FREE Full text] [doi: 10.1002/eat.22087] [Medline: 23354913]
6. McCausland J, Paparo J, Wootton BM. Treatment barriers, preferences and histories of individuals with symptoms of body dysmorphic disorder. Behav Cogn Psychother 2021;49(5):582-595. [doi: 10.1017/S1352465820000843] [Medline: 33161929]
7. Harvey AG, Gumport NB. Evidence-based psychological treatments for mental disorders: modifiable barriers to access and possible solutions. Behav Res Ther 2015;68:1-12 [FREE Full text] [doi: 10.1016/j.brat.2015.02.004] [Medline: 25768982]

8. Wolitzky-Taylor K, Fenwick K, Lengnick-Hall R, Grossman J, Bearman SK, Arch J, et al. A preliminary exploration of the barriers to delivering (and receiving) exposure-based cognitive behavioral therapy for anxiety disorders in adult community mental health settings. Community Ment Health J 2018;54(7):899-911 [FREE Full text] [doi: 10.1007/s10597-018-0252-x] [Medline: 29524078]
9. Andrilla CH, Patterson DG, Garberson LA, Coulthard C, Larson EH. Geographic variation in the supply of selected behavioral health providers. Am J Prev Med 2018;54(6 Suppl 3):S199-S207 [FREE Full text] [doi: 10.1016/j.amepre.2018.01.004] [Medline: 29779543]
10. 2019 national survey of drug use and health (NSDUH) releases. Substance Abuse and Mental Health Services Administration. 2019. URL: https://www.samhsa.gov/data/release/2019-national-survey-drug-use-and-health-nsduh-releases [accessed 2004-11-21]
11. Twenge JM. The age of anxiety? Birth cohort change in anxiety and neuroticism, 1952-1993. J Pers Soc Psychol 2000;79(6):1007-1021. [doi: 10.1037//0022-3514.79.6.1007] [Medline: 11138751]
12. Twenge JM, Cooper AB, Joiner TE, Duffy ME, Binau SG. Age, period, and cohort trends in mood disorder indicators and suicide-related outcomes in a nationally representative dataset, 2005-2017. J Abnorm Psychol 2019;128(3):185-199. [doi: 10.1037/abn0000410] [Medline: 30869927]
13. Wilhelm S, Weingarden H, Greenberg JL, McCoy TH, Ladis I, Summers BJ, et al. Development and pilot testing of a cognitive-behavioral therapy digital service for body dysmorphic disorder. Behav Ther 2020;51(1):15-26 [FREE Full text] [doi: 10.1016/j.beth.2019.03.007] [Medline: 32005333]
14. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
15. Lattie EG, Schueller SM, Sargent E, Stiles-Shields C, Tomasino KN, Corden ME, et al. Uptake and usage of IntelliCare: a publicly available suite of mental health and well-being apps. Internet Interv 2016;4(2):152-158 [FREE Full text] [doi: 10.1016/j.invent.2016.06.003] [Medline: 27398319]
16. Wasil AR, Gillespie S, Shingleton R, Wilks CR, Weisz JR. Examining the reach of smartphone apps for depression and anxiety. Am J Psychiatry 2020;177(5):464-465. [doi: 10.1176/appi.ajp.2019.19090905] [Medline: 32354266]
17. Leigh S, Flatt S. App-based psychological interventions: friend or foe? Evid Based Ment Health 2015;18(4):97-99. [doi: 10.1136/eb-2015-102203] [Medline: 26459466]
18. Sucala M, Cuijpers P, Muench F, Cardo R, Soflau R, Dobrean A, et al. Anxiety: there is an app for that. A systematic review of anxiety apps. Depress Anxiety 2017;34(6):518-525. [doi: 10.1002/da.22654] [Medline: 28504859]
19. 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;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
20. Torous J, Wisniewski H, Liu G, Keshavan M. Mental health mobile phone app usage, concerns, and benefits among psychiatric outpatients: comparative survey study. JMIR Ment Health 2018;5(4):e11715 [FREE Full text] [doi: 10.2196/11715] [Medline: 30446484]
21. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
22. 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;72(6):677-683. [doi: 10.1176/appi.ps.202000561] [Medline: 33467872]
23. 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;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
24. Marshall JM, Dunstan DA, Bartik W. The role of digital mental health resources to treat trauma symptoms in Australia during COVID-19. Psychol Trauma 2020;12(S1):S269-S271. [doi: 10.1037/tra0000627] [Medline: 32496103]
25. Stawarz K, Preist C, Coyle D. Use of smartphone apps, social media, and web-based resources to support mental health and well-being: online survey. JMIR Ment Health 2019;6(7):e12546 [FREE Full text] [doi: 10.2196/12546] [Medline: 31301126]
26. Munn Z, Peters MD, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018;18(1):143 [FREE Full text] [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
27. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
28. Dahne J, Collado A, Lejuez CW, Risco C, Diaz VA, Kustanowitz J, et al. ¡Aptívate!: a Spanish-language behavioral activation mobile application for delivery via primary care. Psychol Serv 2019;16(2):271-275 [FREE Full text] [doi: 10.1037/ser0000304] [Medline: 30431309]

29. Christoforou M, Sáez Fonseca JA, Tsakanikos E. Two novel cognitive behavioral therapy-based mobile apps for agoraphobia: randomized controlled trial. J Med Internet Res 2017;19(11):e398 [FREE Full text] [doi: 10.2196/jmir.7747] [Medline: 29175809]
30. Whiteside SP, Ale CM, Vickers Douglas K, Tiede MS, Dammann JE. Case examples of enhancing pediatric OCD treatment with a smartphone application. Clin Case Stud 2014;13(1):80-94. [doi: 10.1177/1534650113504822]
31. Whiteside SP, Biggs BK, Tiede MS, Dammann JE, Hathaway JC, Blasi ME, et al. An online- and mobile-based application to facilitate exposure for childhood anxiety disorders. Cogn Behav Pract 2019;26(3):478-491 [FREE Full text] [doi: 10.1016/j.cbpra.2019.01.002] [Medline: 33828399]
32. Goldin PR, Lindholm R, Ranta K, Hilgert O, Helteenvuori T, Raevuori A. Feasibility of a therapist-supported, mobile phone-delivered online intervention for depression: longitudinal observational study. JMIR Form Res 2019;3(1):e11509 [FREE Full text] [doi: 10.2196/11509] [Medline: 30682726]
33. Economides M, Ranta K, Nazander A, Hilgert O, Goldin PR, Raevuori A, et al. Long-term outcomes of a therapist-supported, smartphone-based intervention for elevated symptoms of depression and anxiety: quasiexperimental, pre-postintervention study. JMIR Mhealth Uhealth 2019;7(8):e14284 [FREE Full text] [doi: 10.2196/14284] [Medline: 31452521]
34. Economides M, Lehrer P, Ranta K, Nazander A, Hilgert O, Raevuori A, et al. Feasibility and efficacy of the addition of heart rate variability biofeedback to a remote digital health intervention for depression. Appl Psychophysiol Biofeedback 2020;45(2):75-86 [FREE Full text] [doi: 10.1007/s10484-020-09458-z] [Medline: 32246229]
35. Berg CJ, Vanderpool RC, Getachew B, Payne JB, Johnson MF, Sandridge Y, et al. A hope-based intervention to address disrupted goal pursuits and quality of life among young adult cancer survivors. J Cancer Educ 2020;35(6):1158-1169 [FREE Full text] [doi: 10.1007/s13187-019-01574-7] [Medline: 31297743]
36. Dahne J, Kustanowitz J, Lejuez CW. Development and preliminary feasibility study of a brief behavioral activation mobile application (behavioral apptivation) to be used in conjunction with ongoing therapy. Cogn Behav Pract 2018;25(1):44-56 [FREE Full text] [doi: 10.1016/j.cbpra.2017.05.004] [Medline: 29725198]
37. Lenhard F, Andersson E, Mataix-Cols D, Rück C, Vigerland S, Högström J, et al. Therapist-guided, internet-delivered cognitive-behavioral therapy for adolescents with obsessive-compulsive disorder: a randomized controlled trial. J Am Acad Child Adolesc Psychiatry 2017;56(1):10-9.e2 [FREE Full text] [doi: 10.1016/j.jaac.2016.09.515] [Medline: 27993223]
38. Stiles-Shields C, Montague E, Kwasny MJ, Mohr DC. Behavioral and cognitive intervention strategies delivered via coached apps for depression: pilot trial. Psychol Serv 2019;16(2):233-238 [FREE Full text] [doi: 10.1037/ser0000261] [Medline: 30407055]
39. Callan JA, Dunbar Jacob J, Siegle GJ, Dey A, Thase ME, DeVito Dabbs A, et al. CBT MobileWork©: user-centered development and testing of a mobile mental health application for depression. Cogn Ther Res 2021;45(2):287-302. [doi: 10.1007/s10608-020-10159-4]
40. Menezes P, Quayle J, Garcia Claro H, da Silva S, Brandt LR, Diez-Canseco F, et al. Use of a mobile phone app to treat depression comorbid with hypertension or diabetes: a pilot study in Brazil and Peru. JMIR Ment Health 2019;6(4):e11698 [FREE Full text] [doi: 10.2196/11698] [Medline: 31025949]
41. Giosan C, Cobeanu O, Mogoaşe C, Szentagotai A, Mureşan V, Boian R. Reducing depressive symptomatology with a smartphone app: study protocol for a randomized, placebo-controlled trial. Trials 2017;18(1):215 [FREE Full text] [doi: 10.1186/s13063-017-1960-1] [Medline: 28494802]
42. Litvin S, Saunders R, Maier MA, Lüttke S. Gamification as an approach to improve resilience and reduce attrition in mobile mental health interventions: a randomized controlled trial. PLoS One 2020;15(9):e0237220 [FREE Full text] [doi: 10.1371/journal.pone.0237220] [Medline: 32877425]
43. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The use and effectiveness of mobile apps for depression: results from a fully remote clinical trial. J Med Internet Res 2016;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]
44. Watts S, Mackenzie A, Thomas C, Griskaitis A, Mewton L, Williams A, et al. CBT for depression: a pilot RCT comparing mobile phone vs. computer. BMC Psychiatry 2013;13:49 [FREE Full text] [doi: 10.1186/1471-244X-13-49] [Medline: 23391304]
45. Ebenfeld L, Kleine Stegemann S, Lehr D, Ebert DD, Funk B, Riper H, et al. A mobile application for panic disorder and agoraphobia: insights from a multi-methods feasibility study. Internet Interv 2020;19:100296 [FREE Full text] [doi: 10.1016/j.invent.2019.100296] [Medline: 31890640]
46. Ebenfeld L, Lehr D, Ebert DD, Kleine Stegemann S, Riper H, Funk B, et al. Evaluating a hybrid web-based training program for panic disorder and agoraphobia: randomized controlled trial. J Med Internet Res 2021;23(3):e20829 [FREE Full text] [doi: 10.2196/20829] [Medline: 33661121]
47. Otero P, Hita I, Torres Á, Vázquez FL. Brief psychological intervention through mobile app and conference calls for the prevention of depression in non-professional caregivers: a pilot study. Int J Environ Res Public Health 2020;17(12):4578 [FREE Full text] [doi: 10.3390/ijerph17124578] [Medline: 32630541]
48. Ham K, Chin S, Suh YJ, Rhee M, Yu ES, Lee HJ, et al. Preliminary results from a randomized controlled study for an app-based cognitive behavioral therapy program for depression and anxiety in cancer patients. Front Psychol 2019;10:1592 [FREE Full text] [doi: 10.3389/fpsyg.2019.01592] [Medline: 31402881]

49. Martínez-Miranda J, Martínez A, Ramos R, Aguilar H, Jiménez L, Arias H, et al. Assessment of users' acceptability of a mobile-based embodied conversational agent for the prevention and detection of suicidal behaviour. J Med Syst 2019;43(8):246. [doi: 10.1007/s10916-019-1387-1] [Medline: 31240494]
50. Chow PI, Drago F, Kennedy EM, Cohn WF. A novel mobile phone app intervention with phone coaching to reduce symptoms of depression in survivors of women's cancer: pre-post pilot study. JMIR Cancer 2020;6(1):e15750 [FREE Full text] [doi: 10.2196/15750] [Medline: 32027314]
51. Chen AT, Wu S, Tomasino KN, Lattie EG, Mohr DC. A multi-faceted approach to characterizing user behavior and experience in a digital mental health intervention. J Biomed Inform 2019;94:103187 [FREE Full text] [doi: 10.1016/j.jbi.2019.103187] [Medline: 31026595]
52. Graham AK, Greene CJ, Kwasny MJ, Kaiser SM, Lieponis P, Powell T, et al. Coached mobile app platform for the treatment of depression and anxiety among primary care patients: a randomized clinical trial. JAMA Psychiatry 2020;77(9):906-914 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.1011] [Medline: 32432695]
53. Mohr DC, Schueller SM, Tomasino KN, Kaiser SM, Alam N, Karr C, et al. Comparison of the effects of coaching and receipt of app recommendations on depression, anxiety, and engagement in the IntelliCare platform: factorial randomized controlled trial. J Med Internet Res 2019;21(8):e13609 [FREE Full text] [doi: 10.2196/13609] [Medline: 31464192]
54. Mohr DC, Tomasino KN, Lattie EG, Palac HL, Kwasny MJ, Weingardt K, et al. IntelliCare: an eclectic, skills-based app suite for the treatment of depression and anxiety. J Med Internet Res 2017;19(1):e10 [FREE Full text] [doi: 10.2196/jmir.6645] [Medline: 28057609]
55. Orr LC, Graham AK, Mohr DC, Greene CJ. Engagement and clinical improvement among older adult primary care patients using a mobile intervention for depression and anxiety: case studies. JMIR Ment Health 2020;7(7):e16341 [FREE Full text] [doi: 10.2196/16341] [Medline: 32673236]
56. Lee MD, Kang X, Hanrahan N. Addressing cultural contexts in the management of stress via narrative and mobile technology. Stud Health Technol Inform 2014;199:173-177. [Medline: 24875715]
57. Mantani A, Kato T, Furukawa TA, Horikoshi M, Imai H, Hiroe T, et al. Smartphone cognitive behavioral therapy as an adjunct to pharmacotherapy for refractory depression: randomized controlled trial. J Med Internet Res 2017;19(11):e373 [FREE Full text] [doi: 10.2196/jmir.8602] [Medline: 29101095]
58. Watanabe N, Horikoshi M, Yamada M, Shimodera S, Akechi T, Miki K, Steering Committee of the Fun to Learn to Act and Think through Technology Project. Adding smartphone-based cognitive-behavior therapy to pharmacotherapy for major depression (FLATT project): study protocol for a randomized controlled trial. Trials 2015;16:293 [FREE Full text] [doi: 10.1186/s13063-015-0805-z] [Medline: 26149441]
59. Newman MG, Jacobson NC, Rackoff GN, Bell MJ, Taylor CB. A randomized controlled trial of a smartphone-based application for the treatment of anxiety. Psychother Res 2021;31(4):443-454. [doi: 10.1080/10503307.2020.1790688] [Medline: 32662323]
60. Oser M, Wallace ML, Solano F, Szigethy EM. Guided digital cognitive behavioral program for anxiety in primary care: propensity-matched controlled trial. JMIR Ment Health 2019;6(4):e11981 [FREE Full text] [doi: 10.2196/11981] [Medline: 30946022]
61. Raevuori A, Vahlberg T, Korhonen T, Hilgert O, Aittakumpu-Hyden R, Forman-Hoffman V. A therapist-guided smartphone app for major depression in young adults: a randomized clinical trial. J Affect Disord 2021;286:228-238 [FREE Full text] [doi: 10.1016/j.jad.2021.02.007] [Medline: 33743385]
62. Newton A, Bagnell A, Rosychuk R, Duguay J, Wozney L, Huguet A, et al. A mobile phone-based app for use during cognitive behavioral therapy for adolescents with anxiety (MindClimb): user-centered design and usability study. JMIR Mhealth Uhealth 2020;8(12):e18439 [FREE Full text] [doi: 10.2196/18439] [Medline: 33289671]
63. Rauseo-Ricupero N, Torous J. Technology enabled clinical care (TECC): protocol for a prospective longitudinal cohort study of smartphone-augmented mental health treatment. JMIR Res Protoc 2021;10(1):e23771 [FREE Full text] [doi: 10.2196/23771] [Medline: 33296869]
64. Tønning ML, Faurholt-Jepsen M, Frost M, Martiny K, Tuxen N, Rosenberg N, et al. The effect of smartphone-based monitoring and treatment on the rate and duration of psychiatric readmission in patients with unipolar depressive disorder: the RADMIS randomized controlled trial. J Affect Disord 2021;282:354-363. [doi: 10.1016/j.jad.2020.12.141] [Medline: 33421863]
65. Gershkovich M, Middleton R, Hezel DM, Grimaldi S, Renna M, Basaraba C, et al. Integrating exposure and response prevention with a mobile app to treat obsessive-compulsive disorder: feasibility, acceptability, and preliminary effects. Behav Ther 2021;52(2):394-405. [doi: 10.1016/j.beth.2020.05.001] [Medline: 33622508]
66. Broglia E, Millings A, Barkham M. Counseling with guided use of a mobile well-being app for students experiencing anxiety or depression: clinical outcomes of a feasibility trial embedded in a student counseling service. JMIR Mhealth Uhealth 2019;7(8):e14318 [FREE Full text] [doi: 10.2196/14318] [Medline: 31418424]
67. Clough BA, Casey LM. Therapy on the move: the development of a therapeutic smartphone application. Int J Cyber Behav Psychol Learn 2015;5(1):3-41. [doi: 10.4018/ijcbpl.2015010103]

68. Pacella-LaBarbara ML, Suffoletto BP, Kuhn E, Germain A, Jaramillo S, Repine M, et al. A pilot randomized controlled trial of the PTSD coach app following motor vehicle crash-related injury. Acad Emerg Med 2020;27(11):1126-1139 [FREE Full text] [doi: 10.1111/acem.14000] [Medline: 32339359]
69. Possemato K, Kuhn E, Johnson E, Hoffman JE, Owen JE, Kanuri N, et al. Using PTSD coach in primary care with and without clinician support: a pilot randomized controlled trial. Gen Hosp Psychiatry 2016;38:94-98. [doi: 10.1016/j.genhosppsych.2015.09.005] [Medline: 26589765]
70. Tiet QQ, Duong H, Davis L, French R, Smith CL, Leyva YE, et al. PTSD coach mobile application with brief telephone support: a pilot study. Psychol Serv 2019;16(2):227-232. [doi: 10.1037/ser0000245] [Medline: 30407056]
71. Hong JS, Kim SM, Aboujaoude E, Han DH. Investigation of a mobile "serious game" in the treatment of obsessive-compulsive disorder: a pilot study. Games Health J 2018;7(5):317-326. [doi: 10.1089/g4h.2017.0158] [Medline: 30129775]
72. 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;22(2):e16715 [FREE Full text] [doi: 10.2196/16715] [Medline: 32044751]
73. Silk JS, Pramana G, Sequeira SL, Lindhiem O, Kendall PC, Rosen D, et al. Using a smartphone app and clinician portal to enhance brief cognitive behavioral therapy for childhood anxiety disorders. Behav Ther 2020;51(1):69-84 [FREE Full text] [doi: 10.1016/j.beth.2019.05.002] [Medline: 32005341]
74. Pramana G, Parmanto B, Kendall PC, Silk JS. The SmartCAT: an m-health platform for ecological momentary intervention in child anxiety treatment. Telemed J E Health 2014;20(5):419-427 [FREE Full text] [doi: 10.1089/tmj.2013.0214] [Medline: 24579913]
75. Werner-Seidler A, Huckvale K, Larsen ME, Calear AL, Maston K, Johnston L, et al. A trial protocol for the effectiveness of digital interventions for preventing depression in adolescents: the future proofing study. Trials 2020;21(1):2 [FREE Full text] [doi: 10.1186/s13063-019-3901-7] [Medline: 31898512]
76. 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;1:96 [FREE Full text] [doi: 10.1186/s43058-020-00072-y] [Medline: 33145495]
77. Børøsund E, Mirkovic J, Clark MM, Ehlers SL, Andrykowski MA, Bergland A, et al. A stress management app intervention for cancer survivors: design, development, and usability testing. JMIR Form Res 2018;2(2):e19 [FREE Full text] [doi: 10.2196/formative.9954] [Medline: 30684438]
78. Harrer M, Adam SH, Fleischmann RJ, Baumeister H, Auerbach R, Bruffaerts R, et al. Effectiveness of an internet- and app-based intervention for college students with elevated stress: randomized controlled trial. J Med Internet Res 2018;20(4):e136 [FREE Full text] [doi: 10.2196/jmir.9293] [Medline: 29685870]
79. Roy MJ, Costanzo ME, Highland KB, Olsen C, Clayborne D, Law W. An app a day keeps the doctor away: guided education and training via smartphones in subthreshold post traumatic stress disorder. Cyberpsychol Behav Soc Netw 2017;20(8):470-478. [doi: 10.1089/cyber.2017.0221] [Medline: 28737954]
80. Venkatesan A, Rahimi L, Kaur M, Mosunic C. Digital cognitive behavior therapy intervention for depression and anxiety: retrospective study. JMIR Ment Health 2020;7(8):e21304 [FREE Full text] [doi: 10.2196/21304] [Medline: 32845246]
81. Dagöö J, Asplund RP, Bsenko HA, Hjerling S, Holmberg A, Westh S, et al. Cognitive behavior therapy versus interpersonal psychotherapy for social anxiety disorder delivered via smartphone and computer: a randomized controlled trial. J Anxiety Disord 2014;28(4):410-417. [doi: 10.1016/j.janxdis.2014.02.003] [Medline: 24731441]
82. Imamura K, Tran TT, Nguyen HT, Kuribayashi K, Sakuraya A, Nguyen AQ, et al. Effects of two types of smartphone-based stress management programmes on depressive and anxiety symptoms among hospital nurses in Vietnam: a protocol for three-arm randomised controlled trial. BMJ Open 2019;9(4):e025138 [FREE Full text] [doi: 10.1136/bmjopen-2018-025138] [Medline: 30962230]
83. Ly KH, Carlbring P, Andersson G. Behavioral activation-based guided self-help treatment administered through a smartphone application: study protocol for a randomized controlled trial. Trials 2012;13:62 [FREE Full text] [doi: 10.1186/1745-6215-13-62] [Medline: 22607302]
84. Ly KH, Janni E, Wrede R, Sedem M, Donker T, Carlbring P, et al. Experiences of a guided smartphone-based behavioral activation therapy for depression: a qualitative study. Internet Interv 2015;2(1):60-68. [doi: 10.1016/j.invent.2014.12.002]
85. Ly KH, Trüschel A, Jarl L, Magnusson S, Windahl T, Johansson R, et al. Behavioural activation versus mindfulness-based guided self-help treatment administered through a smartphone application: a randomised controlled trial. BMJ Open 2014;4(1):e003440 [FREE Full text] [doi: 10.1136/bmjopen-2013-003440] [Medline: 24413342]
86. Springgate BF, Arevian AC, Wennerstrom A, Johnson AJ, Eisenman DP, Sugarman OK, et al. Community resilience learning collaborative and research network (C-LEARN): study protocol with participatory planning for a randomized, comparative effectiveness trial. Int J Environ Res Public Health 2018;15(8):1683 [FREE Full text] [doi: 10.3390/ijerph15081683] [Medline: 30720791]
87. Stolz T, Schulz A, Krieger T, Vincent A, Urech A, Moser C, et al. A mobile app for social anxiety disorder: a three-arm randomized controlled trial comparing mobile and PC-based guided self-help interventions. J Consult Clin Psychol 2018;86(6):493-504. [doi: 10.1037/ccp0000301] [Medline: 29781648]

88. Uwatoko T, Luo Y, Sakata M, Kobayashi D, Sakagami Y, Takemoto K, et al. Healthy campus trial: a multiphase optimization strategy (MOST) fully factorial trial to optimize the smartphone cognitive behavioral therapy (CBT) app for mental health promotion among university students: study protocol for a randomized controlled trial. Trials 2018;19(1):353 [FREE Full text] [doi: 10.1186/s13063-018-2719-z] [Medline: 29973252]
89. Vázquez FL, Torres Á, Díaz O, Páramo M, Otero P, Blanco V, et al. Cognitive behavioral intervention via a smartphone app for non-professional caregivers with depressive symptoms: study protocol for a randomized controlled trial. Trials 2018;19(1):414 [FREE Full text] [doi: 10.1186/s13063-018-2793-2] [Medline: 30064466]
90. Wilansky P, Eklund JM, Milner T, Kreindler D, Cheung A, Kovacs T, et al. Cognitive behavior therapy for anxious and depressed youth: improving homework adherence through mobile technology. JMIR Res Protoc 2016;5(4):e209 [FREE Full text] [doi: 10.2196/resprot.5841] [Medline: 27833071]
91. Andersson G, Carlbring P, Ljótsson B, Hedman E. Guided internet-based CBT for common mental disorders. J Contemp Psychother 2013;43(4):223-233. [doi: 10.1007/s10879-013-9237-9]
92. 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]
93. Johansson R, Andersson G. Internet-based psychological treatments for depression. Expert Rev Neurother 2012;12(7):861-870. [doi: 10.1586/ern.12.63] [Medline: 22853793]
94. 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;78(4):361-371 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.4364] [Medline: 33471111]
95. Newman MG, Szkodny LE, Llera SJ, Przeworski A. A review of technology-assisted self-help and minimal contact therapies for anxiety and depression: is human contact necessary for therapeutic efficacy? Clin Psychol Rev 2011;31(1):89-103. [doi: 10.1016/j.cpr.2010.09.008] [Medline: 21130939]
96. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. J Affect Disord 2020;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
97. Wu A, Scult MA, Barnes ED, Betancourt JA, Falk A, Gunning FM. Smartphone apps for depression and anxiety: a systematic review and meta-analysis of techniques to increase engagement. NPJ Digit Med 2021;4(1):20 [FREE Full text] [doi: 10.1038/s41746-021-00386-8] [Medline: 33574573]
98. Mohr DC, Duffecy J, Ho J, Kwasny M, Cai X, Burns MN, et al. A randomized controlled trial evaluating a manualized TeleCoaching protocol for improving adherence to a web-based intervention for the treatment of depression. PLoS One 2013;8(8):e70086 [FREE Full text] [doi: 10.1371/journal.pone.0070086] [Medline: 23990896]
99. Titov N, Dear BF, Ali S, Zou JB, Lorian CN, Johnston L, et al. Clinical and cost-effectiveness of therapist-guided internet-delivered cognitive behavior therapy for older adults with symptoms of depression: a randomized controlled trial. Behav Ther 2015;46(2):193-205. [doi: 10.1016/j.beth.2014.09.008] [Medline: 25645168]
100. Titov N, Dear BF, Schwencke G, Andrews G, Johnston L, Craske MG, et al. Transdiagnostic internet treatment for anxiety and depression: a randomised controlled trial. Behav Res Ther 2011;49(8):441-452. [doi: 10.1016/j.brat.2011.03.007] [Medline: 21679925]
101. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on internet-based mental health interventions — a systematic review. Internet Interv 2014;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
102. Ahern E, Kinsella S, Semkovska M. Clinical efficacy and economic evaluation of online cognitive behavioral therapy for major depressive disorder: a systematic review and meta-analysis. Expert Rev Pharmacoecon Outcomes Res 2018;18(1):25-41. [doi: 10.1080/14737167.2018.1407245] [Medline: 29145746]
103. Hadjistavropoulos HD, Schneider LH, Klassen K, Dear BF, Titov N. Development and evaluation of a scale assessing therapist fidelity to guidelines for delivering therapist-assisted internet-delivered cognitive behaviour therapy. Cogn Behav Ther 2018;47(6):447-461. [doi: 10.1080/16506073.2018.1457079] [Medline: 29693533]
104. Kocsis JH, Gerber AJ, Milrod B, Roose SP, Barber J, Thase ME, et al. A new scale for assessing the quality of randomized clinical trials of psychotherapy. Compr Psychiatry 2010;51(3):319-324. [doi: 10.1016/j.comppsych.2009.07.001] [Medline: 20399343]
105. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: an experimental design for developing just-in-time adaptive interventions. Health Psychol 2015;34S:1220-1228 [FREE Full text] [doi: 10.1037/hea0000305] [Medline: 26651463]
106. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
107. Peiris D, Wright L, News M, Rogers K, Redfern J, Chow C, et al. A smartphone app to assist smoking cessation among aboriginal Australians: findings from a pilot randomized controlled trial. JMIR Mhealth Uhealth 2019;7(4):e12745 [FREE Full text] [doi: 10.2196/12745] [Medline: 30938691]

108. Johnston L, Titov N, Andrews G, Spence J, Dear BF. A RCT of a transdiagnostic internet-delivered treatment for three anxiety disorders: examination of support roles and disorder-specific outcomes. PLoS One 2011;6(11):e28079 [FREE Full text] [doi: 10.1371/journal.pone.0028079] [Medline: 22132216]
109. Wildenbos GA, Jaspers MW, Schijven MP, Dusseljee-Peute LW. Mobile health for older adult patients: using an aging barriers framework to classify usability problems. Int J Med Inform 2019;124:68-77. [doi: 10.1016/j.ijmedinf.2019.01.006] [Medline: 30784429]
110. Graham SA, Stein N, Shemaj F, Branch OH, Paruthi J, Kanick SC. Older adults engage with personalized digital coaching programs at rates that exceed those of younger adults. Front Digit Health 2021;3:642818 [FREE Full text] [doi: 10.3389/fdgth.2021.642818] [Medline: 34713112]
111. Aschbrenner KA, Naslund JA, Tomlinson EF, Kinney A, Pratt SI, Brunette MF. Adolescents' use of digital technologies and preferences for mobile health coaching in public mental health settings. Front Public Health 2019;7:178 [FREE Full text] [doi: 10.3389/fpubh.2019.00178] [Medline: 31312629]
112. LeRouge C, Dickhut K, Lisetti C, Sangameswaran S, Malasanos T. Engaging adolescents in a computer-based weight management program: avatars and virtual coaches could help. J Am Med Inform Assoc 2016;23(1):19-28 [FREE Full text] [doi: 10.1093/jamia/ocv078] [Medline: 26159466]
113. Farrer L, Leach L, Griffiths KM, Christensen H, Jorm AF. Age differences in mental health literacy. BMC Public Health 2008;8:125 [FREE Full text] [doi: 10.1186/1471-2458-8-125] [Medline: 18423049]
114. Wetherell JL, Petkus AJ, McChesney K, Stein MB, Judd PH, Rockwell E, et al. Older adults are less accurate than younger adults at identifying symptoms of anxiety and depression. J Nerv Ment Dis 2009;197(8):623-626 [FREE Full text] [doi: 10.1097/NMD.0b013e3181b0c081] [Medline: 19684501]
115. Ramos G, Chavira DA. Use of technology to provide mental health care for racial and ethnic minorities: evidence, promise, and challenges. Cogn Behav Pract 2022;29(1):15-40. [doi: 10.1016/j.cbpra.2019.10.004]
116. Rozbroj T, Lyons A, Pitts M, Mitchell A, Christensen H. Assessing the applicability of e-therapies for depression, anxiety, and other mood disorders among lesbians and gay men: analysis of 24 web- and mobile phone-based self-help interventions. J Med Internet Res 2014;16(7):e166 [FREE Full text] [doi: 10.2196/jmir.3529] [Medline: 24996000]
117. Hofmann SG, Hayes SC. The future of intervention science: process-based therapy. Clin Psychol Sci 2019;7(1):37-50 [FREE Full text] [doi: 10.1177/2167702618772296] [Medline: 30713811]
118. Schueller SM, Tomasino KN, Mohr DC. Integrating human support into behavioral intervention technologies: the efficiency model of support. Clin Psychol Sci Pract 2017;24(1):27-45. [doi: 10.1111/cpsp.12173]
119. Hayes SC, Hofmann SG, Stanton CE, Carpenter JK, Sanford BT, Curtiss JE, et al. The role of the individual in the coming era of process-based therapy. Behav Res Ther 2019;117:40-53. [doi: 10.1016/j.brat.2018.10.005] [Medline: 30348451]
120. Ly KH, Ly AM, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. Internet Interv 2017;10:39-46 [FREE Full text] [doi: 10.1016/j.invent.2017.10.002] [Medline: 30135751]

Abbreviations

CBT: cognitive behavioral therapy
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
PTSD: posttraumatic stress disorder
PTSS: posttraumatic stress symptoms

Edited by A Mavragani; submitted 01.09.21; peer-reviewed by T Bose, M Kraepelien, J Beames; comments to author 28.11.21; revised version received 18.01.22; accepted 31.01.22; published 08.04.22.

Please cite as:
Bernstein EE, Weingarden H, Wolfe EC, Hall MD, Snorrason I, Wilhelm S
Human Support in App-Based Cognitive Behavioral Therapies for Emotional Disorders: Scoping Review
J Med Internet Res 2022;24(4):e33307
URL: https://www.jmir.org/2022/4/e33307
doi:10.2196/33307
PMID:35394434

Review

Effectiveness of Digital Interventions for Preventing Alcohol Consumption in Pregnancy: Systematic Review and Meta-analysis

Sarah Soyeon Oh[1,2,3], PhD; Jong Youn Moon[4,5*], MD, PhD; Doukyoung Chon[3], PhD; Carol Mita[6], MLIS; Jourdyn A Lawrence[7], MScPH, PhD; Eun-Cheol Park[2,8], MD, PhD; Ichiro Kawachi[1*], MD, PhD

[1]Department of Social & Behavioral Sciences, Harvard TH Chan School of Public Health, Boston, MA, United States

[2]Institute of Health Services Research, Yonsei University College of Medicine, Seoul, Republic of Korea

[3]Center for Public Health, Gil Medical Center, Gachon University College of Medicine, Incheon, Republic of Korea

[4]Department of Preventive Medicine, Gachon University College of Medicine, Incheon, Republic of Korea

[5]Artificial Intelligence and Big-Data Convergence Center, Gil Medical Center, Gachon University College of Medicine, Incheon, Republic of Korea

[6]Research and Instruction Services, Countway Library of Medicine, Harvard Medical School, Boston, MA, United States

[7]François-Xavier Bagnoud Center for Health and Human Rights, Harvard University, Boston, MA, United States

[8]Department of Preventive Medicine & Public Health, Yonsei University College of Medicine, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Jong Youn Moon, MD, PhD
Department of Preventive Medicine
Gachon University College of Medicine
191 Hambangmoe-ro, Yeonsu-gu
Incheon, 21936
Republic of Korea
Phone: 82 8572647167
Email: moonjy@gachon.ac.kr

Abstract

Background: Alcohol consumption in pregnancy has been associated with serious fetal health risks and maternal complications. While previous systematic reviews of digital interventions during pregnancy have targeted smoking cessation and flu vaccine uptake, few studies have sought to evaluate their effectiveness in preventing alcohol consumption during pregnancy.

Objective: This systematic review aims to assess (1) whether digital interventions are effective in preventing alcohol consumption during the pregnancy/pregnancy-planning period, and (2) the differential effectiveness of alternative digital intervention platforms (ie, computers, mobiles, and text messaging services).

Methods: PubMed, Embase, CINAHL, and Web of Science were searched for studies with digital interventions aiming to prevent alcohol consumption among pregnant women or women planning to become pregnant. A random effects primary meta-analysis was conducted to estimate the combined effect size and extent to which different digital platforms were successful in preventing alcohol consumption in pregnancy.

Results: Six studies were identified and included in the final review. The primary meta-analysis produced a sample-weighted odds ratio (OR) of 0.62 (95% CI 0.42-0.91; *P*=.02) in favor of digital interventions decreasing the risk of alcohol consumption during pregnancy when compared to controls. Computer/internet-based interventions (OR 0.59, 95% CI 0.38-0.93) were an effective platform for preventing alcohol consumption. Too few studies of text messaging (OR 0.29, 95% CI 0.29-2.52) were available to draw a conclusion.

Conclusions: Overall, our review highlights the potential for digital interventions to prevent alcohol consumption among pregnant women and women planning to become pregnant. Considering the advantages of digital interventions in promoting healthy behavioral changes, future research is necessary to understand how certain platforms may increase user engagement and intervention effectiveness to prevent women from consuming alcohol during their pregnancies.

(J Med Internet Res 2022;24(4):e35554) doi:10.2196/35554

KEYWORDS

fetal alcohol spectrum disorders; fetal alcohol syndrome; digital health; pregnancy; alcohol consumption; text message; text messaging; alcohol; digital intervention; mother; systematic review; meta-analysis; mobile health; mHealth; computer-based intervention; internet-based intervention

Introduction

Background

Alcohol consumption during pregnancy is a major public health concern, and it has explicit links to fetal alcohol spectrum disorders (FASDs) and adverse birth-related outcomes like miscarriage and stillbirth [1]. Yet globally, 9.8% of women are estimated to consume alcohol during pregnancy, resulting in more than 630,000 babies being born each year with life-long neurodevelopmental abnormalities and central nervous system damage, and this makes FASDs the most common preventable form of developmental disability in the Western world [2]. In the United States, around 1% (9.1 cases per 1000 live births) of all babies are born with alcohol-related birth defects [3]. Socioeconomic costs pertaining to health care, special education, disability-adjusted life years, and premature mortality are believed to be more than US $24,000 per individual, which exceed the costs for autism and asthma by 26% and 87%, respectively [1].

Barriers to alcohol abstinence during pregnancy range from lack of awareness about health consequences to low socioeconomic status and/or ability to access necessary health care services [4]. According to a report by the New Zealand Ministry of Health, while 91% of mothers-to-be reduce their alcohol intake upon learning about their pregnancy, more than half only do so after their pregnancy has commenced [4]. Furthermore, many pregnant women who drink throughout all 3 trimesters may have a history of trauma or violence, physical health concerns, lack of mental health support, and/or fear of accessing health care services due to social stigmatization [5].

Social inequalities are also a fundamental risk factor for alcohol consumption during pregnancy, with women of low socioeconomic status and racial/ethnic minority backgrounds at greater risk of bearing children with severe forms of FASDs like fetal alcohol syndrome [6]. Alarming statistics have reported that certain indigenous communities in British Columbia (190 cases per 1000 live births) and the Manitoba First Nations reserve (55-101 cases per 1000 live births), for example, have a significantly higher proportion of children with FASDs than the general population [7]. Population-based studies of FASDs in South Africa have shown that women living in poor rural farms where living conditions are the poorest and binge drinking is a regular practice, have the greatest odds of bearing children with FASDs [8].

With digital technologies having considerable potential to deliver health care interventions at a low cost and with easy accessibility [9], innovative approaches in the field of preventive and personalized medicine are targeting pregnant women. Lifestyle change interventions empowering women and men to adopt healthy nutrition behaviors, as well as mobile apps for self-monitoring gestational diabetes [10], hypertension [11], and depression [12] have all shown improved health outcomes upon use. The removal of social pressures derived from face-to-face interactions with health care providers may also reduce social desirability bias, as seen in computer-based interventions for smoking cessation, which can decrease the odds of smoking during pregnancy by more than three-fold [13].

Prior Work

To our knowledge, no systematic review to date has evaluated the effectiveness of digital interventions for preventing alcohol consumption among pregnant women. By contrast, multiple systematic reviews and meta-analyses have examined the effectiveness of digital interventions for smoking cessation [13-15]. Only systematic reviews on the effectiveness of nondigital interventions for preventing alcohol consumption during pregnancy, such as cognitive-behavioral therapy and motivational interviewing [16-19], were found in our analyses. By contrast, a number of reviews examined the effectiveness of digital and computer-based alcohol intervention programs in primary care [20,21] or for patients recovering from substance use disorders [22,23], but such studies did not target pregnant women or women planning to become pregnant.

Goal of This Study

This systematic review sought to (1) identify the current studies describing the above-mentioned digital interventions, (2) assess whether these digital interventions are effective in preventing alcohol consumption among the target population, and (3) examine the extent to which digital interventions on various platforms, such as computers (web-based, internet, eHealth, etc), mobiles, and text messaging services, may vary in their degree of effectiveness in preventing alcohol consumption.

Methods

Search Strategy and Data Sources

Studies that discussed digital interventions to prevent alcohol consumption among pregnant women or women planning to become pregnant were identified by searching MEDLINE/PubMed (National Library of Medicine, NCBI), Embase (Elsevier), Cumulative Index of Nursing and Allied Health Literature (CINAHL Plus, EBSCO), and Web of Science Core Collection (Clarivate). Controlled vocabulary terms (ie, MeSH, Emtree, and CINAHL subject headings) were used when available and appropriate. The search strategies were designed and executed by a librarian (CM). Searches were not limited to a specific region, language, study design, or time period. The exact search terms used in each of the databases, and corresponding result numbers, are provided in Multimedia Appendix 1. The reference lists of identified studies were manually reviewed by SSO and DC to prevent relevant studies from being excluded in our search for relevant articles. Endnote X9 and Covidence software were used for database management.

Eligibility Criteria

We included studies that (1) targeted pregnant women or women planning to become pregnant, (2) measured the use of a digital intervention aiming to prevent alcohol consumption during pregnancy, (3) involved a digital interaction between the patient and a health care provider or professionally developed service (social media where subjects communicated with one another were excluded), and (4) reported rates of alcohol abstinence.

Data Management, Screening Process, and Data Extraction

Using these eligibility criteria, 2 independent investigators (SSO and DC) examined all studies reporting the use of a digital intervention to prevent alcohol consumption among pregnant women. All studies were screened at the title and abstract levels and excluded if the main target population did not consist of pregnant women or women planning to become pregnant, or if they did not include a digital intervention or a control group/preintervention comparison group. Subsequently, full-text reviews were performed to ensure that all articles measured and reported alcohol abstinence, and involved a digital interaction with a health care provider or professionally developed service. Any discrepancies were resolved by discussion. For the extraction of data regarding intervention characteristics and outcome measures (effect size), an online data extraction sheet was employed so that 2 independent investigators (SSO and JYM) could extract the necessary information. Regarding interrater reliability, kappa values (κ) of 0.78 for the title and abstract screening, and 0.84 for the full-text review were obtained. As a kappa coefficient exceeding 0.75 indicates strong agreement according to Fleiss et al [24], no further calibration was required.

Data Analyses

Rates of alcohol abstinence during pregnancy were extracted and presented as crude odds ratios (ORs) to maximize similarity between different studies. To examine the extent to which a digital intervention was effective, a random effects primary meta-analysis was conducted to determine the combined effect size and extent to which each digital intervention affected overall alcohol abstinence. An exploratory subgroup analysis was carried out to determine whether different platforms of digital interventions differed in the extent to which they affected the effect size. A random effects model was adopted for all meta-analyses to estimate intervention effects with 95% CIs that fall on a distribution of effect sizes. The Cohen *Q* test for chi-squared distribution and an inconsistency index (I^2) were implemented to test for heterogeneity among studies. Visual inspection of funnel plot asymmetry and the Egger test were used to assess the possibility of publication bias. All meta-analyses were performed using RStudio.

Quality Assessment

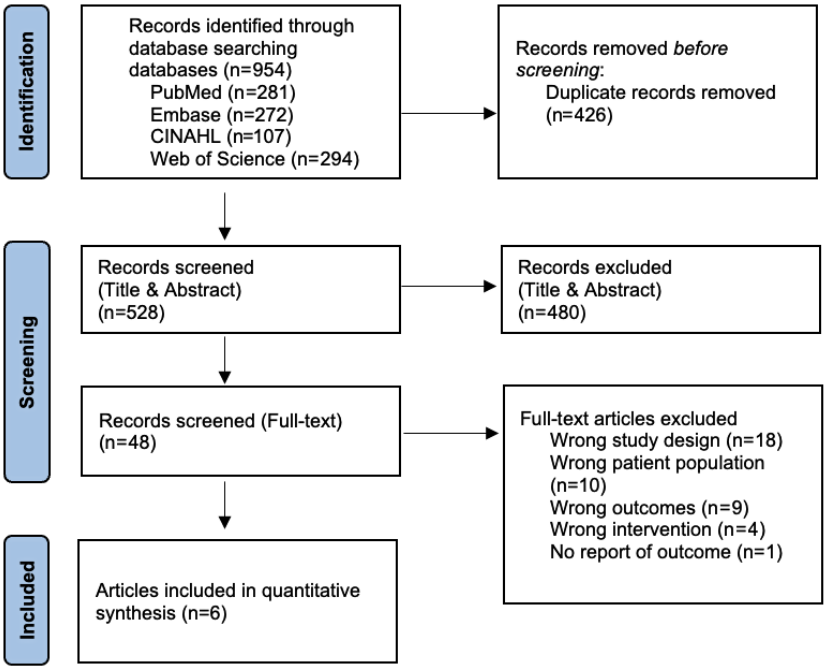
We assessed study quality in terms of potential bias using the Cochrane Collaboration tool for randomized controlled trials to assess the validity of the included studies [25]. A statistic of heterogeneity was calculated to quantify the proportion of variation across studies due to variability in the effect size rather than sampling variance (I^2). Cochran Q was used to formally test for heterogeneity. Publication bias was assessed through visual assessments of funnel plot asymmetry and was tested using the Egger test.

Results

Identification of Studies

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart in Figure 1 summarizes the search results and selection process of all studies included in our synthesis. Overall, the number of records identified by our database searches was 954. Of these records, 480 were removed during the title and abstract screening process, and a further 48 were screened for the full-text review.

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



Study Characteristics

Of the 48 articles assessed for eligibility, 42 were excluded for the following reasons: (1) weak study design in terms of the absence of a control group (pertaining to usual care or a preintervention baseline) or no targeting of alcohol consumption prevention during pregnancy (n=18); (2) no targeting of currently pregnant women or women with plans to become pregnant (n=10); (3) no outcome measure for alcohol abstinence (n=9); (4) no use of a digital intervention (n=4), and (5) no report of the outcome of interest (n=1). Ultimately, 6 studies were included in our final review. Table 1 provides a general summary of the included papers. Trials took place in the United States (n=5) or the Netherlands (n=1) between 2012 and 2018 [26,27,29-32].

Table 1. Characteristics of the included studies.

Author	Country	Sample size, n	Mean age (years)	Population sample	Control	Digital intervention	Follow-up assessment
Evans et al, 2014 [26]	United States	459	26.5	Pregnant military health care beneficiaries aged 18-45 years presenting for care (>14 weeks' gestation)	Usual care only	Text4Baby: Text messaging service on nutrition, smoking, taking vitamins, alcohol use, flu shots, health care appointments, health information seeking, and related risk prevention behaviors	4 weeks (pilot study)
Evans et al, 2012 [27]	United States	86	27.6	Pregnant women first presenting for care at the Fairfax County, Virginia Health Department	Usual care only	Text4Baby Pilot: Text messaging service with immediate "just-in-time" tips about prenatal and postpartum health outcomes	28 weeks of the baby's gestational age
Ingersoll et al, 2018 [32]	United States	71	27.8	Pregnant women *and* women of childbearing age between the ages of 18 and 44 years, recruited for study online	Patient education	CHOICES intervention: Automated internet intervention providing 6 web-based cores of information, videos, and interactive activities (eg, diaries) regarding alcohol-exposed pregnancies	24 weeks post-treatment
Ondersma et al, 2015 [29]	United States	48	__[a]	Pregnant women seeking services at a prenatal care clinic affiliated with the Henry Ford Health System in Detroit, Michigan	Time-matched (20 minutes) and moderately interactive intervention focused on infant nutrition, with no mention of alcohol use during pregnancy	e-SBI intervention facilitating self-change and/or treatment-seeking through a 20-minute interactive session, using techniques such as education about alcohol-related pregnancies and feedback regarding proactive problem-solving	Postpartum, for the past 90 days (22-23 weeks)
van der Wulp et al, 2014 [30]	The Netherlands	258	32.6	Pregnant women seeking services at midwifery practices in the Netherlands	Usual care only	Both computer tailoring internet-based feedback and offline health counseling based on the I-Change model (promote awareness, motivation, and action for behavioral change)	24 weeks post-treatment
Wernette et al, 2018 [31]	United States	50	24.4	Pregnant women visiting a prenatal clinic in a large inner-city hospital	Time- and attention-matched control group (watched segments of popular television shows and received brochures about health risks during pregnancy postintervention)	Computer-delivered single-session brief motivational intervention plus booster session addressing both substance use and sexually transmitted infection risk	4 months post-treatment

[a]Not reported.

Digital Interventions

Two studies delivered digital content via a text messaging service called "Text4Baby," which provides weekly tips about prenatal care, emotional support, alcohol and drugs, infectious diseases, and exercise to pregnant women and new mothers [26,27]. In the prenatal message module, which was used in both studies in our review, 3 free-text messages were sent to participants weekly throughout their pregnancies [28]. Each message was around 150 characters long (eg, "Free msg: Give your baby a good start by not drinking alcohol, smoking, or using drugs. For help, call 800-784-8669 (smoking); 800-662-4357 (drugs & alcohol)") and was designed to be understandable to low-literacy populations [28]. Messages were developed in advance for varying stages of gestation by a team of epidemiologists and experts in obstetrics, pediatrics, family practice, and health communication [28].

Four studies included computer/internet-based interventions consisting of interactive counseling sessions, educational videos, and interactive activities (ie, diary writing, meditation, etc) [29-31]. Counseling sessions consisted of various interactions with midwives or health care professionals, such as regular "feedback letters" from midwives via email (eg, "Drinking alcohol can be harmful to your unborn baby, even if it's just a sip. The type of alcohol you drink (beer, wine or spirits) does not matter") [30]. One electronic screening and brief intervention

(e-SBI) consisted of educational videos featuring mothers who avoided alcohol use during pregnancy, or health care professionals informing participants about health care risks and cost-savings [29].

Control Groups

Three studies used usual care in the form of a standard physician, obstetrician, or nurse-midwife/midwife providing advice [26,27,30] as the control group arm. One study used offline "patient education" as the control group [32], while 2 studies developed a time- and attention-matched intervention for the control group that did not mention any information about the harms of prenatal alcohol exposure (eg, viewing of a segment of a popular television show) [29,31].

Primary Outcome

Alcohol consumption during pregnancy was employed as the primary outcome. Studies administered self-reported questionnaires via telephone/email asking participants whether or not they had consumed any alcohol during the pregnancy period (eg, "Since you found out about your pregnancy, have you consumed alcoholic beverages?" [yes/no]). Participants were mostly questioned at 16 [31] to 24 weeks posttreatment [30,32], or after 28 weeks of gestation [27]. However, in 1 pilot study, the short-term effects of a 4-week text messaging intervention were examined [26], while in another study, alcohol consumption within the past 90 days was questioned postpartum via an AUDIO Computer-Assisted Self Interview [29].

Statistical Analyses

A primary meta-analysis including 6 trial arms from 6 studies was performed. The sample-weighted OR indicated that digital interventions decreased the odds of alcohol consumption during pregnancy compared with control groups (OR 0.62, 95% CI 0.42-0.91; *P*=.02) (Figure 2). In 1 study, there was no difference in the effect estimate between the intervention and control groups [26]; however, all other studies showed that alcohol consumption decreased among women using digital interventions. Tests of heterogeneity suggested that we failed to reject the null hypothesis of differences in the effect being a result of sampling variation (I^2=0%; *P*=.85).

A stratified analysis examining the influence of different intervention platforms revealed that computer-based interventions (OR 0.59, 95% CI 0.38-0.93) were effective for preventing alcohol consumption; however, too few studies of text messaging (OR 0.85, 95% CI 0.29-2.52) were available to draw a conclusion regarding the effect of this platform (Figure 3).

When studies were stratified according to each publication's quality risk of bias, point estimates (OR 0.62) were identical across study quality (Figure 4). However, due to the small number of studies analyzed, estimates were presumed to be imprecise.

Figure 2. Effectiveness of digital interventions for preventing alcohol consumption in pregnancy [26,27,29-32].

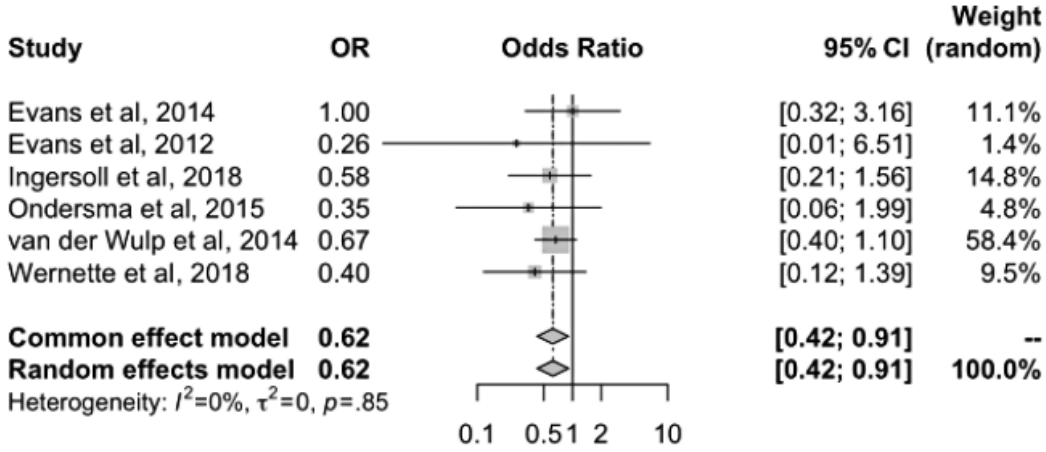


Figure 3. Effectiveness of digital interventions by platform [26,27,29-32].

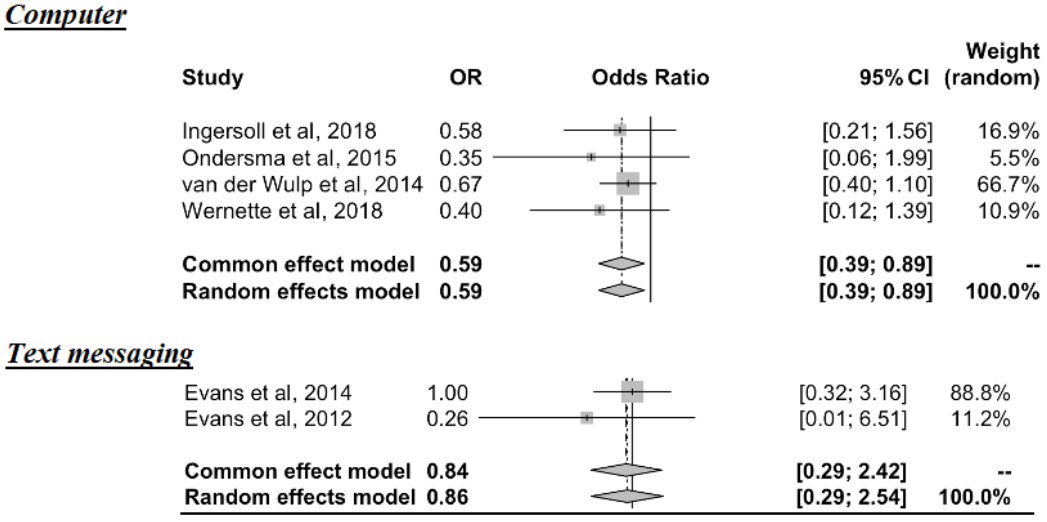
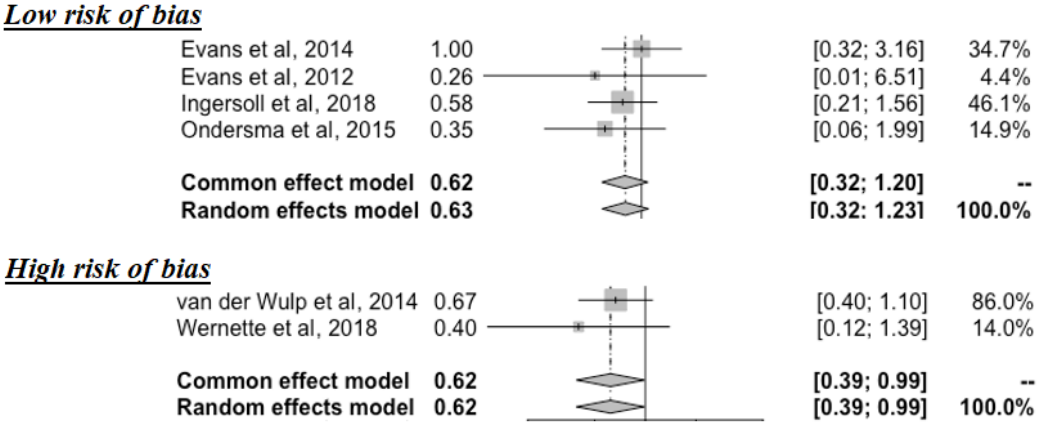


Figure 4. Effectiveness of digital interventions by quality risk of bias [26,27,29-32].



Quality Assessment

A summary of the quality assessment can be found in Figure 5. All studies had a high risk of bias in at least one key domain. All studies were randomized, and all but 1 [33] used a randomizing algorithm or software program to maintain research assistant blinding [26,27,29-32]. Studies had limited information regarding the extent to which trial participants were blinded about their allocation; however, most studies had various mechanisms for blinding clinicians. For example, in 2 studies, it was reported that clinicians who met with patients were blinded so that the randomization occurred outside the actual clinical visit and the trial data were not accessed by clinicians during the study [26,27]. Another study ensured that follow-up evaluators at childbirth were blinded so that evaluations would not be subject to any detection bias [29].

In a high-risk study, the authors reported that the blinding of both participants and researchers was not possible because they had to keep track of whether participants received additional counseling from their midwives or tailored feedback via the computer [30]. Another study also reported problems regarding an imbalance in the computerized randomization, and the presence of an unblinded research assistant who gave instructions to certain participants and may have contributed to the intervention effect [31]. All studies were at high risk of incomplete outcome data, as measures for drinking were all self-reported and loss to follow-up ranged from approximately 20% [29] to 50% [27]. Selective reporting was of concern in 1 study [32], where prespecified outcomes regarding certain continuous drinking variables were not reported [32].

Results from the Egger test for funnel plot asymmetry were not statistically significant (t_5=−1.66; P=.16; Figure 6), suggesting the absence of publication bias; however, such results should be interpreted with caution as the Egger method has limited power when used in smaller samples (n<10) [34].

Figure 5. Risk of bias summary [26,27,29-32].

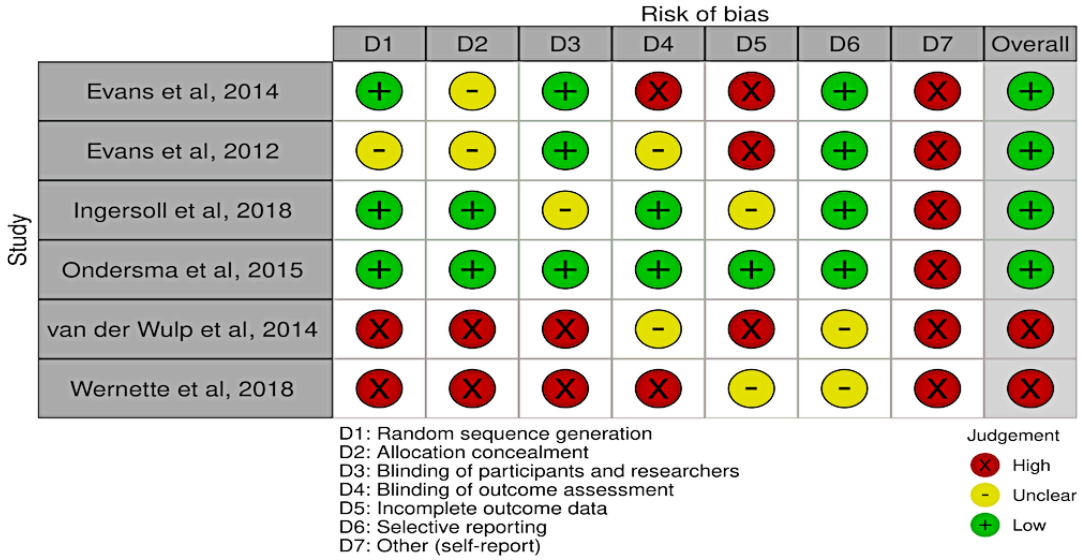
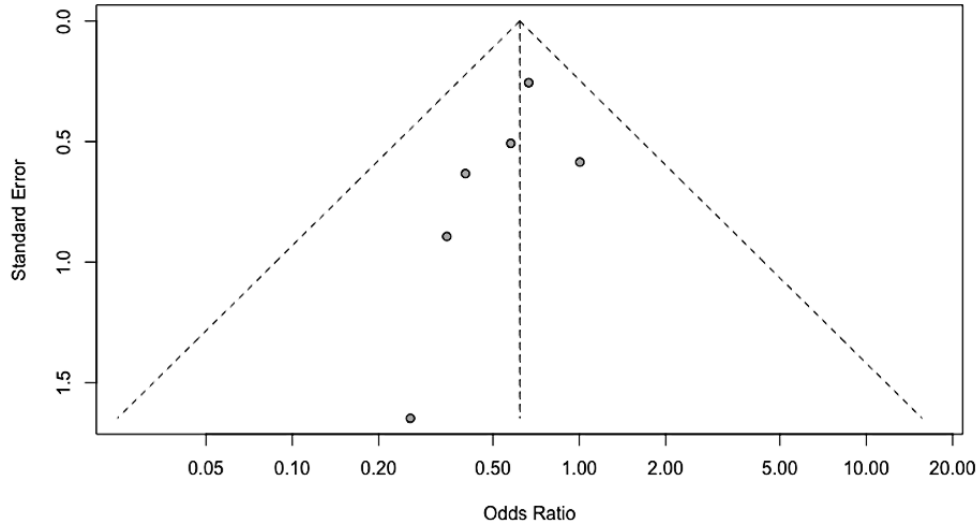


Figure 6. Funnel plot assessing publication bias.



Discussion

Principal Findings

In this systematic review, we found that digital interventions for preventing alcohol consumption during pregnancy may be effective in preventing alcohol consumption, especially on computer/internet-based platforms. Excluding a pilot Text4Baby study [27], all studies showed that digital interventions may decrease the odds of drinking during pregnancy relative to comparison groups. However, our findings must be interpreted with caution as it may not hold for interventions with a low risk of bias. As the first systematic review to assess the effectiveness of digital interventions targeting pregnant drinkers, our review is timely as it supports the claim that more technological interventions, possibly in combination with offline counseling strategies, should be incorporated into existing prenatal care services.

Comparison With Prior Work

Regarding text messaging platforms, there were too few studies in our review to draw a conclusion regarding their effectiveness as digital platforms for alcohol abstinence. Previous studies on the use of text messaging services to raise awareness about smoking cessation and flu vaccinations among pregnant women have shown mixed results, with some studies reporting promise compared to nontailored or internet platforms [35,36], and others claiming that they are less effective than visually engaging interventions like videos and iBooks [35]. It should be noted that in the 2 text messaging trials in our review, the entire evaluation period only lasted for 4 weeks, which was relatively shorter than the period of the other platforms [37]. Scholars of technology-based strategies to improve health outcomes among pregnant women have noted that short-term interventions (approximately <16 weeks) may not be successful in bringing about behavioral change [38], which may explain why 4 weeks was not enough to examine the effect. While the most vulnerable period for brain volume reduction and FASDs is during the first trimester [39], FASDs may occur from any alcohol intake during all 3 trimesters of pregnancy, regardless of the timing or exposure amount. Thus, more research is warranted to examine how text messaging services, which are not only cost-effective

but also flexible and accessible, may be employed to deliver longer-lasting interventions throughout pregnancy.

As expected, the most effective interventions in our review were those that incorporated both offline house counseling and internet or mobile-based feedback (ie, "blended" care) for individuals [30,31]. In the study by van der Wulp et al comparing 6 months of computer-tailored programs to usual care and health counseling, computer-tailored programs were more effective in reducing prenatal alcohol use than face-to-face counseling sessions [30]. Such findings show that because digital tailoring has the potential to decrease social pressure that may arise from face-to-face interactions with health care providers, many pregnant women may prefer it to other offline platforms [30].

Strengths and Limitations

As the first systematic review to question the effectiveness of digital interventions for preventing alcohol consumption during pregnancy, the findings of this review are novel. However, there are some limitations of our review. While our assessment of funnel plot symmetry did not formally detect publication bias (by significance testing), the sample of studies was small. It is possible that underpowered studies with null results are missing (a "file drawer" problem). Cultural differences between the United States and the Netherlands may also have affected study outcomes; while both the United States and the Netherlands officially recommend that pregnant women completely abstain from alcohol, the prevalence of alcohol consumption during pregnancy is higher in the Netherlands (19%-21%) than in the United States (15%) [40].

Most concerningly, the primary outcome for alcohol abstinence during pregnancy was self-reported, and follow-up methods/timing differed among all studies. The absence of validation by biomarkers to assess abstinence was a fundamental limitation of the included trials, which is concerning as self-reports of alcohol consumption may be affected by memory loss from alcohol abuse and underreporting due to a fear of negative consequences like being reported to Child Protective Services [41]. In many states, for example, health care providers are required by the federal Child Abuse Prevention and Treatment Act legislation to notify Child Protective Services when they are involved in the delivery or care of infants with FASDs [42].

Intervention duration, quality, and intensity could not be controlled for, with some studies, such as the e-SBI trial, specifically targeting high-risk individuals via professional counseling methods (eg, motivational interviewing) [29], and other studies incorporating alcohol intake monitoring in a larger more generalized program for pregnant women in general (ie, Text4Baby) [26,27]. As seen in the quality assessment of various biases, some studies had large losses to follow-up, lack of information about the extent to which patients/evaluators were blinded with regard to the randomization process, and possible risk of incomplete outcome data [43]. Some studies had trouble blinding instructors [31] and participants [30]. All studies had difficulty retaining participants for long-term follow-up, with 1 study having a retention rate of less than 50% [27].

Future Directions

Future studies would benefit from controlling for discrepancies among varying trials regarding the quality of usual care provided in the control group, assessment of alcohol abstinence, and intervention duration/quality. However, in our study, this was not possible due to the limited descriptions provided by the included studies regarding these factors. In future studies when more trials targeting alcohol abstinence during pregnancy are available for review, a more consistent and thorough subgroup analysis of intervention techniques, involving video, counseling, blended care, etc, is warranted.

Conclusions

More studies are required to assess the extent to which digital interventions targeting pregnant drinkers may be effective for women from disadvantaged backgrounds and/or a low socioeconomic status. While few programs and trials are currently available to review, digital technologies are being embraced rapidly for personalized health care. Future studies would benefit from assessing how better allocation of both online and offline resources may help pregnant women and women planning to become pregnant avoid consuming alcohol and other teratogenic substances during their pregnancies.

Acknowledgments

This research was supported by a grant of the Korea Health Promotion R&D Project, funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HS21C0037).

Authors' Contributions

IK and SSO conceptualized and supervised the study. IK, SSO, JAL, and JYM developed the methodology. CM provided the literature search. SSO, JAL, and DC screened the studies and performed the formal analysis/quality assessments. SSO and IK performed the validation. SSO, JAL, CM, and IK wrote, reviewed, and edited the manuscript. All authors read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Summary of database search terms.
[DOCX File , 17 KB - jmir_v24i4e35554_app1.docx]

References

1. Bailey BA, Sokol RJ. Prenatal alcohol exposure and miscarriage, stillbirth, preterm delivery, and sudden infant death syndrome. Alcohol Res Health 2011;34(1):86-91 [FREE Full text] [Medline: 23580045]
2. Popova S, Lange S, Probst C, Gmel G, Rehm J. Global prevalence of alcohol use and binge drinking during pregnancy, and fetal alcohol spectrum disorder. Biochem Cell Biol 2018 Apr;96(2):237-240. [doi: 10.1139/bcb-2017-0077] [Medline: 28834683]
3. Seguin D, Gerlai R. Fetal alcohol spectrum disorders: Zebrafish in the analysis of the milder and more prevalent form of the disease. Behav Brain Res 2018 Oct 15;352:125-132 [FREE Full text] [doi: 10.1016/j.bbr.2017.10.005] [Medline: 28988969]
4. Sellman D, Connor J. In utero brain damage from alcohol: a preventable tragedy. N Z Med J 2009 Nov 20;122(1306):6-8. [Medline: 20145681]
5. Lyall V, Wolfson L, Reid N, Poole N, Moritz KM, Egert S, et al. "The Problem Is that We Hear a Bit of Everything…": A Qualitative Systematic Review of Factors Associated with Alcohol Use, Reduction, and Abstinence in Pregnancy. Int J Environ Res Public Health 2021 Mar 26;18(7):3445 [FREE Full text] [doi: 10.3390/ijerph18073445] [Medline: 33810338]
6. Vorgias D, Bernstein B. Fetal Alcohol Syndrome. In: StatPearls. Treasure Island, FL: StatPearls Publishing; 2021.
7. Banerji A, Shah C. Ten-year experience of fetal alcohol spectrum disorder; diagnostic and resource challenges in Indigenous children. Paediatr Child Health 2017 Jun;22(3):143-147 [FREE Full text] [doi: 10.1093/pch/pxx052] [Medline: 29479201]
8. May PA, Gossage JP. Maternal risk factors for fetal alcohol spectrum disorders: not as simple as it might seem. Alcohol Res Health 2011;34(1):15-26 [FREE Full text] [Medline: 23580036]
9. Mu A, Deng Z, Wu X, Zhou L. Does digital technology reduce health disparity? Investigating difference of depression stemming from socioeconomic status among Chinese older adults. BMC Geriatr 2021 Apr 21;21(1):264 [FREE Full text] [doi: 10.1186/s12877-021-02175-0] [Medline: 33882865]
10. Larbi D, Randine P, Årsand E, Antypas K, Bradway M, Gabarron E. Methods and Evaluation Criteria for Apps and Digital Interventions for Diabetes Self-Management: Systematic Review. J Med Internet Res 2020 Jul 06;22(7):e18480 [FREE Full text] [doi: 10.2196/18480] [Medline: 32628125]
11. Band R, Hinton L, Tucker KL, Chappell LC, Crawford C, Franssen M, et al. Intervention planning and modification of the BUMP intervention: a digital intervention for the early detection of raised blood pressure in pregnancy. Pilot Feasibility Stud 2019;5:153 [FREE Full text] [doi: 10.1186/s40814-019-0537-z] [Medline: 31890265]
12. Genovez M, Vanderkruik R, Lemon E, Dimidjian S. Psychotherapeutic Treatments for Depression During Pregnancy. Clin Obstet Gynecol 2018 Sep;61(3):562-572. [doi: 10.1097/GRF.0000000000000388] [Medline: 29965823]
13. Griffiths SE, Parsons J, Naughton F, Fulton EA, Tombor I, Brown KE. Are digital interventions for smoking cessation in pregnancy effective? A systematic review and meta-analysis. Health Psychol Rev 2018 Dec;12(4):333-356. [doi: 10.1080/17437199.2018.1488602] [Medline: 29912621]
14. Griffiths SE, Brown KE, Fulton EA, Tombor I, Naughton F. Are digital interventions for smoking cessation in pregnancy effective? A systematic review protocol. Syst Rev 2016 Dec 01;5(1):207 [FREE Full text] [doi: 10.1186/s13643-016-0390-6] [Medline: 27906071]
15. Tombor I, Neale J, Shahab L, Ruiz M, West R. Healthcare Providers' Views on Digital Smoking Cessation Interventions for Pregnant Women. J. Smok Cessat 2014 Mar 10;10(2):116-123. [doi: 10.1017/jsc.2014.6]
16. Erng MN, Smirnov A, Reid N. Prevention of Alcohol-Exposed Pregnancies and Fetal Alcohol Spectrum Disorder Among Pregnant and Postpartum Women: A Systematic Review. Alcohol Clin Exp Res 2020 Dec;44(12):2431-2448. [doi: 10.1111/acer.14489] [Medline: 33119893]
17. Fergie L, Campbell KA, Coleman-Haynes T, Ussher M, Cooper S, Coleman T. Identifying Effective Behavior Change Techniques for Alcohol and Illicit Substance Use During Pregnancy: A Systematic Review. Ann Behav Med 2019 Jul 17;53(8):769-781 [FREE Full text] [doi: 10.1093/abm/kay085] [Medline: 30379989]
18. Gilinsky A, Swanson V, Power K. Interventions delivered during antenatal care to reduce alcohol consumption during pregnancy: A systematic review. Addiction Research & Theory 2010 Aug 31;19(3):235-250. [doi: 10.3109/16066359.2010.507894]
19. Ujhelyi Gomez K, Goodwin L, Jackson L, Jones A, Chisholm A, Rose AK. Are psychosocial interventions effective in reducing alcohol consumption during pregnancy and motherhood? A systematic review and meta-analysis. Addiction 2021 Jul;116(7):1638-1663. [doi: 10.1111/add.15296] [Medline: 33067887]
20. Nair NK, Newton NC, Shakeshaft A, Wallace P, Teesson M. A Systematic Review of Digital and Computer-Based Alcohol Intervention Programs in Primary Care. Curr Drug Abuse Rev 2015;8(2):111-118. [doi: 10.2174/1874473708666150916113538] [Medline: 26373848]
21. Ramsey AT, Satterfield JM, Gerke DR, Proctor EK. Technology-Based Alcohol Interventions in Primary Care: Systematic Review. J Med Internet Res 2019 Apr 08;21(4):e10859 [FREE Full text] [doi: 10.2196/10859] [Medline: 30958270]

22. Bendtsen M, McCambridge J, Åsberg K, Bendtsen P. Text messaging interventions for reducing alcohol consumption among risky drinkers: systematic review and meta-analysis. Addiction 2021 May;116(5):1021-1033 [FREE Full text] [doi: 10.1111/add.15294] [Medline: 33047865]
23. Nesvåg S, McKay JR. Feasibility and Effects of Digital Interventions to Support People in Recovery From Substance Use Disorders: Systematic Review. J Med Internet Res 2018 Aug 23;20(8):e255 [FREE Full text] [doi: 10.2196/jmir.9873] [Medline: 30139724]
24. Fleiss J, Levin B, Paik M. Statistical Methods for Rates and Proportions. Hoboken, NJ: John Wiley & Sons; 2013.
25. 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]
26. Evans WD, Wallace Bihm J, Szekely D, Nielsen P, Murray E, Abroms L, et al. Initial outcomes from a 4-week follow-up study of the Text4baby program in the military women's population: randomized controlled trial. J Med Internet Res 2014 May 20;16(5):e131 [FREE Full text] [doi: 10.2196/jmir.3297] [Medline: 24846909]
27. Evans WD, Wallace JL, Snider J. Pilot evaluation of the text4baby mobile health program. BMC Public Health 2012 Nov 26;12:1031 [FREE Full text] [doi: 10.1186/1471-2458-12-1031] [Medline: 23181985]
28. Whittaker R, Matoff-Stepp S, Meehan J, Kendrick J, Jordan E, Stange P, et al. Text4baby: development and implementation of a national text messaging health information service. Am J Public Health 2012 Dec;102(12):2207-2213. [doi: 10.2105/AJPH.2012.300736] [Medline: 23078509]
29. Ondersma SJ, Beatty JR, Svikis DS, Strickler RC, Tzilos GK, Chang G, et al. Computer-Delivered Screening and Brief Intervention for Alcohol Use in Pregnancy: A Pilot Randomized Trial. Alcohol Clin Exp Res 2015 Jul;39(7):1219-1226 [FREE Full text] [doi: 10.1111/acer.12747] [Medline: 26010235]
30. van der Wulp NY, Hoving C, Eijmael K, Candel MJJ, van Dalen W, De Vries H. Reducing alcohol use during pregnancy via health counseling by midwives and internet-based computer-tailored feedback: a cluster randomized trial. J Med Internet Res 2014 Dec 05;16(12):e274 [FREE Full text] [doi: 10.2196/jmir.3493] [Medline: 25486675]
31. Wernette GT, Plegue M, Kahler CW, Sen A, Zlotnick C. A Pilot Randomized Controlled Trial of a Computer-Delivered Brief Intervention for Substance Use and Risky Sex During Pregnancy. J Womens Health (Larchmt) 2018 Jan;27(1):83-92 [FREE Full text] [doi: 10.1089/jwh.2017.6408] [Medline: 28981379]
32. Ingersoll K, Frederick C, MacDonnell K, Ritterband L, Lord H, Jones B, et al. A Pilot RCT of an Internet Intervention to Reduce the Risk of Alcohol-Exposed Pregnancy. Alcohol Clin Exp Res 2018 Jun;42(6):1132-1144 [FREE Full text] [doi: 10.1111/acer.13635] [Medline: 29741798]
33. Waterson E, Murray-Lyon IM. Preventing fetal alcohol effects; a trial of three methods of giving information in the antenatal clinic. Health Educ Res 1990;5(1):53-61. [doi: 10.1093/her/5.1.53]
34. Song F, Khan KS, Dinnes J, Sutton AJ. Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. Int J Epidemiol 2002 Feb;31(1):88-95. [doi: 10.1093/ije/31.1.88] [Medline: 11914301]
35. Parsons J, Griffiths SE, Thomas N, Atherton H. How effective are digital interventions in increasing flu vaccination among pregnant women? A systematic review and meta-analysis. J Public Health (Oxf) 2021 Jun 23:fdab220. [doi: 10.1093/pubmed/fdab220] [Medline: 34164686]
36. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016 Apr 10;4:CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
37. Van Dijk MR, Huijgen NA, Willemsen SP, Laven JS, Steegers EA, Steegers-Theunissen RP. Impact of an mHealth Platform for Pregnancy on Nutrition and Lifestyle of the Reproductive Population: A Survey. JMIR Mhealth Uhealth 2016 May 27;4(2):e53 [FREE Full text] [doi: 10.2196/mhealth.5197] [Medline: 27234926]
38. Sherifali D, Nerenberg KA, Wilson S, Semeniuk K, Ali MU, Redman LM, et al. The Effectiveness of eHealth Technologies on Weight Management in Pregnant and Postpartum Women: Systematic Review and Meta-Analysis. J Med Internet Res 2017 Oct 13;19(10):e337 [FREE Full text] [doi: 10.2196/jmir.8006] [Medline: 29030327]
39. Maier SE, Chen WJ, Miller JA, West JR. Fetal alcohol exposure and temporal vulnerability regional differences in alcohol-induced microencephaly as a function of the timing of binge-like alcohol exposure during rat brain development. Alcohol Clin Exp Res 1997 Nov;21(8):1418-1428. [doi: 10.1111/j.1530-0277.1997.tb04471.x] [Medline: 9394113]
40. Lanting CI, van Dommelen P, van der Pal-de Bruin KM, Bennebroek Gravenhorst J, van Wouwe JP. Prevalence and pattern of alcohol consumption during pregnancy in the Netherlands. BMC Public Health 2015 Jul 29;15:723 [FREE Full text] [doi: 10.1186/s12889-015-2070-1] [Medline: 26219278]
41. Roberts SCM, Nuru-Jeter A. Women's perspectives on screening for alcohol and drug use in prenatal care. Womens Health Issues 2010;20(3):193-200 [FREE Full text] [doi: 10.1016/j.whi.2010.02.003] [Medline: 20457407]
42. Williams JF, Smith VC, COMMITTEE ON SUBSTANCE ABUSE. Fetal Alcohol Spectrum Disorders. Pediatrics 2015 Nov;136(5):e1395-e1406. [doi: 10.1542/peds.2015-3113] [Medline: 26482673]
43. van der Windt M, van der Kleij RM, Snoek KM, Willemsen SP, Dykgraaf RHM, Laven JSE, et al. Impact of a Blended Periconception Lifestyle Care Approach on Lifestyle Behaviors: Before-and-After Study. J Med Internet Res 2020 Sep 30;22(9):e19378 [FREE Full text] [doi: 10.2196/19378] [Medline: 32996885]

Abbreviations

e-SBI: electronic screening and brief intervention
FASD: fetal alcohol spectrum disorder
OR: odds ratio

Edited by T Leung; submitted 08.12.21; peer-reviewed by H Mehdizadeh; comments to author 06.01.22; revised version received 20.01.22; accepted 01.03.22; published 11.04.22.

Please cite as:
Oh SS, Moon JY, Chon D, Mita C, Lawrence JA, Park EC, Kawachi I
Effectiveness of Digital Interventions for Preventing Alcohol Consumption in Pregnancy: Systematic Review and Meta-analysis
J Med Internet Res 2022;24(4):e35554
URL: https://www.jmir.org/2022/4/e35554
doi:10.2196/35554
PMID:35404257

Review

Computerized Cognitive Behavioral Therapy for Treatment of Depression and Anxiety in Adolescents: Systematic Review and Meta-analysis

Alice Wickersham[1*], MSc; Tamara Barack[1*], MSc; Lauren Cross[2], MPhil; Johnny Downs[1], PhD

[1]Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

[2]Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Alice Wickersham, MSc
Institute of Psychiatry, Psychology & Neuroscience
King's College London
16 De Crespigny Park
London, SE5 8AF
United Kingdom
Phone: 44 (0)20 7848 0002
Email: alice.wickersham@kcl.ac.uk

Abstract

Background: Depression and anxiety are major public health concerns among adolescents. Computerized cognitive behavioral therapy (cCBT) has emerged as a potential intervention, but its efficacy in adolescents remains unestablished.

Objective: This review aimed to systematically review and meta-analyze findings on the efficacy of cCBT for the treatment of adolescent depression and anxiety.

Methods: Embase, PsycINFO, and Ovid MEDLINE were systematically searched for randomized controlled trials in English, which investigated the efficacy of cCBT for reducing self-reported depression or anxiety in adolescents aged 11 to 19 years. Titles, abstracts, and full texts were screened for eligibility by 2 independent researchers (TB and LC). A random-effects meta-analysis was conducted to pool the effects of cCBT on depression and anxiety symptom scores compared with the control groups. Study quality was assessed using the Cochrane Collaboration Risk of Bias tool.

Results: A total of 16 randomized controlled trials were eligible for inclusion in this review, of which 13 (81%) were included in the meta-analysis. The quality of the studies was mixed, with 5 (31%) studies rated as good overall, 2 (13%) rated as fair, and 9 (56%) rated as poor. Small but statistically significant effects of cCBT were detected, with cCBT conditions showing lower symptom scores at follow-up compared with control conditions for both anxiety (standardized mean difference −0.21, 95% CI −0.33 to −0.09; I^2=36.2%) and depression (standardized mean difference −0.23, 95% CI −0.39 to −0.07; I^2=59.5%). Secondary analyses suggested that cCBT may be comparable with alternative, active interventions (such as face-to-face therapy or treatment as usual).

Conclusions: This meta-analysis reinforces the efficacy of cCBT for the treatment of anxiety and depression and is the first to examine this exclusively in adolescents. Future research could aim to identify the active components of these interventions toward optimizing their development and increasing the feasibility and acceptability of cCBT in this age group.

Trial Registration: PROSPERO CRD42019141941; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=141941

(J Med Internet Res 2022;24(4):e29842) doi:10.2196/29842

KEYWORDS

adolescent; anxiety; depression; meta-analysis

Introduction

Background

Depression and anxiety represent significant public health concerns among adolescents [1]. Depression and anxiety in adolescence are associated with negative outcomes which can extend into adulthood, such as suicidal behavior, risk of mental health disorders, substance abuse, poorer educational attainment, and poorer social functioning [2-5]. Cognitive behavioral therapy (CBT), a collaborative model of therapy, is widely considered the gold standard psychological intervention for the treatment of adolescent depression and anxiety [6,7]. Several reviews and meta-analyses have found it to be an effective treatment method, with improvements maintained during long-term follow-up [8,9].

However, many young people do not receive the help they need [10]. There are barriers to accessing treatment at the service level such as a lack of resources, inadequate staff training, limited availability of age-specific treatments, and difficulties in liaising with families [11]. Barriers at the patient level may include perceived stigma, confidentiality concerns, inability to access services, and a preference for self-reliance [12].

Computerized CBT (cCBT) programs, which deliver CBT digitally, may overcome some of these barriers. As heavy users of technology [13], cCBT programs may present various advantages over traditional face-to-face therapies for adolescents, such as reduced costs, accessibility, convenience, flexibility, and the avoidance of stigma associated with accessing traditional mental health services [14,15]. These programs vary widely in design and access. Many combine different types of media such as text, pictures, videos, activities, and gamification to engage adolescent audiences; some also include homework activities, personalization, and chat functionalities with trained therapists or administrators [14,16].

Previous systematic reviews and meta-analyses of the efficacy of digitally delivered CBT for anxiety or depression in children and young people have yielded promising results [14,17-23]. However, these reviews have typically been insensitive to age-specific effects, making it difficult to determine whether cCBT is equally effective for children, adolescents, and young adults. Ebert et al [18] stratified their meta-analysis by age group and found that studies examining cCBT among adolescents achieved better outcomes than those targeting children or mixed age groups. However, they examined studies published over 6 years ago (much has changed in digital health since then), with few available studies focusing on adolescents. More recently, Grist et al [19] stratified their meta-analysis by age group but did not differentiate cCBT from other technological interventions in doing so. Given that the National Institute for Health and Care Excellence (NICE) guidelines for depression in children and young people have recently recommended the use of digital CBT for mild depression, a strong and cCBT-specific evidence base for these technologies is crucial [7].

Objectives

The aim of this systematic review of randomized controlled trials (RCTs) is to update previous meta-analyses and focus exclusively on the adolescent age group to investigate the efficacy of cCBT for treating anxiety and depression. Adolescence represents a distinct developmental period of biological and social transition [24] during which the prevalence of mental health disorders increases [25,26]. Furthermore, adolescents have distinctive technology and internet use habits compared with other age groups [13,27]. Therefore, it is important that adolescence be investigated as a distinct developmental period to evaluate whether cCBT is an effective intervention in this age group.

Methods

The study was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Table S1 in Multimedia Appendix 1) [28] and was registered on PROSPERO (CRD42019141941).

Ethics Approval

This review makes use of already published data, so ethical approval was not required or sought.

Eligibility Criteria

Studies were included if (1) participants were aged 11 to 19 years (inclusive); (2) they examined any digital, computerized, or web-based CBT program; (3) they included depression or anxiety-related primary outcome measures; (4) they were completed RCTs; and (5) they were in English and published in a peer-reviewed journal.

The age criterion ranged from 11 years, matching the typical starting age for secondary school (United Kingdom) or middle school (United States), to 19 years, matching the upper threshold for adolescence used by the World Health Organization [29]. The RCT criterion was assessed based on the NICE definition of RCT [30]. No specific restrictions were placed on how depression or anxiety was operationalized; for example, they could be measured using self- or informant-rated symptom questionnaires or diagnostic criteria. In addition, any anxiety subtype was eligible for inclusion. No restrictions were placed on how cCBT was used as an intervention in the trials, such that the included programs could be preventive or treatment focused.

Search Strategy

Reviewer 1 (TB) conducted the database and reference list searches. PsycINFO, Embase, and Ovid MEDLINE were searched up to July 1, 2019. Keyword searches were grouped around three concepts: (1) age group, (2) cCBT interventions, and (3) depression or anxiety. Database filters were not applied (Table S2 in Multimedia Appendix 2). The reference lists of included articles and relevant existing systematic reviews were also searched for potentially relevant studies. Reviewers 1 (TB) and 2 (LC) independently conducted title, abstract, and full-text screening according to the study eligibility criteria. Disagreements were discussed until consensus was reached.

Reviewers 1 (TB) and 2 (LC) fully agreed on the final papers to be included in the review.

Data Extraction

Reviewer 1 (TB) extracted study characteristics from the included full texts, such as information on the participants (country, study population, exclusion criteria, sample size, and participant age and gender), cCBT intervention or interventions, and control group or groups. Reviewer 3 (AW) extracted data to be used in the meta-analysis, such as information on the outcomes (outcome measurement, means, SDs, and SEs) and study design (randomization type, analysis type, and follow-up sample sizes). Wherever possible, data were extracted from intention-to-treat analyses. If studies reported outcomes for multiple follow-up periods, data were extracted for the longest follow-up period for which outcomes were reported in all relevant study arms (Cochrane Handbook section 9.3.4 [31]). If studies reported data from multiple relevant depression or anxiety measures, we extracted the study's primary outcome measure by default, unless one of the secondary outcome measures was more similar to those used by the other included studies or if the outcome measure showed strong evidence of skew (Cochrane Handbook section 9.4.5.3 [31]). Reviewer 2 (LC) independently extracted the study characteristics and meta-analysis data of 10% of the included studies. The extraction agreement was high (81%). Following recommendations arising from peer reviews, reviewer 3 (AW) further extracted information on whether each study assessed treatment versus prevention, the role of parents in each intervention, and contacted corresponding authors for information on whether the interventions were available or delivered outside of the research setting (eg, in schools or in routine clinical care).

Study Quality

Study quality was assessed using the Cochrane Collaboration Risk of Bias tool [32]. This tool assessed seven areas for possible bias (scored as low risk, high risk, or unclear risk): random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, and other biases. The overall study quality score (good, fair, or poor) was then calculated using the thresholds for converting the Cochrane Collaboration tool to the Agency for Healthcare Research and Quality standards [32,33]. Reviewers 1 (TB) and 2 (LC) independently assessed the risk of bias. Disagreements were discussed until consensus was reached.

Statistical Analysis

We conducted meta-analyses to pool differences in follow-up depression and anxiety scores between the treatment and control arms. All analyses were conducted using Stata (version 15.0; StataCorp) [34]. Owing to anticipated heterogeneity in the study designs and scales used, we specified a random-effects model to pool standardized mean differences (SMDs) between the treatment and control arms using the Hedges correction [35,36]. For the main meta-analysis, if studies reported on multiple relevant study arms, we combined them to ensure that a single average treatment score was compared with a single average control score (Cochrane Handbook section 16.5.4 [31]). Heterogeneity was investigated using Cochran Q and the I^2 statistic, and publication bias was investigated using funnel plots and the Egger test for funnel plot asymmetry [37,38].

We conducted post hoc sensitivity analyses to investigate the effects of pooling different intervention types (treatment vs prevention), pooling different randomization techniques (clustered vs other), and pooling different analysis types (intention to treat vs other or unclear). We also investigated the effects of study quality, stratifying analyses by the overall study quality score (good, fair, or poor). Finally, we conducted a post hoc sensitivity analysis to investigate whether the pooled effect size varied according to the type of control group under study: active treatment (specific interventions or treatment as usual) versus other (waitlist, attentional, and no intervention controls). For this analysis, we did not combine multiple control arms into a single average control score as described above, such that studies could appear twice if they included 2 control arms (both active treatment and others).

Results

Study Characteristics

A total of 16 studies were included in this review (Figure 1). They included 4012 participants, with all participants aged in the range of 11 to 19 years and with varying gender balances (Table 1). The studies targeted various populations (Table S3 in Multimedia Appendix 3 [39-54]) and were conducted in a wide range of countries, including New Zealand, China, Japan, Australia, the United Kingdom, the Netherlands, Denmark, and Sweden. Studies had a range of participant exclusion criteria, most often excluding participants who had severe symptoms, had other disorders, were at high risk of self-harm or suicide or were already receiving treatment (Table S3 in Multimedia Appendix 3 [39-54]).

Interventions were conducted at school, at the participant's home, or in a setting of the participant's choice, such as a local child and adolescent mental health service, local general practice, and community center. The extent of clinician or therapist input varied among interventions but was typically minimal. In addition, most interventions did not require parents to take an active role, with the exception of 19% (3/16) of the studies in which parents received additional guidance to support their adolescents through the course of treatment [39-41]. Some of the studies investigated interventions that have been made publicly available, whereas others have not been implemented beyond the research setting. A description of each cCBT program is provided in Multimedia Appendix 4 (Table S4) [39-54].

Figure 1. PRISMA (Preferred Reporting Item for Systematic Reviews and Meta-Analyses) flow diagram for study inclusion. RCT: randomized controlled trial.

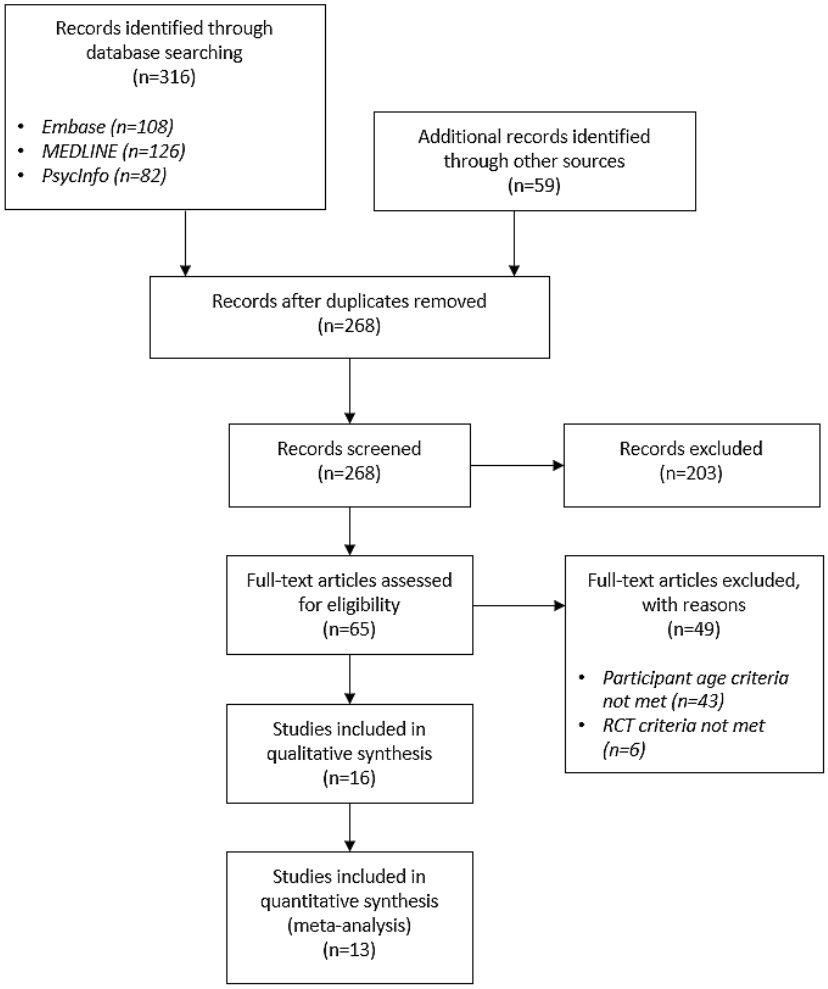


Table 1. Study characteristics.

Study	Country	Participants, n	Age (years), mean; range	Gender, male (%)	Intervention arm or arms	Control arm or arms
Calear et al [42]	Australia	1477	14.34; 12-17	44	MoodGYM	Waitlist
Fleming et al [43]	New Zealand	32	14.9; 13-16	56	SPARX[a]	Waitlist
Ip et al [44]	China	257	14.63; 13-17	32	Grasp the Opportunity (culturally modified from CATCH-IT[b])	Attention control (anti-smoking website without mental health prevention components)
Merry et al [45]	New Zealand	187	15.6; 12-19	34	SPARX	Treatment as usual (face-to-face therapy)
Poppelaars et al [46]	Netherlands	208	13.35; 11-16	0	SPARX	School-based CBT[c]; monitoring control[d]
Sekizaki et al [47]	Japan	80	15.75; 15	100	Group iCBT[e] program	No intervention
Smith et al [48]	United Kingdom	112	Not reported; 12-16	Not reported	Stressbusters	Waitlist
Spence et al [39]	Australia	115	13.98; 12-18	41	BRAVE-Online	Clinic-based CBT; waitlist
Sportel et al [49]	Netherlands	240	14.21; 12-15	28	Online Cognitive Bias Modification	Face-to-face CBT; no intervention
Stallard et al [50]	United Kingdom	20	Not reported; 11-17	67	Think, Feel, Do	Waitlist
Stasiak et al [51]	New Zealand	34	15.2; 13-18	59	The Journey	Attention control (computerized psychoeducation)
Stjerneklar et al [40]	Denmark	70	Not reported; 13-17	79	ChilledOut Online	Waitlist
Topooco et al [52]	Sweden	70	17.04; 15-19	Not reported	Blended approach with weekly therapist chats	Attention control (restricted access to platform and therapist chat but not the CBT component)
Wong et al [53]	Australia	976	Not reported; 14-16	30	ThisWayUp Schools (modules: “Overcoming Anxiety” or “Combating Depression”)	Usual health classes
Wright et al [54]	United Kingdom	91	15.34; 12-18	34	Stressbusters	Attention control (self-help website)
Wuthrich et al [41]	Australia	43	15.7; 14-17	37	Cool Teens CD-ROM	Waitlist

[a]SPARX: Smart, Positive, Active, Realistic, X-factor thoughts.

[b]CATCH-IT: Competent Adulthood Transition with Cognitive Behavioral Humanistic and Interpersonal Training.

[c]CBT: cognitive behavioral therapy.

[d]A fourth arm consisted of both SPARX and school-based CBT but was not analyzed as part of this review, as for our purposes it was a combined treatment and control group.

[e]iCBT: internet-based CBT.

Study Quality

Study quality was mixed, with 31% (5/16) of the studies rated as good overall, 13% (2/16) rated as fair, and 56% (9/16) rated as poor (Table 2). There was a low risk of bias in 88% (14/16) of the included studies for the completeness of outcome data and other biases, 81% (13/16) for random sequence generation, and 56% (9/16) for allocation concealment and blinding of the participants and personnel. There was a low risk of bias among only 50% (8/16) of studies for the blinding of outcome assessment and 19% (3/16) of studies for selective reporting. The risk of bias in the remaining studies was typically unclear.

Table 2. Results of risk of bias assessment.

Study	Random sequence generation	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Other bias	Overall study quality
Calear et al [42]	+[a]	+	?[b]	?	+	?	+	Poor
Fleming et al [43]	+	+	+	+	+	+	?	Good
Ip et al [44]	+	+	+	+	+	?	+	Good
Merry et al [45]	+	+	+	+	–[c]	+	+	Fair
Poppelaars et al [46]	+	?	+	+	+	?	+	Poor
Sekizaki et al [47]	–	?	–	–	+	?	+	Poor
Smith et al [48]	+	?	?	?	+	?	+	Poor
Spence et al [39]	+	?	+	+	+	?	+	Fair
Sportel et al [49]	+	+	+	+	+	+	+	Good
Stallard et al [50]	?	+	?	?	+	?	+	Poor
Stasiak et al [51]	+	+	+	+	+	?	+	Good
Stjerneklar et al [40]	+	+	+	+	+	?	+	Good
Topooco et al [52]	+	+	+	–	+	?	+	Poor
Wong et al [53]	?	?	?	?	–	?	+	Poor
Wright et al [54]	+	?	?	?	+	?	+	Poor
Wuthrich et al [41]	+	?	?	?	+	?	–	Poor

[a]+: low risk of bias.

[b]?: unclear risk of bias.

[c]–: high risk of bias.

Efficacy of cCBT

In all, 19% (3/16; 2 of poor quality and 1 of good quality) did not report adequate data to be included in the meta-analyses [43,47,50]. Here, we briefly summarized the findings of these studies. Compared with waitlist control groups, Sekizaki et al [47] found better depression and anxiety symptoms following cCBT, whereas Fleming et al [43] found this only for depression. Stallard et al [50] found symptom improvements in both cCBT and waitlist control groups but did not conduct sufficient analyses to assess whether either group had better outcomes than the other group. For the remaining studies, we present the meta-analysis results separately for anxiety and depression outcomes. The data used in the meta-analyses can be found in Multimedia Appendix 5 (Table S5) [39-42,44-46,48,49,51-54].

Anxiety Meta-analysis

A total of 11 studies were included in the anxiety meta-analysis (3/11, 27% good quality; 2/11, 18% fair quality; and 6/11, 55% poor quality) [39-42,44,45,48,49,52-54]. The pooled SMD for the anxiety random-effects meta-analysis demonstrated a small but statistically significant effect of cCBT, with treatment arms showing lower anxiety scores at follow-up than the control arms (SMD −0.21, 95% CI −0.33 to −0.09; Figure 2). There was evidence of moderate, but not statistically significant, heterogeneity (I^2=36.2%; Q_{10}=15.68; P=.11). Visual inspection of the funnel plot (Figure 3) and the Egger test for funnel plot asymmetry (P=.80) did not show strong evidence of publication bias.

Figure 2. Forest plot for anxiety meta-analysis. SMD: standardized mean difference.

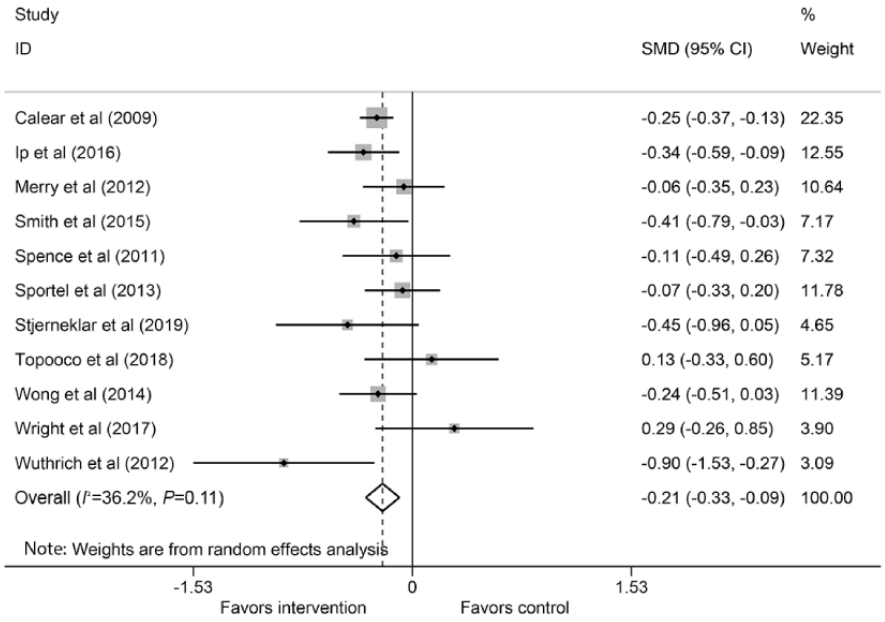
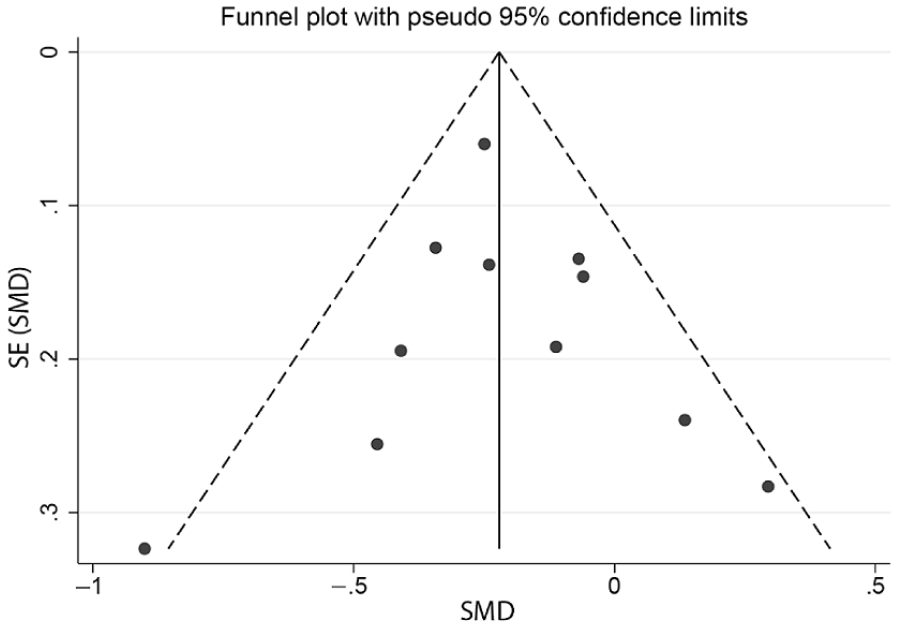


Figure 3. Funnel plot for anxiety meta-analysis. SMD: standardized mean difference.



Sensitivity analyses yielded similar small effect sizes when the meta-analysis was limited to different intervention, randomization, and analysis types (Table 3). However, the pooled effect size varied according to the risk of bias ratings, with those rated *fair* finding the weakest evidence for an effect of cCBT (Figure S1 in Multimedia Appendix 6 [39-42,44,45,48,49,52-54]). The pooled effect size was substantially smaller and not significant when the control groups were limited to those with an active treatment component.

Table 3. Results of stratified anxiety meta-analyses.

Stratified analyses	Quality of the included studies	Total studies, n	Standardized mean difference (95% CI)	Cochran *Q* (df)	Cochran *Q*, *P* value	I^2 (%)
Randomization type						
Clustered [42,49,53]	1 good; 2 poor	3	−0.22 (−0.32 to −0.12)	1.52 (2)	.47	0.0
Other [39-41,44,45,48,52,54]	2 good; 2 fair; 4 poor	8	−0.22 (−0.42 to −0.02)	14.16 (7)	.048	50.6
Analysis type						
Intention to treat [39,44,45,52]	1 good; 2 fair; 1 poor	4	−0.14 (−0.34 to 0.05)	4.10 (3)	.25	26.9
Other or unclear [40-42,48,49,53,54]	2 good; 5 poor	7	−0.25 (−0.41 to −0.09)	10.84 (6)	.09	44.7
Intervention type						
Treatment [39-42,45,48,49,52,54]	2 good; 2 fair; 5 poor	9	−0.18 (−0.33 to −0.03)	14.60 (8)	.07	45.2
Prevention [44,53]	1 good; 1 poor	2	−0.30 (−0.48 to −0.11)	0.30 (1)	.59	0.0
Control group type						
Treatment [39,45,49]	3 good; 1 fair; 5 poor	3	−0.07 (−0.26 to 0.12)	0.78 (2)	.68	0.0
Other [39-42,44,48,49,52-54]	1 good; 2 fair	10	−0.23 (−0.36 to 0.10)	14.31 (9)	.11	37.1
Study quality						
Good [40,44,49]	3 good	3	−0.25 (−0.47 to −0.03)	2.98 (2)	.23	32.8
Fair [39,45]	2 fair	2	−0.08 (−0.31 to 0.15)	0.05 (1)	.83	0.0
Poor [41,42,48,52-54]	6 poor	6	−0.22 (−0.43 to −0.02)	10.97 (5)	.05	54.4

Depression Meta-analysis

A total of 10 studies were included in the depression meta-analysis (3/10, 30% good; 1/10, 10% fair; and 6/10, 60% poor) [40,42,44-46,48,51-54]. As with the anxiety meta-analysis, the pooled SMD for the depression random-effects meta-analysis demonstrated a small but statistically significant effect of cCBT, with treatment arms showing lower depression scores at follow-up than the control arms (SMD −0.23; 95% CI −0.39 to −0.07; Figure 4). There was evidence of moderate and statistically significant heterogeneity (I^2=59.5%; Q_9=22.2; P=.008). Visual inspection of the funnel plot (Figure 5) and the Egger test for funnel plot asymmetry (P=.53) did not show strong evidence of publication bias.

Figure 4. Forest plot for depression meta-analysis. SMD: standardized mean difference.

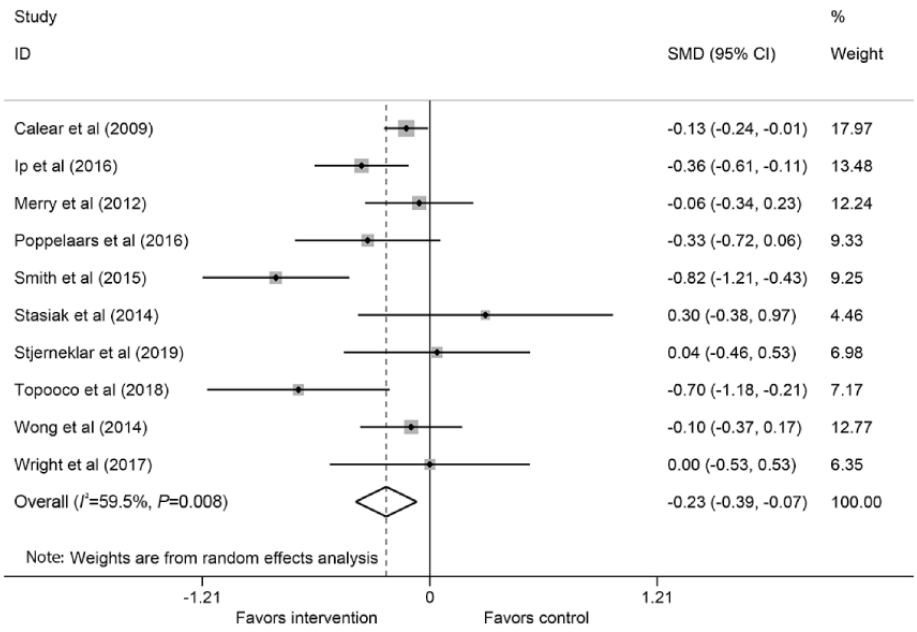
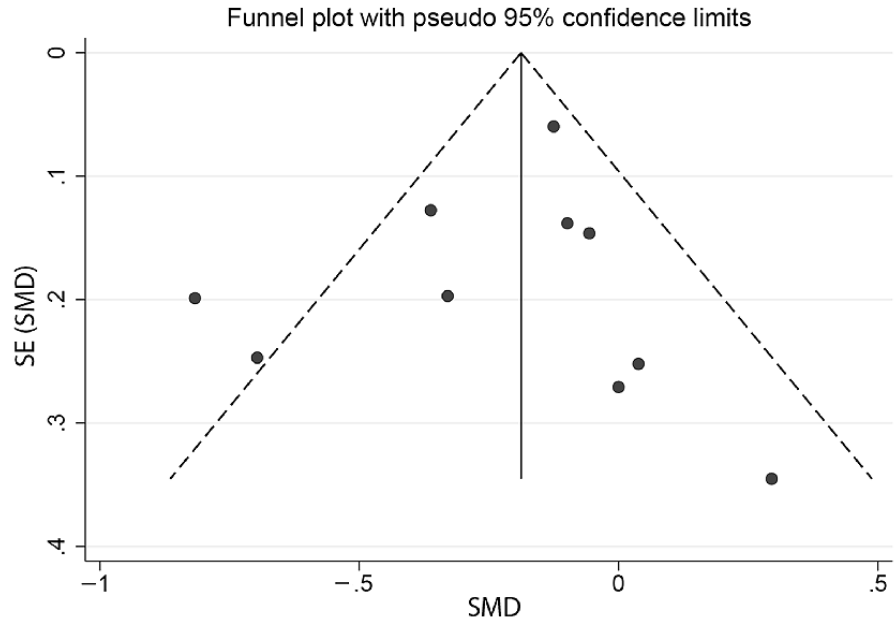


Figure 5. Funnel plot for depression meta-analysis. SMD: standardized mean difference.



Sensitivity analyses yielded similar small effect sizes when the meta-analysis was limited to different intervention types, randomization types, and analysis types (Table 4). The effect size again varied according to the risk of bias ratings, with those rated *poor* finding the strongest evidence for an effect of cCBT compared with control arms (note that only one study was rated *fair* in this meta-analysis; Figure S2 in Multimedia Appendix 7 [40,42,44-46,48,51-54]). Finally, as with anxiety, the effect size was substantially smaller and not significant when the control groups were limited to those with an active treatment component.

Table 4. Results of stratified depression meta-analyses.

Stratified analyses	Quality of the included studies	Total number of studies, N	Standardized mean difference (95% CI)	Cochran *Q* (df)	Cochran *Q*, *P* value	I^2 (%)
Randomization type						
Clustered [42,46,53]	3 poor	3	−0.14 (−0.24 to −0.03)	1.07 (2)	.59	0.0
Other [40,44,45,48,51,52,54]	3 good; 1 fair; 3 poor	7	−0.26 (−0.53 to 0.01)	18.32 (6)	.005	67.2
Analysis type						
Intention to treat [44,45,51,52]	2 good; 1 fair; 1 poor	4	−0.24 (−0.56 to −0.07)	8.29 (3)	.04	63.8
Other or unclear [40,42,46,48,53,54]	1 good; 5 poor	6	−0.22 (−0.43 to −0.01)	13.14 (5)	.02	62.0
Intervention type						
Treatment [40,42,45,48,51,52,54]	2 good; 1 fair; 4 poor	7	−0.22 (−0.46 to 0.03)	19.18 (6)	.004	68.7
Prevention [44,46,53]	1 good; 2 poor	3	−0.26 (−0.43 to −0.09)	2.12 (2)	.35	5.9
Control group type						
Treatment [45,46]	1 fair; 1 poor	2	−0.17 (−0.49 to 0.14)	1.47 (1)	.23	32.1
Other [40,42,44,46,48,51-54]	3 good; 6 poor	9	−0.25 (−0.43 to −0.07)	21.01 (8)	.007	61.9
Study quality						
Good [40,44,51]	3 good	3	−0.10 (−0.49 to 0.29)	4.52 (2)	.11	55.7
Fair [45]	1 fair	1	−0.06 (−0.34 to 0.23)	N/A[a]	N/A	N/A
Poor [42,46,48,52-54]	6 poor	6	−0.32 (−0.56 to −0.08)	16.74 (5)	.005	70.1

[a]N/A: not applicable. Only 1 study was included in this stratification.

Discussion

Principal Findings

This review aims to provide an update to previous meta-analyses and evaluate the efficacy of cCBT in the treatment of anxiety and depression in adolescents. We found evidence of small but significant effects of cCBT on adolescent anxiety and depression. Secondary analyses suggested that the efficacy of cCBT was comparable with that of alternative, active interventions (such as face-to-face therapy or treatment as usual) and resulted in significantly greater symptom reductions than the control groups not receiving such interventions. However, it should also be noted that we identified a large number of poor-quality studies, which could limit the strength of our overall findings, and it would be important for future RCTs in this area to adopt rigorous methodologies to ensure their reliability and validity.

To the best of our knowledge, this is the first meta-analysis of this kind to focus exclusively on the adolescent age range to establish the efficacy of cCBT within this distinct developmental period. This is an important addition to the literature, given the unique biological and social transitions associated with adolescence, as well as the associated prevalence rates of mental health disorders in this age group [24-26]. Limiting the included studies to RCTs further strengthens the evidence for the efficacy of cCBT [55]. Our findings, therefore, support the results of a previous review which found favorable effects of cCBT for the treatment of anxiety and depression in adolescents [18] and support a recent move by the NICE guidelines to recommend the use of digital CBT for children and young people with mild depression [7]. Our findings are also similar to those seen in the adult population; for example, a recent meta-analysis found that internet-based CBT or cCBT was effective for treating depression and anxiety among adults and showed equivalent effects to face-to-face therapy [56]. Furthermore, we identified studies from a wide range of countries, suggesting that cCBT might be an effective intervention in various cultural settings. However, we did not identify any studies published in languages other than English or conducted in low- and middle-income countries, where a lack of resources and mental health services might make the availability of cCBT particularly beneficial.

Challenges remain in the development of effective digital interventions for adolescents; however, our findings suggest that cCBT may overcome various barriers to treatment in the adolescent age group [14,15] and provide reasonable grounds for optimism in promoting adolescent public health. It was found that cCBT may be a less resource-intensive alternative to traditional therapies, with many programs included in this meta-analysis requiring little or no clinician support. However, receiving professional support alongside computerized therapies might be beneficial, such that blended approaches could be optimal [57]. In addition, parental involvement in child CBT may lead to better outcomes, although evidence on their role in adolescent CBT is less conclusive [58,59]. Few of the studies we identified described direct parental involvement in the cCBT intervention; for those that did, the extent of parental involvement was minimal and sometimes optional. Therefore, the potential role of therapists and parents in adolescent cCBT is an important area for future work to optimize the balance of effectiveness, resource requirements, and scalability. It should also be noted that only 6% (1/16) of the included studies included a health economic analysis; a thorough investigation of the cost-effectiveness of cCBT interventions would be an important area for future studies [54].

The feasibility and acceptability of cCBT are not universally positive, necessitating a thorough understanding of the active components of these interventions [14]. Indeed, this issue is not unique to digital interventions targeting anxiety and depression or to adolescents. Going forward, more studies should demonstrate an awareness of the importance of co-design, personalization, and data privacy, and should incorporate principles from various disciplines such as app design and machine learning to further improve the feasibility and acceptability of these interventions [60-62]. Although this meta-analysis did not explore the quality or usability of interventions, future authors assessing cCBT should also consider using process evaluation measures [63], as has been previously highlighted [64]. This would also help prepare interventions for adoption in schools or routine clinical care, as several of the interventions identified in this study are not currently being implemented outside of a research setting.

Limitations

This study had several limitations. In the meta-analysis, we were unable to correct for the effects of clustered randomization because of underreporting in many of the included clustered RCTs. Therefore, the effects of clustering were explored using sensitivity analyses, and effect sizes were found to be similar following the removal of clustered RCTs. In addition, the results from the longest available follow-up period were used in the meta-analysis to ensure that prolonged intervention effects could be captured. However, because the longest follow-up period varies among studies, this can introduce additional heterogeneity (Cochrane Handbook section 9.3.4 [31]). Heterogeneity was generally moderate to high, suggesting that some methodological variation among the included studies was not accounted for. Further heterogeneity may have been introduced by combining prevention and treatment studies and by combining superiority and noninferiority studies; nonetheless, we conducted sensitivity analyses to understand the differential effects of these study designs.

We focused on participants aged 11 to 19 years but acknowledge that the age of adolescence is widely debated, such that some of the studies excluded from this review owing to the age of their participants might have produced findings relevant to adolescent groups [24]. The authors of included studies were not contacted to identify further studies for inclusion; however, our search strategy was otherwise rigorous. Finally, the strength of our findings is necessarily limited by the quality of the included studies, which were very mixed, with more than half (9/16, 56%) receiving overall ratings of poor study quality according to our quality assessments [32,33]. Indeed, evidence for the efficacy of cCBT in treating depression was stronger in studies rated as poor quality.

Conclusions

This meta-analysis reinforces cCBT as an effective intervention for anxiety and depression, showing small, but statistically significant effects. To our knowledge, this is the first study to establish this relationship in an exclusively adolescent group. Most studies were of poor quality and highly heterogeneous, highlighting the need for rigorous and high-quality RCTs in this area. Given the wide variety of available programs and technologies, future research could focus on establishing the active components of cCBT and draw principles from various disciplines, such as design technology and computer science, to optimize feasibility and acceptability. Nonetheless, the clinical potential of cCBT in treating adolescent anxiety and depression is clear and has the scope to address current unmet needs within child and adolescent mental health services.

Acknowledgments

AW is in receipt of a PhD studentship funded by the National Institute for Health Research (NIHR) Biomedical Research Centre in South London and Maudsley National Health Service (NHS) Foundation Trust and King's College London. The views expressed are those of the author and not necessarily those of the NHS, NIHR, or Department of Health and Social Care. JD is supported by the NIHR Clinician Science Fellowship award (CS-2018-18-ST2-014) and has received support from the Medical Research Council Clinical Research Training Fellowship (MR/L017105/1) and Psychiatry Research Trust Peggy Pollak Research Fellowship in Developmental Psychiatry.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOC File , 66 KB - jmir_v24i4e29842_app1.doc]

Multimedia Appendix 2
Search strategy.
[DOCX File , 15 KB - jmir_v24i4e29842_app2.docx]

Multimedia Appendix 3
Included studies' populations and exclusion critieria.
[DOCX File , 17 KB - jmir_v24i4e29842_app3.docx]

Multimedia Appendix 4
Intervention descriptions.
[DOCX File , 24 KB - jmir_v24i4e29842_app4.docx]

Multimedia Appendix 5
Data used in meta-analyses.
[DOCX File , 21 KB - jmir_v24i4e29842_app5.docx]

Multimedia Appendix 6
Anxiety forest plot by study quality.
[DOCX File , 2054 KB - jmir_v24i4e29842_app6.docx]

Multimedia Appendix 7
Depression forest plot by study quality.
[DOCX File , 2108 KB - jmir_v24i4e29842_app7.docx]

References

1. Merikangas KR, Nakamura EF, Kessler RC. Epidemiology of mental disorders in children and adolescents. Dialogues Clin Neurosci 2009;11(1):7-20 [FREE Full text] [doi: 10.31887/dcns.2009.11.1/krmerikangas] [Medline: 19432384]
2. Maughan B, Collishaw S, Stringaris A. Depression in childhood and adolescence. J Can Acad Child Adolesc Psychiatry 2013;22(1):35-40 [FREE Full text] [Medline: 23390431]

3. Wickersham A, Dickson H, Jones R, Pritchard M, Stewart R, Ford T, et al. Educational attainment trajectories among children and adolescents with depression, and the role of sociodemographic characteristics: longitudinal data-linkage study. Br J Psychiatry 2021;218(3):151-157 [FREE Full text] [doi: 10.1192/bjp.2020.160] [Medline: 33028438]
4. Wickersham A, Sugg HV, Epstein S, Stewart R, Ford T, Downs J. Systematic review and meta-analysis: the association between child and adolescent depression and later educational attainment. J Am Acad Child Adolesc Psychiatry 2021;60(1):105-118 [FREE Full text] [doi: 10.1016/j.jaac.2020.10.008] [Medline: 33130250]
5. Woodward LJ, Fergusson DM. Life course outcomes of young people with anxiety disorders in adolescence. J Am Acad Child Adolesc Psychiatry 2001;40(9):1086-1093. [doi: 10.1097/00004583-200109000-00018] [Medline: 11556633]
6. Creswell C, Waite P, Cooper PJ. Assessment and management of anxiety disorders in children and adolescents. Arch Dis Child 2014;99(7):674-678 [FREE Full text] [doi: 10.1136/archdischild-2013-303768] [Medline: 24636957]
7. Depression in children and young people: identification and management. NICE guideline [NG134]. National Institute for Health and Care Excellence. 2019. URL: https://www.nice.org.uk/guidance/ng134 [accessed 2022-03-22]
8. Ishikawa SI, Okajima I, Matsuoka H, Sakano Y. Cognitive behavioural therapy for anxiety disorders in children and adolescents: a meta-analysis. Child Adolesc Ment Health 2007;12(4):164-172. [doi: 10.1111/j.1475-3588.2006.00433.x] [Medline: 32811007]
9. Klein JB, Jacobs RH, Reinecke MA. Cognitive-behavioral therapy for adolescent depression: a meta-analytic investigation of changes in effect-size estimates. J Am Acad Child Adolesc Psychiatry 2007;46(11):1403-1413 [FREE Full text] [doi: 10.1097/chi.0b013e3180592aaa] [Medline: 18049290]
10. Kataoka SH, Zhang L, Wells KB. Unmet need for mental health care among U.S. children: variation by ethnicity and insurance status. Am J Psychiatry 2002;159(9):1548-1555. [doi: 10.1176/appi.ajp.159.9.1548] [Medline: 12202276]
11. O'Brien D, Harvey K, Howse J, Reardon T, Creswell C. Barriers to managing child and adolescent mental health problems: a systematic review of primary care practitioners' perceptions. Br J Gen Pract 2016;66(651):e693-e707 [FREE Full text] [doi: 10.3399/bjgp16X687061] [Medline: 27621291]
12. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
13. Children and parents: media use and attitudes report 2018. Ofcom. 2019. URL: https://www.ofcom.org.uk/research-and-data/media-literacy-research/childrens/children-and-parents-media-use-and-attitudes-report-2018 [accessed 2019-10-07]
14. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual research review: digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
15. Khanna M, Aschenbrand SG, Kendall PC. New frontiers: computer technology in the treatment of anxious youth. Behav Ther 2007;30(1):22-25.
16. Donovan CL, Spence SH, March S. Using new technologies to deliver cognitive behaviour therapy with children and adolescents. In: Graham P, Reynolds S, editors. Cognitive behaviour therapy for children and families. 3rd edition. Cambridge, UK: Cambridge University Press; 2013:351-370.
17. Calear AL, Christensen H. Review of internet-based prevention and treatment programs for anxiety and depression in children and adolescents. Med J Aust 2010;192(S11):S12-S14. [doi: 10.5694/j.1326-5377.2010.tb03686.x] [Medline: 20528700]
18. Ebert DD, Zarski AC, Christensen H, Stikkelbroek Y, Cuijpers P, Berking M, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials. PLoS One 2015;10(3):e0119895 [FREE Full text] [doi: 10.1371/journal.pone.0119895] [Medline: 25786025]
19. Grist R, Croker A, Denne M, Stallard P. Technology delivered interventions for depression and anxiety in children and adolescents: a systematic review and meta-analysis. Clin Child Fam Psychol Rev 2019;22(2):147-171 [FREE Full text] [doi: 10.1007/s10567-018-0271-8] [Medline: 30229343]
20. Pennant ME, Loucas CE, Whittington C, Creswell C, Fonagy P, Fuggle P, Expert Advisory Group. Computerised therapies for anxiety and depression in children and young people: a systematic review and meta-analysis. Behav Res Ther 2015;67:1-18. [doi: 10.1016/j.brat.2015.01.009] [Medline: 25727678]
21. Podina IR, Mogoase C, David D, Szentagotai A, Dobrean A. A meta-analysis on the efficacy of technology mediated CBT for anxious children and adolescents. J Ration Emot Cogn Behav Ther 2016;34(1):31-50. [doi: 10.1007/s10942-015-0228-5]
22. Richardson T, Stallard P, Velleman S. Computerised cognitive behavioural therapy for the prevention and treatment of depression and anxiety in children and adolescents: a systematic review. Clin Child Fam Psychol Rev 2010;13(3):275-290. [doi: 10.1007/s10567-010-0069-9] [Medline: 20532980]
23. Rooksby M, Elouafkaoui P, Humphris G, Clarkson J, Freeman R. Internet-assisted delivery of cognitive behavioural therapy (CBT) for childhood anxiety: systematic review and meta-analysis. J Anxiety Disord 2015;29:83-92. [doi: 10.1016/j.janxdis.2014.11.006] [Medline: 25527900]
24. Sawyer SM, Azzopardi PS, Wickremarathne D, Patton GC. The age of adolescence. Lancet Child Adolesc Health 2018;2(3):223-228. [doi: 10.1016/S2352-4642(18)30022-1] [Medline: 30169257]
25. Costello EJ, Mustillo S, Erkanli A, Keeler G, Angold A. Prevalence and development of psychiatric disorders in childhood and adolescence. Arch Gen Psychiatry 2003;60(8):837-844. [doi: 10.1001/archpsyc.60.8.837] [Medline: 12912767]

26. Merikangas KR, He JP, Burstein M, Swanson SA, Avenevoli S, Cui L, et al. Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication--Adolescent Supplement (NCS-A). J Am Acad Child Adolesc Psychiatry 2010;49(10):980-989 [FREE Full text] [doi: 10.1016/j.jaac.2010.05.017] [Medline: 20855043]
27. Adults: media use and attitudes report 2019. Ofcom. 2019. URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0021/149124/adults-media-use-and-attitudes-report.pdf [accessed 2019-10-07]
28. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009;151(4):264-269 [FREE Full text] [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
29. Adolescent health. World Health Organization. URL: https://www.who.int/health-topics/adolescent-health#tab=tab_1 [accessed 2021-11-06]
30. NICE glossary. National Institute for Health and Care Excellence. 2019. URL: https://www.nice.org.uk/glossary [accessed 2019-10-03]
31. Higgins JP, Green S. Cochrane handbook for systematic reviews of interventions version 5.1.0. London, UK: The Cochrane Collaboration; 2011.
32. 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]
33. Viswanathan M, Patnode CD, Berkman ND, Bass EB, Chang S, Hartling L, et al. Assessing the risk of bias of individual studies in systematic reviews of health care interventions. In: Methods guide for effectiveness and comparative effectiveness reviews. Rockville, MD: Agency for Healthcare Research and Quality; 2017:193-221.
34. Stata statistical software: release 15. StataCorp. College Station, TX: StataCorp; 2017. URL: https://www.stata.com/ [accessed 2022-03-22]
35. Borenstein M, Hedges L, Higgins J, Rothstein H. Effect sizes based on means. In: Borenstein M, Hedges LV, Higgins JP, Rothstein HR, editors. Introduction to meta-analysis. Chichester, UK: John Wiley & Sons; 2009:21-32.
36. Harris RJ, Deeks JJ, Altman DG, Bradburn MJ, Harbord RM, Sterne JA. Metan: fixed- and random-effects meta-analysis. Stata J 2008;8(1):3-28. [doi: 10.1177/1536867x0800800102]
37. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
38. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
39. Spence SH, Donovan CL, March S, Gamble A, Anderson RE, Prosser S, et al. A randomized controlled trial of online versus clinic-based CBT for adolescent anxiety. J Consult Clin Psychol 2011;79(5):629-642. [doi: 10.1037/a0024512] [Medline: 21744945]
40. Stjerneklar S, Hougaard E, McLellan LF, Thastum M. A randomized controlled trial examining the efficacy of an internet-based cognitive behavioral therapy program for adolescents with anxiety disorders. PLoS One 2019;14(9):e0222485 [FREE Full text] [doi: 10.1371/journal.pone.0222485] [Medline: 31532802]
41. Wuthrich VM, Rapee RM, Cunningham MJ, Lyneham HJ, Hudson JL, Schniering CA. A randomized controlled trial of the Cool Teens CD-ROM computerized program for adolescent anxiety. J Am Acad Child Adolesc Psychiatry 2012;51(3):261-270. [doi: 10.1016/j.jaac.2011.12.002] [Medline: 22365462]
42. Calear AL, Christensen H, Mackinnon A, Griffiths KM, O'Kearney R. The YouthMood Project: a cluster randomized controlled trial of an online cognitive behavioral program with adolescents. J Consult Clin Psychol 2009;77(6):1021-1032. [doi: 10.1037/a0017391] [Medline: 19968379]
43. Fleming T, Dixon R, Frampton C, Merry S. A pragmatic randomized controlled trial of computerized CBT (SPARX) for symptoms of depression among adolescents excluded from mainstream education. Behav Cogn Psychother 2012;40(5):529-541. [doi: 10.1017/S1352465811000695] [Medline: 22137185]
44. Ip P, Chim D, Chan KL, Li TM, Ho FK, Van Voorhees BW, et al. Effectiveness of a culturally attuned internet-based depression prevention program for Chinese adolescents: a randomized controlled trial. Depress Anxiety 2016;33(12):1123-1131. [doi: 10.1002/da.22554] [Medline: 27618799]
45. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MF. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. BMJ 2012;344:e2598 [FREE Full text] [doi: 10.1136/bmj.e2598] [Medline: 22517917]
46. Poppelaars M, Tak YR, Lichtwarck-Aschoff A, Engels RC, Lobel A, Merry SN, et al. A randomized controlled trial comparing two cognitive-behavioral programs for adolescent girls with subclinical depression: a school-based program (Op Volle Kracht) and a computerized program (SPARX). Behav Res Ther 2016;80:33-42. [doi: 10.1016/j.brat.2016.03.005] [Medline: 27019280]
47. Sekizaki R, Nemoto T, Tsujino N, Takano C, Yoshida C, Yamaguchi T, et al. School mental healthcare services using internet-based cognitive behaviour therapy for young male athletes in Japan. Early Interv Psychiatry 2019;13(1):79-85. [doi: 10.1111/eip.12454] [Medline: 28613407]

48. Smith P, Scott R, Eshkevari E, Jatta F, Leigh E, Harris V, et al. Computerised CBT for depressed adolescents: randomised controlled trial. Behav Res Ther 2015;73:104-110. [doi: 10.1016/j.brat.2015.07.009] [Medline: 26301756]
49. Sportel BE, de Hullu E, de Jong PJ, Nauta MH. Cognitive bias modification versus CBT in reducing adolescent social anxiety: a randomized controlled trial. PLoS One 2013;8(5):e64355 [FREE Full text] [doi: 10.1371/journal.pone.0064355] [Medline: 23691203]
50. Stallard P, Richardson T, Velleman S, Attwood M. Computerized CBT (Think, Feel, Do) for depression and anxiety in children and adolescents: outcomes and feedback from a pilot randomized controlled trial. Behav Cogn Psychother 2011;39(3):273-284. [doi: 10.1017/S135246581000086X] [Medline: 21272393]
51. Stasiak K, Hatcher S, Frampton C, Merry SN. A pilot double blind randomized placebo controlled trial of a prototype computer-based cognitive behavioural therapy program for adolescents with symptoms of depression. Behav Cogn Psychother 2014;42(4):385-401. [doi: 10.1017/S1352465812001087] [Medline: 23253641]
52. Topooco N, Berg M, Johansson S, Liljethörn L, Radvogin E, Vlaescu G, et al. Chat- and internet-based cognitive-behavioural therapy in treatment of adolescent depression: randomised controlled trial. BJPsych Open 2018;4(4):199-207 [FREE Full text] [doi: 10.1192/bjo.2018.18] [Medline: 29988969]
53. Wong N, Kady L, Mewton L, Sunderland M, Andrews G. Preventing anxiety and depression in adolescents: a randomised controlled trial of two school based internet-delivered cognitive behavioural therapy programmes. Internet Interv 2014;1(2):90-94. [doi: 10.1016/j.invent.2014.05.004]
54. Wright B, Tindall L, Littlewood E, Allgar V, Abeles P, Trépel D, et al. Computerised cognitive-behavioural therapy for depression in adolescents: feasibility results and 4-month outcomes of a UK randomised controlled trial. BMJ Open 2017;7(1):e012834 [FREE Full text] [doi: 10.1136/bmjopen-2016-012834] [Medline: 28132000]
55. Hariton E, Locascio JJ. Randomised controlled trials - the gold standard for effectiveness research: study design: randomised controlled trials. BJOG 2018;125(13):1716 [FREE Full text] [doi: 10.1111/1471-0528.15199] [Medline: 29916205]
56. 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]
57. Gratzer D, Khalid-Khan F. Internet-delivered cognitive behavioural therapy in the treatment of psychiatric illness. CMAJ 2016;188(4):263-272 [FREE Full text] [doi: 10.1503/cmaj.150007] [Medline: 26527829]
58. Cardy JL, Waite P, Cocks F, Creswell C. A systematic review of parental involvement in cognitive behavioural therapy for adolescent anxiety disorders. Clin Child Fam Psychol Rev 2020;23(4):483-509 [FREE Full text] [doi: 10.1007/s10567-020-00324-2] [Medline: 32862330]
59. Walczak M, Esbjørn BH, Breinholst S, Reinholdt-Dunne ML. Parental involvement in cognitive behavior therapy for children with anxiety disorders: 3-year follow-up. Child Psychiatry Hum Dev 2017;48(3):444-454. [doi: 10.1007/s10578-016-0671-2] [Medline: 27405872]
60. Leightley D, Puddephatt JA, Jones N, Mahmoodi T, Chui Z, Field M, et al. A smartphone app and personalized text messaging framework (InDEx) to monitor and reduce alcohol use in ex-serving personnel: development and feasibility study. JMIR Mhealth Uhealth 2018;6(9):e10074 [FREE Full text] [doi: 10.2196/10074] [Medline: 30206054]
61. Scholten H, Granic I. Use of the principles of design thinking to address limitations of digital mental health interventions for youth: viewpoint. J Med Internet Res 2019;21(1):e11528 [FREE Full text] [doi: 10.2196/11528] [Medline: 31344671]
62. Wickersham A, Petrides PM, Williamson V, Leightley D. Efficacy of mobile application interventions for the treatment of post-traumatic stress disorder: a systematic review. Digit Health 2019;5:2055207619842986 [FREE Full text] [doi: 10.1177/2055207619842986] [Medline: 31019722]
63. Donabedian A. Evaluating the quality of medical care. 1966. Milbank Q 2005;83(4):691-729 [FREE Full text] [doi: 10.1111/j.1468-0009.2005.00397.x] [Medline: 16279964]
64. Khan K, Hall CL, Davies EB, Hollis C, Glazebrook C. The effectiveness of web-based interventions delivered to children and young people with neurodevelopmental disorders: systematic review and meta-analysis. J Med Internet Res 2019;21(11):e13478 [FREE Full text] [doi: 10.2196/13478] [Medline: 31682573]

Abbreviations

CBT: cognitive behavioral therapy
cCBT: computerized CBT
NHS: National Health Service
NICE: National Institute for Health and Care Excellence
NIHR: National Institute for Health Research
PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses
RCT: randomized controlled trial
SMD: standardized mean difference

Edited by A Mavragani; submitted 22.04.21; peer-reviewed by C Hollis, S Butler, K Khan; comments to author 31.05.21; revised version received 30.06.21; accepted 03.12.21; published 11.04.22.

Please cite as:
Wickersham A, Barack T, Cross L, Downs J
Computerized Cognitive Behavioral Therapy for Treatment of Depression and Anxiety in Adolescents: Systematic Review and Meta-analysis
J Med Internet Res 2022;24(4):e29842
URL: https://www.jmir.org/2022/4/e29842
doi:10.2196/29842
PMID:35404263

Review

Patient-Generated Health Photos and Videos Across Health and Well-being Contexts: Scoping Review

Bernd Ploderer[1], PhD; Atae Rezaei Aghdam[2], PhD; Kara Burns[3], PhD

[1]School of Computer Science, Queensland University of Technology, Brisbane, Australia

[2]School of Information Systems, Queensland University of Technology, Brisbane, Australia

[3]Centre for Digital Transformation of Health, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Melbourne, Australia

Corresponding Author:
Bernd Ploderer, PhD
School of Computer Science
Queensland University of Technology
Gardens Point 2 George St
Brisbane, 4001
Australia
Phone: 61 731384927
Email: b.ploderer@qut.edu.au

Abstract

Background: Patient-generated health data are increasingly used to record health and well-being concerns and engage patients in clinical care. Patient-generated photographs and videos are accessible and meaningful to patients, making them especially relevant during the current COVID-19 pandemic. However, a systematic review of photos and videos used by patients across different areas of health and well-being is lacking.

Objective: This review aims to synthesize the existing literature on the health and well-being contexts in which patient-generated photos and videos are used, the value gained by patients and health professionals, and the challenges experienced.

Methods: Guided by a framework for scoping reviews, we searched eight health databases (CINAHL, Cochrane Library, Embase, PsycINFO, PubMed, MEDLINE, Scopus, and Web of Science) and one computing database (ACM), returning a total of 28,567 studies. After removing duplicates and screening based on the predefined inclusion criteria, we identified 110 relevant articles. Data were charted and articles were analyzed following an iterative thematic approach with the assistance of NVivo software (version 12; QSR International).

Results: Patient-generated photos and videos are used across a wide range of health care services (39/110, 35.5% articles), for example, to diagnose skin lesions, assess dietary intake, and reflect on personal experiences during therapy. In addition, patients use them to self-manage health and well-being concerns (33/110, 30%) and to share personal health experiences via social media (36/110, 32.7%). Photos and videos create significant value for health care (59/110, 53.6%), where images support diagnosis, explanation, and treatment (functional value). They also provide value directly to patients through enhanced self-determination (39/110, 35.4%), social (33/110, 30%), and emotional support (21/110, 19.1%). However, several challenges emerge when patients create, share, and examine photos and videos, such as limited accessibility (16/110, 14.5%), incomplete image sets (23/110, 20.9%), and misinformation through photos and videos shared on social media (17/110, 15.5%).

Conclusions: This review shows that photos and videos engage patients in meaningful ways across different health care activities (eg, diagnosis, treatment, and self-care) for various health conditions. Although photos and videos require effort to capture and involve challenges when patients want to use them in health care, they also engage and empower patients, generating unique value. This review highlights areas for future research and strategies for addressing these challenges.

(J Med Internet Res 2022;24(4):e28867) doi:10.2196/28867

KEYWORDS

patient engagement; patient-generated health data; consumer-generated health data; personal health information; patient empowerment; mobile phone

Introduction

Background

There has been a growing interest in patient-generated health data (PGHD) in recent years, where patients create and collect personal information about some aspects of their own health outside the health care setting [1]. This interest has been spurred by technological developments, most notably by sensors embedded in smartphones and wearable devices that allow people to automatically generate a wide range of health data, from physical activity to heart rate to sleep [2-4]. At the same time, patient perspectives are progressively changing from passive recipients of health care to active agents, with an emphasis on proactive well-being, rather than reactive clinical care [5].

Current evidence suggests that for patients, PGHD support the self-management of disease, promote partnership with providers, enable people to gain social support within the peer network, and facilitate the creation of different types of value [6-9]. Health service providers are also increasingly interested in assessing patient health outside the health care setting, for example, through patient-reported outcome measures (PROMs) [10]. In contrast to PROMs, PGHD can be initiated by patients rather than by health care providers. Not only are patients responsible for capturing personal data but they can also direct the sharing of this information and retain ownership of their data [1]. Furthermore, PROMs are often survey-based, whereas PGHD can be diverse, including sensor data, personal diaries, photos, and histories [1,8,11].

This paper focuses on patient-generated photos and videos because they are more accessible and meaningful for patients than other forms of PGHD. First, accessibility stems from the widespread availability of cameras in smartphones, which allows patients to capture photos or videos of their bodies, lifestyles, and experiences relevant for their health and well-being [12]. Photos are also accessible as a medium that patients can readily use and understand across different languages and cultures, without requiring in-depth medical or technical expertise. For example, patients tracking their diet may find it easier to take a photo of each meal consumed than to keep a diary of the ingredients and nutritional value of each meal [13]. In writing that “seeing comes before words,” Berger [14] highlights that photos and videos are accessible on a more fundamental level, because we experience the world, and thereby our health, primarily through our senses, including our visual sense. Second, photos and videos are meaningful for patients because they can communicate something that they cannot directly express, as suggested by Haines et al [15]: “photographs can reveal the gap between ‘what we see and what we know’, and show aspects of experience not easily captured through words alone.” Videos allow patients to discuss and record what they see and experience. Both photos and videos can aid patients in capturing and discussing unique information during consultations, and conversely, they offer prompts to health care professionals to ask questions that may not be asked otherwise. Furthermore, social media (eg, YouTube and PatientsLikeMe) allow patients to share not only data but also personal experiences and knowledge through photos and videos, which make them interesting resources for other patients, health care professionals, and health care organizations [6,16].

During the current COVID-19 global pandemic, the accessibility and meaningfulness of photos and videos are especially relevant. With unprecedented stress on health systems and risk of infection spread, patients and health care providers are looking for tools that are easy to use and accessible for diagnosis and ongoing care via telehealth [17]. However, current systematic reviews on the use of photography and videos for health and well-being concerns have been limited to specific populations [18], a single clinical assessment [13], or one type of content [19]. A comprehensive assessment of the extent of research evidence and the potential scope of patient-generated photos and videos in different areas of health and well-being is lacking.

Objectives

The overarching objective of this review is to synthesize the literature on patient-generated photos and videos across health and well-being contexts. Specific objectives include (1) providing an overview of the different contexts in which photos and videos are used, (2) examining the value gained for patients and health care professionals, and (3) examining the challenges experienced by these groups in creating, sharing, and examining photos and videos. Throughout the review, we examine the differences between photos and videos. On the basis of these insights, this study seeks to offer practical implications for patients and health care professionals, as well as future research directions for medical informatics researchers.

Methods

Overview

This study was guided by the 5-step framework for scoping reviews by Arksey and O’Malley [20]. Scoping reviews aim to comprehensively assess the size and scope of available research literature to convey the breadth of a nascent field. Similar to systematic reviews, scoping reviews aim to be systematic, transparent, and replicable [21]. However, a scoping review protocol has not yet been published. In contrast to systematic reviews, scoping reviews do not assess the quality of included studies because of the paucity of randomized controlled studies [22], and the review also requires analytical reinterpretation of the literature [23]. In the following sections, we describe each of the 5 steps taken to conduct a scoping review of patient-generated photos and videos for health and well-being. For a succinct summary via a scoping review checklist, see Multimedia Appendix 1 [24]. Although the steps are presented in a linear order, it is important to note that the scoping process is iterative and requires a back-and-forth within and between steps as researchers gain a better understanding of the literature [20,22].

Step 1: Identifying the Research Question

The research questions for this review were as follows: (1) *In which health and well-being contexts are patient-generated photos and videos used?* (2) *What value and challenges do patient-generated photos and videos hold for patients and health care professionals?* These questions were based on our shared

interest in photos taken by patients for health, also known as *medical selfies* [12,25]. We refined the questions over time as we became acquainted with the literature to focus on patient-generated photos, rather than selfies, to align with the widespread use of the term *PGHD* in the literature [7,26,27]. Videos were also included because they were similarly captured through smartphones and used in ways similar to photos. Our primary concern has always been with the experiences of patients and their caregivers, as well as their photo-mediated interactions with health care professionals (eg, clinicians, allied health, and nurses) and peers (eg, via social media), rather than a health system or pure technology perspective. On the basis of the literature reviewed, we refined the research question from experience to the more specific study objectives of (1) contexts, (2) value, and (3) challenges.

Step 2: Identifying Relevant Studies

We devised a systematic search strategy to identify relevant studies. The strategy was based on the literature review of the PhD thesis of the third author (KB) and the support of a librarian. The search terms described in Textbox 1 were based on keywords in the research question and were developed in consultation with a research librarian. Full search strings with particular terms for each database can be found in Multimedia Appendix 2.

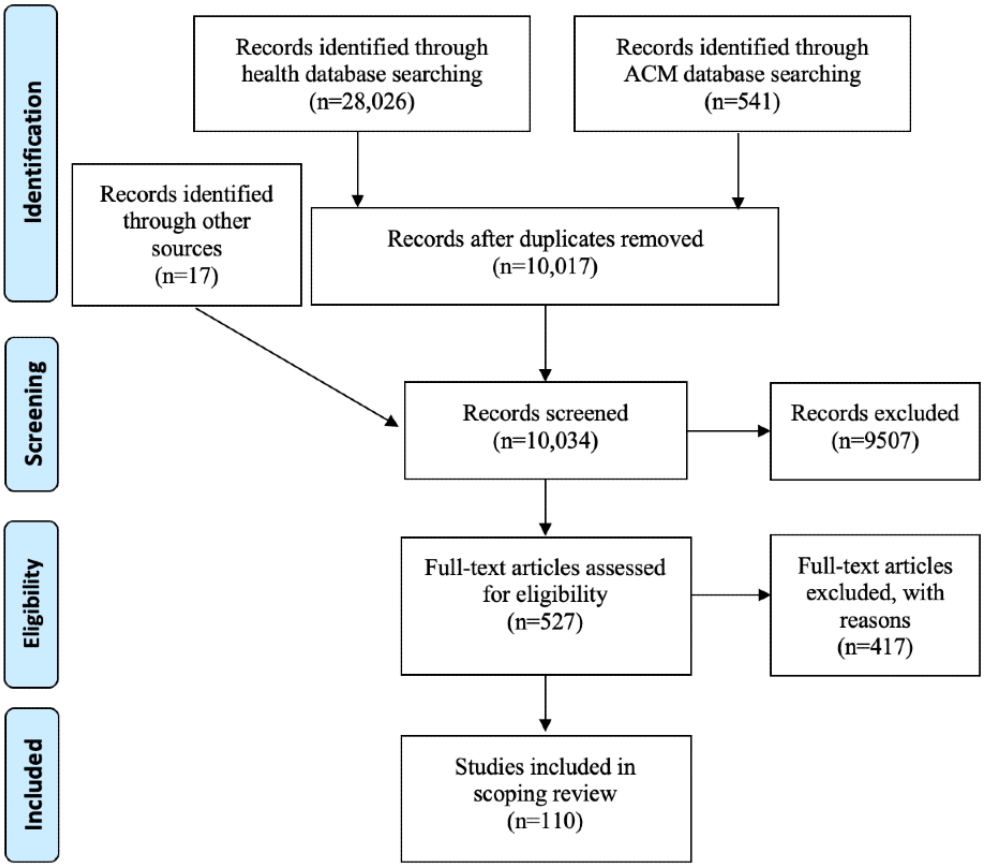
Textbox 1. Search terms.

Search terms
(image* OR pictur* OR photo* OR video* OR selfie* OR portrait* OR snap* OR shot* OR depict* OR data* OR info*)
AND
(patient* OR consumer* OR care* OR customer* OR veteran* OR client* OR self* OR crowd*)
AND
(generate* OR record* OR creat* OR captur* OR document* OR evidence* OR story OR report* OR track* OR initiat* OR monitor* OR take*)

The search included articles from January 2008 to January 31, 2021, written in the English language. The start date was chosen because the major brands of smartphones—iPhone (Apple Inc) and Android (Open Handset Alliance)—were first released in 2007 and 2008, respectively, which provide the platform for patient-generated photos and videos. Articles written in other languages were excluded because of the cost and time required for translation. Only peer-reviewed articles that included primary research were selected to ensure that the conclusions were supported by an evidence base.

According to the objectives of this study and the focus on patients as technology users, we conducted our search strategy in both health and computing databases. Furthermore, we considered social science databases such as Embase and PsycINFO to cover special studies in psychology and behavioral science. We searched eight health databases (CINAHL, Cochrane Library, Embase, PsycINFO, Web of Science, PubMed, MEDLINE, and Scopus) and one computing database (ACM). To ensure that we did not neglect any relevant articles, we broadened the search by using Medical Subject Headings terms and synonyms to collect a comprehensive pool of relevant articles. As illustrated in Figure 1, the health database search yielded 28,026 results, and the ACM search yielded 541 results. In addition, 2 authors (BP and KB) hand searched the reference lists of related review articles [7,13,18,19,28] and JMIR archives, which returned 17 additional articles. After removing duplicates, 10,017 articles remained.

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



Step 3: Study Selection

The study selection was performed by 2 authors (BP and KB) based on inclusion and exclusion criteria to ensure consistency and replicability. As illustrated in Figure 1, we followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) procedure [29] to ensure systematic selection. On the basis of the literature review and discussion of the research question, we established several inclusion criteria (Textbox 2). Articles needed to fulfill all the inclusion criteria to be included in the review. After screening the titles of articles, 527 remained for review. The 2 authors independently reviewed the abstracts of each remaining article and, if necessary, downloaded the entire article to check if it fit the criteria. Papers that were potentially eligible were discussed during meetings among the authors. Through these discussions, we also established exclusion criteria (Textbox 2) to disambiguate decisions on potentially relevant articles.

Textbox 2. Inclusion and exclusion criteria for study selection.

Inclusion criteria

1. Articles describe patient-generated photography or videos that reflect personal information and experiences to help address a health and well-being concern
2. Photos or videos are taken by patients, carers, or other participants who are not health care professionals or researchers
3. Findings report on photography or videos as a collection mechanism, intervention, or unit of analysis

Exclusion criteria

1. Publications without primary research, such as editorials, opinions, perspectives, reviews, and research protocols
2. Secondary analysis of photos and videos, for example, from social media, that have been shared by individuals without an explicit health or well-being intent
3. Automatic video recordings of consultations or teleconsultations as well as images generated by clinician, surveillance, and patient monitoring systems

On the basis of this process, we selected 110 relevant articles for inclusion in this review. Owing to the large number of papers involved, we only kept track of the number of papers at each stage of the selection process, but we did not record the reason for excluding each paper.

Step 4: Charting the Data

NVivo (version 12; QSR International), a qualitative data analysis software package, was used to store and manage the charted data. Initially, we charted the data in a predefined form (Multimedia Appendix 3 [15,30-138]), collecting publication data to allow numerical coding and extracting qualitative information relevant to our research questions (eg, author information, year published, aims, target group, research methods, results, number of photos or videos, and values). However, with the large number of articles involved, the diversity of studies, and particularly the breadth of qualitative results presented, the spreadsheet became impractical.

To manage the large volume of data generated through charting, NVivo (version 12) was used to code the content from the PDF version of each article. This process also enabled the next step of collating results. Publication data were extracted verbatim from each article by 2 authors (BP and ARA), whereas coding and critical analysis for the research questions was completed by all authors. Extracted information was discussed at regular meetings of all authors to ensure that the research questions were still relevant, and the articles could answer the research questions and to explore any discrepancies to clarify key concepts and identify major gaps.

Step 5: Collating, Summarizing, and Reporting Results

Following the recommendation of Arksey and O'Malley [20], we collated and reported the results based on a thematic analysis approach [139] with an analytic framework [140]. Our thematic analysis followed the steps described by Braun and Clarke [139]. We started by reading articles to familiarize ourselves with the data, recorded notes through the *memo* and annotation features of NVivo, and discussed ideas for coding. One author (ARA) manually coded a subset of the 110 articles to generate an initial list of 102 codes relevant to our research questions of health and well-being contexts, value generated, and challenges. These initial codes gave us an overview of the data, but they also highlighted the diversity of study designs and results, which made the aggregation of findings impossible. Instead, we needed a framework to structure and report the results according to our research questions.

To structure the results around health contexts, we initially coded articles according to the International Classification of Diseases, 10th revision [141], a medical classification established by the World Health Organization consisting of 21 chapters. For example, chapter 1 describes infectious and parasitic diseases, which relates to photos and videos used to describe vaccine information and experiences. However, we found that this framework was limited because it presented a medical perspective and did not fit well with articles that reported well-being outcomes or social media contexts. Hence, we revised the structure around the primary contexts presented in the articles: (1) health care services, where patients share images with a health care professional to observe and treat health and well-being concerns; (2) self-management, where patients use images to independently track and manage health concerns; (3) social media, where patients share personal health information and experiences with peers on the web; (4) education, where images are used for health education in schools and waiting rooms; and (5) service improvement, where patients are invited to take images to reflect on their health service experience and express their needs.

To analyze the value of photos and videos reported in our article collection, we used a health consumer engagement framework [140] that highlights six key values of PGHD: functional, emotional, social, transactional, efficiency, and self-determination. For example, the functional value describes how images are used by health care professionals to support health outcomes through diagnosis, explanation, treatment, therapy, and health promotion. The values from this framework were chosen because they originated from a study of patient-generated photos and allowed value to be considered from the perspective of both patients and providers across different health and well-being domains. We chose this framework over benefit-risk models of the health care value, which aim to promote strategic reform [142,143], because photographs and videos are not routinely used in clinical practice, and quantification of value was not demonstrated in the articles retrieved.

To analyze these challenges, we identified several frameworks that describe data challenges [27,144]. Although none of these frameworks captured the range of challenges identified in our initial codes, we selectively applied relevant concepts from these frameworks for our analysis. For example, accessibility is a key challenge for patients [144], which includes lack of access to camera phones, lack of access due to poor app usability, and difficulty in taking photos of feet or the back. From existing frameworks [27,144], we also included the challenges of privacy, interpretability, and relevancy, and we structured the challenges according to different stages of their use: collection, sharing, and examination of photos and videos. In addition, we inductively coded other challenges that emerged from the articles, such as poor photo quality when photos were not in focus or when they did not clearly show the relevant details.

A selection of 15 articles was coded independently by all 3 authors using the chosen frameworks. Regular meetings were held to discuss the suitability of the frameworks for our objectives and to explore any discrepancies in how we applied them in our analysis, especially on how to distinguish between values that appear interrelated (eg, the social and transactional values). Once agreement was reached on how to apply the frameworks and how to structure the challenges, one author (BP) coded the remaining papers. The naming of themes and subthemes was further refined by all the authors while writing the report. The full coding tree is provided in Multimedia Appendix 4. The results present the overall number of articles identified in each theme, as well as the number of articles reporting on photos and videos.

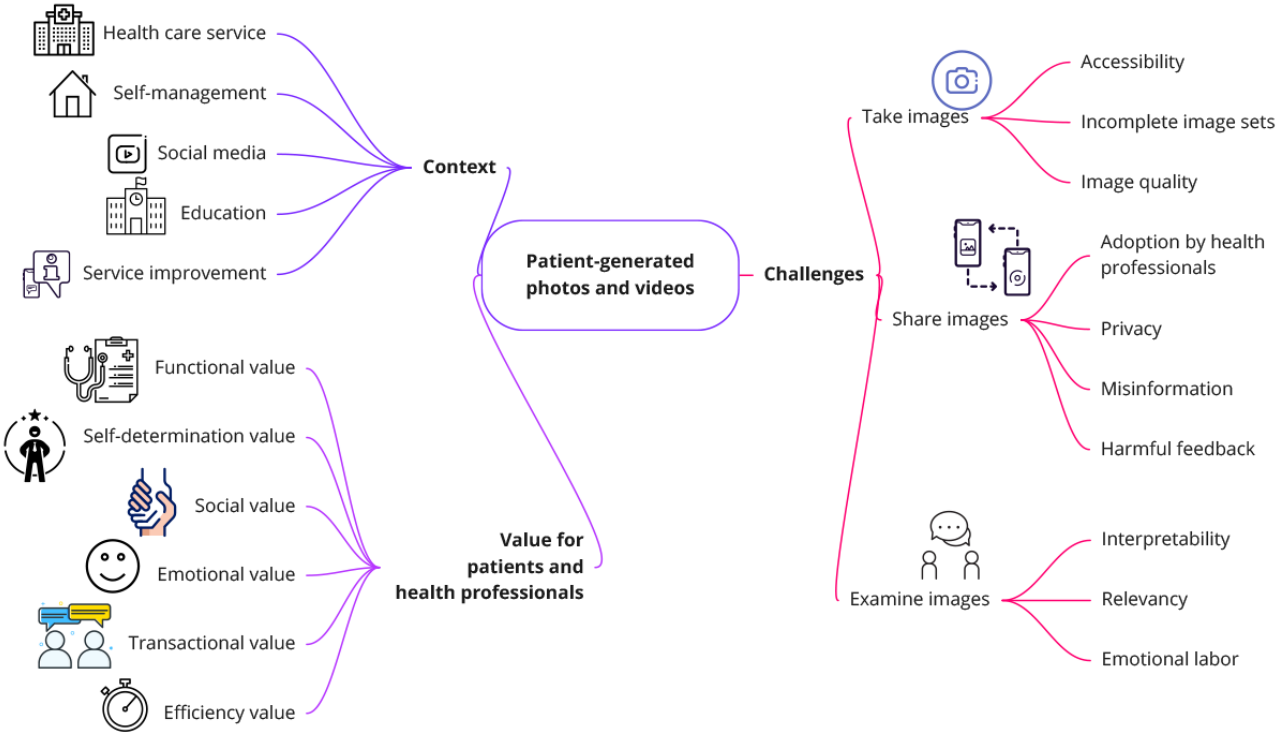
Results

Overview

Of the 110 articles identified in this review, 90 (81.8%) reported on photos, 23 (20.9%) used videos, and 3 (2.7%) used both photos and videos. Figure 2 provides an overview of the key

themes revealed in our review, showing the contexts in which photos and videos were used, values gained by patients, and challenges when taking, sharing, and examining photos and videos. The following sections provide further details of each theme.

Figure 2. Overview of key themes identified in this review, presenting the contexts in which patients use health photos and videos, the value gained by patients, and challenges experienced.



Use of Photos and Videos Across Health Contexts

We categorized articles based on the context in which the patient-generated photos and videos were used. As summarized in Table 1, images were largely used in health care services, self-management, and social media contexts. Multimedia Appendix 3 provides a more detailed table that also lists who captured the images (patient or carer), the technologies used to capture and share images, and the audiences receiving them.

Table 1. The use of patient-generated photos and videos across health contexts (N=110 articles).[a]

Context	Description	Articles, n (%)	Photos, n (%)	Videos, n (%)	Image information
Health care service	Patients share images with a health care service to observe and treat health and well-being concerns	39 (35.5)	39 (35.5)	2 (1.8)	Skin photos showing potential cancer [30-34], hernia [35], rash [36-38], and wounds [39-50]; foods and beverages consumed [51-62]; experiences related to mental health (eg, death of a parent) [63], emotions such as hope [64], goals for the near future [65] for therapy; and health equipment [66,67] and medication [68]
Self-management	Patients use images to independently track and manage health concerns	33 (30)	33 (30)	1 (0.9)	Foods and beverages consumed [69-92]; nature, people, and events to reflect on emotions [93,94] and lifestyle [95-97]; and smoking and quitting [15,98-100]
Social media	Patients share personal health information and experiences with peers on the web on platforms such as Instagram, Facebook, Flickr, and YouTube	36 (32.7)	19 (17.3)	17 (15.5)	Foods and beverages consumed [88-92]; disease experience (cancer [101-106], cardiovascular [107], diabetes [104,108], kidney stone [109], and multiple sclerosis [110]); mental health (depression [111-113], suicidal thought [114], and other [93,115,116]); proanorexia images [117-120]; medical procedures [121,122]; smoking and quitting [99]; vaccine information [123-127]; and various health concerns [97,128]
Education	Images are used for health education in schools, waiting rooms, and community centers or at home	7 (6.4)	6 (5.5)	1 (0.9)	Digitally altered selfies showing impact of smoking [129,130] and UV exposure [131,132]; healthy eating ideas [133]; toothbrushing behavior videos [134]; and vagina selfies [135]
Service improvement	Patients are invited to take images to reflect on their health service experience and to express their needs	3 (2.7)	1 (0.9)	2 (1.8)	Children's experiences and challenges in the hospital (eg, needing to process new information) [136,137] and in transitioning to their homes (eg, manage medications) [138]

[a]Several articles reported results on multiple contexts, or they included both photos and videos.

Health care service contexts were described in 35.4% (39/110) of the articles, where patients created photos to document a health concern to share them with a health care professional. The three most common contents in this context were skin photos, food photos, and photos capturing mental health experiences. Patient-generated skin photos were used by dermatologists and general practitioners to review skin lesions and assess potential melanoma [30-34] or rashes [36-38]. Surgeons have also used patient-generated photos to diagnose inguinal hernia [35] and to follow up on surgical wounds or injuries [39-50]. Patient-generated photos showing food and beverages consumed were commonly used by dietitians to support patients with diabetes [54,57,58,60,61], patients with irritable bowel syndrome [53], and pregnant women [51]. Therapists and counselors collaborated with patients to discuss photos capturing events and experiences that affected the patient's mental health (eg, the death of a parent) [63], emotions such as hope [64], and goals for the near future (eg, to go on a holiday) [65]. In addition, of the 110 studies, 2 (1.8%) used wearable cameras that automatically took photos throughout the day to document dietary intake [55,59], which provided the dietitian with more comprehensive data and alleviated the effort required for patients. Only 1.8% (2/110) of the articles in this context used videos either to document intermittent hand twitching for diagnosis [42] or to record feelings and thoughts about mental health issues between therapy sessions [63].

Self-management contexts accounted for almost a third of the articles (33/110, 30%), where patients used images in their day-to-day lives to track and control a health concern. These images had a clinical or therapeutic context; however, the studies did not report any image sharing with health care services. The most prevalent concern that people self-manage through photos is dietary intake. Food photos provide rich information to recall details of the foods consumed, with whom they were eaten, and the context [74]. People can use food photos to provide accurate energy intake estimates, which do not differ significantly from the gold standard, doubly labeled water, over short periods (6 days) [81]. However, over longer periods (6 months), adherence to photographic food diaries diminishes [73]. Several studies have explored the feasibility of photos with children [69,86] and adolescents [70,71,80,83-85]. In addition to dietary intake, people also use photos of nature, surroundings, people, objects (including foods), and past events to reflect on their emotional state [93,94] or their current lifestyle and well-being [95-97]. Photos are also used by smokers to capture places, events, and routines associated with smoking or quitting cigarettes [15,98-100].

Social media contexts featured in a third of all articles (36/110, 32.7%), where patients share videos and photos with an audience of peers on the web. The summary in Table 1 shows 2 key differences with social media, compared with other contexts. First, almost half of the studies in this context report on videos generated by patients, where they talk about personal health

experiences on YouTube. A good example is the study by Liu et al [104], which presents insights from 36 video bloggers who share their experiences with chronic conditions that require self-management, such as diabetes and HIV. The findings show that these videos are often used to teach others about self-management or to keep a personal journal to share their physical and emotional updates in their illness journey. Videos (unlike photos or text alone) allow patients to build rapport with their audiences by filming themselves talking, showing emotions, introducing other people, and showing their health care environments and significant events [104]. A second key difference in the social media context is that images are used to present a broad range of health concerns, including cancer experiences [101-106], mental health [93,111-116], and vaccinations [123-127]. This is partly a result of the focus on experience sharing, where people talk about a disease rather than depict a symptom. It also results from social media, allowing patients to find and join web-based communities dedicated to a shared health concern. A poignant example is proanorexia communities on Flickr, Instagram, and YouTube, which use images and videos to promote eating disorders as a desirable lifestyle rather than as a disease [117-120].

Social media contexts overlapped with self-management contexts (8/110, 7.3% articles), where patients used photos predominantly to self-manage a health concern; however, they also shared these photos with peers on the web. Instagram was used to self-monitor diet [88-92] and emotional well-being [93]. Facebook was used to share photos depicting reasons for quitting cigarettes [99]. A bespoke platform (Staccato) was used to capture and share photos of healthy lifestyle choices such as taking steps instead of an escalator [97].

Educational contexts were described in 6.4% (7/110) of the articles, in which the aim was to educate patients about a health concern. In the school context, a face-aging app was used as an educational intervention to promote smoking cessation [129] and sun protection [131,132]. The app allowed students to take a face selfie and to see the potential impact of smoking cigarettes [129] and UV exposure without sunscreen [131,132] on the way that their face will age. A similar educational intervention has been deployed in the context of a physician's waiting room to promote smoking cessation [130]. In a home context, *vagina selfies* were used to let women explore and learn about their own intimate anatomy [135], and patient-recorded toothbrushing videos were used to educate dental residents and to refine their toothbrushing behaviors [134].

Finally, health service improvements were described in 2.7% (3/110) of the articles. In this context, health service providers asked patients and their family members to take photos to better understand their patients' health care experiences with the aim of improving their service delivery. All studies were conducted in pediatric services. Children and parents were invited to take videos or photos to describe their experiences inside the hospital [136,137] and after their transition to their homes [138]. Videos of hospital experiences showed that patients desire better information and ways to share experiences and reflect on feelings [136]. Photos taken at home showed challenges, such as children having to share responsibility for managing medication, and fears and uncertainties, as children adjust to living with a chronic health condition [138].

The Value of Photos and Videos

Patient-generated photos and videos create significant value when used for health and well-being. On the basis of an engagement framework [140], our analysis identified six key values: functional, self-determination, social, emotional, transactional, and efficiency. Table 2 provides a summary of each value and the number of relevant articles.

Table 2. The value of patient-generated photos and videos (N=110 articles).[a]

Value	Description	Articles, n (%)	Photos, n (%)	Videos, n (%)
Functional	Support health outcomes through diagnosis, explanation, treatment, therapy, and health promotion	59 (53.6)	54 (49.1)	7 (6.4)
Self-determination	Empower patient through knowledge, form a personal narrative, and share experiences	39 (35.5)	28 (25.5)	12 (10.9)
Social	Share experience and support with peers, family members, and web-based community members	33 (30)	22 (20)	12 (10.9)
Emotional	Express, understand, and regulate emotions; capture significant moments for therapy	21 (19.1)	18 (16.4)	5 (4.5)
Efficiency	Eliminate unnecessary appointments; replace paper diaries and forms with photographic records	19 (17.3)	19 (17.3)	1 (0.9)
Transactional	Enrich transactions through increased patient engagement and by providing health professionals with a more holistic view of their patients	18 (16.4)	14 (12.7)	6 (5.5)

[a]Several articles reported results on multiple values or on photos and videos.

The most prominent value reported is the functional value (59/110, 53.6%), where photos and videos are used as an aid to support health outcomes through diagnosis, explanation, treatment, therapy, and health promotion. In terms of diagnosis, photos provide important data for health care professionals to diagnose hernias [35], rashes [36], injuries [38], lesions [30,47], and cysts and angioedema [49]. Patients also use photos to self-diagnose skin lesions [32,34] and monitor lesions over time [31]. Photos and videos can provide valuable explanations that lead to new insights for patients about the functioning of their own body [135] and to come to terms with new diagnoses, for example, to cope with cancer [103] and kidney stone disease

[109]. For patients with diabetes, photos and videos provide new knowledge about the impact of lifestyle factors such as diet, alcohol consumption, and exercise on their diabetes management [57,58,60,61,104]. Photos can enhance treatment by showing biopsy sites to decrease wrong-site surgery in dermatology [33], medication monitoring [68], and documentation of the healing of postoperative wounds [41,43,45], ulcers [50], and soft-tissue injuries [44]. The therapeutic value of photos and videos was illustrated in reminiscence therapy in patients with Alzheimer disease, where photos were used to support remembering and reminiscing on personal memories [95], as well as in mental health therapy to reflect on past experiences [113]. Finally, photos and videos support health promotion. This is most common with food photos, which help health professionals and patients create an awareness of patterns of eating, food choices, and portion sizes [55,59,62,73,78,82,84-86]; decide on diet changes to promote healthier food choices [37,51,72,76,87]; and aid in weight loss [52,77].

Functional value often went hand in hand with efficiency value (19/110, 17.3%), where the data provided through photos saved time, money, and effort [140]. Commonly reported with photos of skin conditions, time and money are saved when health professionals assess photos instead of assessing patients in person [41,42,44,45] or when patients can self-diagnose skin lesions and rashes [31,32,34,36,40]. Similarly, patients save time and effort when they are allowed to capture their dietary intake through photos, rather than through pen and paper diaries [56,69,71,76,77,79,80,92].

Several values—self-determination, social, and emotional—come from patients using photos and videos to reflect upon, capture, and share personal health experiences, rather than specific data. Self-determination value (39/110, 35.5% articles) arises when patients "confirm and integrate their beliefs (cognitive, spiritual, or other) into health care services, asserting a degree of control over a health care situation congruent with psychological empowerment" [6]. We identified self-determination value from photos and videos through enhanced knowledge, for example, by examining the personal meanings of smoking and related social influences when quitting smoking [15,98,100]. Health professionals sometimes encourage patients to take on more responsibility by monitoring their condition through photos to shift the power in consultations so that patients become more informed and assertive [33,38,39]. Self-determination also arises when patients use videos to form a personal narrative to make sense of a new diagnosis, such as diabetes [58] or cancer [42,101,102], and what is occurring with their bodies, emotions, and social identity before and after medical interventions. Finally, several studies showed self-determination value from sharing personal health experiences, achievements, resources, and advice with other patients through social media [104,133,136]. This was common for mental health conditions, where negative self-perception is a challenge for many patients. In this context, photos can help empower patients through the expression of emotions and negative self-perception as well as through seeing oneself as part of a (web-based) group with same condition [93,112,113,116,117].

Social value (33/110, 30%) comes from sharing health photos and videos with other patients, family members, and friends. Videos are commonly used to share personal experiences and help with others who manage the same illness, for example, diabetes, HIV, cancer, and multiple sclerosis [58,61,101,103,104,106,107,110,121,136]. Patients report that they were motivated to share personal videos because they could not find the web-based information and guidance they wanted [108] and because they gained additional motivation by being able to help other patients [42]. Photo sharing on social media is also common for general healthy living, for example, to share insights about how to eat healthier meals [88-90,133], stay physically active [97], and give up smoking [98,99]. People with mental health conditions also gain value from posting photos on the web to ask questions, call for help, show empathy, and offer support to others [111-114,118,120]. Many patients reported a sense of community with other social media users who are experiencing similar health challenges [67,104,111,112,115,116,119,133].

Emotional value (21/110, 19.1%) can arise from capturing personal experiences with an illness to better understand and regulate emotions [74,93,94,96]. Emotions reported in the studies include a wide range of emotions: sympathy [97], humor [135], hopefulness [64,101,112], fear [101], hopelessness [112], pain [116], suicidal feelings [114], and ambivalent feelings such as simultaneously feeling joyful and worried [98]. Images also allow patients to express emotions and garner support from family members [138], health care providers [136], and web-based audiences [42,102,108,113,114,116]. Patients report that they feel better when they see other social media users who share similar emotions and that photos are more visually stimulating than written text [88]. Therapies involve patient-generated photos to help clients reflect on coping strategies [64,65] and reminisce about past events and emotions [95].

Finally, photos and videos can enrich transactions between patients and health care professionals (18/110, 16.4%). On the one hand, photos can increase patient engagement. Capturing photos together with personal notes helps patients prepare for consultations and take on a more active role in their interactions with health care professionals, for example, by recalling information about their diet [53], skin lesions [46], and experiences with mental illness [65]. People with aphasia can use photos to support expressive communication with health care professionals [66]. On the other hand, patient-generated photos and videos can empower health care professionals. Reviewing photos during consultation can prompt health care professionals to ask questions about health experiences [42,138], triggers for adverse reactions [53], and adherence to treatment plans [42,138]. Photos used in consultations are not always limited to clinical data, as shown in a study with general practitioners who reported that they also see social images (new babies and holidays) that provide them with insights into the broader lives of their patients that impact their health [38]. Photos and videos help health care professionals gain a more holistic view of their patients and empathize better with their patients, for example, in general practice [38], dementia care

[95], children's hospital [136-138], and cancer prevention and treatment [15,102].

Challenges With Photos and Videos

Overview

The final part of our analysis describes the barriers and challenges faced by patients with health-related photographs and videos. Here, our analysis is structured based on the process of working with photos and videos, starting with challenges that patients face when they take photos, when they share them with peers and health professionals, and when they are examined. These challenges are interrelated, meaning that challenges in taking photos and sharing them, in turn, can also affect examination. Table 3 provides a summary of these challenges.

Table 3. The challenges faced by patients in taking, sharing, and examining images (N=110 articles).

Challenge	Description	Articles, n (%)	Photos, n (%)	Videos, n (%)
Image-taking challenges				
Accessibility	Lack of access to camera phone; poor app usability; difficulty in taking photos of feet or back	16 (14.5)	16 (14.5)	0 (0)
Incomplete image sets	Lapses in food photos over long periods or when people (fail to) reach goal; camera error	23 (20.9)	22 (20)	1 (0.9)
Image quality	Image not in focus or not well lit; image not showing relevant details (body part or food)	16 (14.5)	15 (13.6)	1 (0.9)
Sharing challenges				
Adoption by health professionals	Time and effort required; increased sense of responsibility; limited technical support	4 (3.6)	4 (3.6)	1 (0.9)
Privacy	Potential risk to patients and health care professionals captured; lack of safe image transfer; invisible social media audiences	10 (9.1)	8 (7.3)	3 (2.7)
Misinformation	Inaccurate or misleading social media images (vaccination); unhealthy behaviors (anorexia)	17 (15.5)	7 (6.4)	10 (9.1)
Harmful feedback	Web-based feedback harming people who quit smoking or who share stories of depression	7 (6.4)	4 (3.6)	3 (2.7)
Examination challenges				
Interpretability	Not enough information in images to assess dietary intake or to diagnose skin lesions	10 (9.1)	10 (9.1)	1 (0.9)
Relevancy	Clinicians do not examine images; patients stop when food photos show no new information	6 (5.5)	6 (5.5)	0 (0)
Emotional labor	Anxiety about potential infection or cancer diagnosis; stress from revisiting past struggles with surgery or mental illness	7 (6.4)	6 (5.5)	2 (1.8)

Image-Taking Challenges

Image-taking challenges were largely reported with photos. Challenges with capturing videos rarely surfaced in our review, despite the potentially large burden for video (and audio) capture, storage, and editing.

A major challenge in taking images is accessibility (16/110, 14.5%). For example, patients reported difficulty in accessing body parts such as their feet or their backs with a camera phone [30,31] or felt it inappropriate to access their cameras to capture photos of the groin area [31,43,135] and in social situations (to take food photos in public settings such as a restaurant) [54,55,62,74]. Not all patients have access to a camera phone, or they do not know how to use them, particularly children [69] and older patients [33,66]. Usability issues of bespoke apps also limited the accessibility of photos and videos, especially when instructions for taking photos were unclear [33,70,78,135].

In total, 20.9% (23/110) of the articles reported incomplete sets of images as a challenge. This was particularly a concern when patients took photos of their food over long periods [73,74] or when food photos needed to be taken both before and after having a meal to show what has been consumed [55,56,60,69,71,78,79,81,84,85]. Participants reported that the time, effort, and training required to take good images were causes of incomplete sets of data [57,78,82] or simply that they forgot [54,56,57,60,69,71,74,78,82]. People stop taking images when they reach a health goal or when they fail to do so [48,88] or because of life disruptions such as moving to a new home [61].

Patients had difficulties with taking high-quality images (16/110, 14.5%). Photos were not in focus [30,31,33], or photos were too dark to show the relevant body part [31] or food [55,59,82,87]. Often, images did not present all relevant details. For example, photos did not show all ingredients of a meal [60,78,91,133], and videos lacked details on how to complete preparation for a medical procedure [122]. Poor-quality photos of wounds [41,44] and cancer biopsy sites [33] led authors to conclude that patients require further guidance to take high-quality images.

Sharing Challenges

Several challenges arise when patients share photos and videos with health professionals and peers. A first barrier is the lack of adoption by health professionals (4/110, 3.6%). Attending to photos and videos takes time and effort [82], with health professionals indicating that they need support from medical assistants to review and identify relevant photos [53]. Adoption is also limited by an increased sense of responsibility for health professionals, for example, when patients assume that health professionals are available all the time and that they take responsibility as soon as photos or videos have been shared [42]. The institutional environment also prohibits adoption, for example, when electronic medical records do not support images taken by patients [45].

Sharing health concerns through photos and videos introduces various privacy risks to patients and their carers (10/110, 9.1%). There is a risk that people may gain access to images on the patient's phone, for example, patients may accidentally show health photos when showing other images to family and friends [31]. In the context of health services, privacy is at risk when secure and encrypted options for transferring patient photos are not available or when there is a lack of information on who has access to patient photos stored in electronic medical records [45]. Therefore, some clinicians advise their patients to bring photos on their phones instead of sending them, which leaves patients in control and allows them to retain ownership [38]. The privacy of health professionals is also at risk, for example, when patients take images during consultations [38]. Finally, the context of social media introduces privacy risks because the audience is large and unknown, and information can be taken out of context and misinterpreted. For example, videos describing personal experiences with diabetes [108] or memories for people with Alzheimer disease [95] can be seen by not only strangers but also friends and relatives, which can be painful and make them worry.

Misinformation on social media is a common challenge (17/110, 15.5%). This is the only area where videos are more prevalent than photos (10/110, 9.1%, vs 7/110, 6.4% articles). YouTube videos detailing patient experiences can act as a useful source of health information; however, from a medical perspective, these videos can often be inaccurate. For example, patient videos of bowel preparation for colonoscopy often miss important information, such as types of preparation purgatives, disgust, and embarrassment [122]. Videos reporting on breast reconstruction can provide unrealistic expectations [105]. Some videos present unreliable and potentially misleading information about treatments that have no evidence for being effective, such as home remedies for skin cancer [106] and herbal medicines used to treat kidney stone disease [109]. Patient photos and videos posted on social media commonly present vaccinations in a negative light [126], and they receive a higher number of likes than images with positive views toward vaccination [123,124]. Social media are also used to promote harmful behaviors through images of self-injury [116], suicide [114], and eating disorders [117-120].

Patients often share photos and videos on social media to create social value, but such sharing also carries the risk of receiving harmful feedback (7/110, 6.3%). For example, people who quit smoking can gain valuable social support from Facebook groups, but photos posted by current smokers can be counterproductive to quitting attempts [99]. Similarly, people who shared personal experiences with depression [112], rape [115], and thoughts of suicide [114] on social media reported harmful feedback that blamed the victim or even encouraged suicide.

Examination Challenges

When patients and health professionals examine photos and videos, a first challenge is interpretability (10/110, 9.1%). The risk of misinterpretation is related to food photos, where photos and accompanying self-reports did not provide sufficient information to accurately assess intake, that is, items of a meal, portion size, and nutritional value, often remained unclear [52,78,80,82,91]. Health professionals expressed concerns about potential misdiagnosis when they rely solely on photos or videos from patients [38,42], and patients also recognize that this is a possibility [31]. Potential misdiagnosis was raised, particularly in the context of skin lesions. Overestimating the significance of a particular lesion may lead to anxiety, but, more importantly, underestimating its significance carries the risk of missed melanoma [30].

A second examination challenge lies in the relevancy of photos and videos (6/110, 5.4%). Consultation times are limited, and health providers do not always see patient photos as relevant enough to examine them [53]. Patients stop taking food photos and sharing food photos on the web when they think they provide no new information and become irrelevant [61,88]. A lack of gender and racial diversity can diminish the relevance of photos and videos on social media for a particular person or target group; for example, they may fail to encourage human papillomavirus vaccination among African American individuals when they do not see themselves represented on the web [127]. Time delays between capturing and examining images can also diminish the relevancy of photos for patients, for example, when reflecting on diet or mental health [87,94].

Finally, the papers also highlighted the challenge of emotional labor, where examining photos triggers emotions that patients and caregivers find difficult to manage (7/110, 6.3%). Photos can add stress to patients, particularly when they already feel stressed from having to manage a chronic illness [137]. Patients also report anxiety about possible health issues raised by photos, such as infection [42] or a cancer diagnosis [34]. Emotional labor can also result from photos that bring back stressful memories from the past, such as an unpleasant surgery or struggles with mental illness [31,65,96]. Revisiting photos from the past was a challenge for people with dementia, as photos used for reminiscing triggered positive emotions of happiness as well as negative emotions of sadness and distress [95].

Discussion

Principal Findings

This is the first review to better understand how patient-generated photographs and videos are used across different health and well-being contexts, and what value and challenges they hold for patients and health professionals. In

many ways, photos and videos reflect the characteristics of other PGHD; that is, they capture data related to medical conditions or general wellness, are generated by patients or their caregivers, and are often shared with health care professionals, peers, and other stakeholders [1,26]. However, our results highlight several key messages that show that photos and videos are not merely a subset of PGHD but are a powerful medium to engage patients as active partners in their health care, which generates unique value and challenges.

First, photos and videos not only are used in health care services, in education, and for self-management at home but also play an important role in social media contexts. According to the traditional notion of PGHD [1], images offer valuable health data to aid with health decisions in health care services, self-management, and health education. The most common areas in our review were skin photos that assist with the diagnosis of melanoma, food photos to help assess dietary intake, and information related to mental health for discussion with therapists. We also observed images used in unique and unexpected ways. For example, under education, we found that women were invited to take vagina selfies to explore their own intimate anatomy, which can be awkward but helps break associated taboos [135]. Very few studies reported on videos to aid with health decisions, but videos are needed for decisions relating to body movement, such as diagnosing twitching [42] and assessing toothbrushing skills [134]. By contrast, in social media contexts, videos were more common. Instead of presenting data, patient-generated videos (and, to some extent, photos) were used as a medium to communicate personal health knowledge, experiences, and stories to social media audiences. This has also been characterized as health video blogging [104] or visual narrative [116]. Both concepts describe when patients use images simultaneously for personal purposes, particularly to keep a journal and to reminisce, and for communicative purposes, particularly to document their health journey and teach others. Such experience videos are not limited to health concerns that can be easily captured using a camera. Hence, we found a broad range of health and well-being topics discussed on social media, including cancer [104], eating disorders [118,119], and vaccination [123,124].

Second, photos and videos do not only offer functional value to aid with diagnosis and treatment but also provide value to engage and empower patients. On the one hand, the functional value was the most mentioned (59/110, 53.6% articles), where photos (rather than videos) primarily aid with diagnosis, explanation, or treatment. This result aligns with traditional notions of photos as data that offer insights to health providers and patients to address a health concern [1] or even for health providers to monitor patients remotely [144]. On the other hand, our results highlighted several different types of value—self-determination, social, emotional, and transactional—that directly benefit the patient. These types of value arise from active engagement with photos and videos, both through personal reflection (self-determination and emotional values) and when patients interact with health professionals (transactional value) and peers (social value). In particular, self-determination value can lead to patients feeling a higher degree of control in their health care, congruent with psychological empowerment [145]. Overall, our review shows that the different types of value described in the framework of Burns et al [140] are applicable across different health and well-being domains.

Third, although reviews of PGHD emphasize that patients benefit from technologies that reduce the effort required by automatically collecting data such as physical activity, heart rate, and sleep [7,8], our review highlights the opposite: active engagement to record and interpret images is important to generate self-determination value (39/110, 35.5% articles), where patients gain a sense of control over their disease and feel empowered in their health care. Only very few studies explored wearable cameras that automatically take images [55,59,95], but even these studies emphasized the importance of active engagement in reviewing images with caregivers and health professionals. Here, we see a parallel between the papers in our review and visual research methods used in public health, such as photovoice [146,147] and photo-elicitation interviews [148], which show that the effort of representing one's health through photos pays off because it gives a voice to people that can be empowering [147]. Moreover, visual research methods [146-149] highlight that images can encourage a critical dialogue between different stakeholders to interpret the meaning of an image in a particular social context (eg, a consultation or an online community) and to achieve mutual understanding, which can result in social (33/110, 30%), emotional (21/110, 19.1%), and transactional values (18/110, 16.4%).

Fourth, challenges with photos and videos largely reflect PGHD challenges; however, there are several unique aspects. On the one hand, our results highlight challenges that are reflective of PGHD as discussed in previous work, such as the time and effort required for patients and clinicians [27], incomplete data [27], privacy concerns [7], and limited interpretability and relevancy for clinicians [8,144]. On the other hand, our review highlighted several unique challenges specific to photos and videos. Photos and videos pose unique challenges for data quality, for example, their quality can be diminished by low lighting, lack of focus, and lack of details [30,31,33]. The privacy risks associated with images are potentially higher than those associated with other PGHD because photographic images are more likely to identify the patient than numerical data of physical activity, sleep, and so on. Furthermore, photos and videos are often posted on mainstream social media where privacy is a particular concern, because unlike that in a face-to-face consultation, information on social media is permanent, searchable, copyable, and accessible to invisible audiences [150]. We also found that photos and videos on social media can lead to problematic discourse either through misinformation presented in these images or through harmful feedback from other social media users. These challenges could also affect nonvisual social media data; however, videos allow patients to create a narrative that connects with audiences in ways that are arguably different from numerical or textual health data. For example, video narratives can be persuasive because they personalize information, create dramatic tension, and foster emotional engagement [102], which explains why misinformation was more commonly reported with videos than with photos, despite the smaller number of video articles overall.

Finally, throughout the results, we identified several advantages and disadvantages of photos compared with videos. Photos were more commonly used than videos (90/110, 81.8%, vs 23/110, 20.9% articles) because photos capture the information required to aid decisions in health care service, self-management, and education contexts. Furthermore, photos generally require less effort for capture and examination than videos. However, videos offer a unique advantage through their richness. As explained in media richness theory [151,152], additional details in videos help reduce uncertainty and equivocality for the task at hand. Videos can address uncertainty by providing additional temporal information, which is required to capture and aid with decisions on body movements [42,134] and to provide education on the different steps in a health care procedure [122]. By contrast, equivocality refers to confusion that cannot be clarified by more information but only through a higher quality or richness of information [151,152]. Our review highlighted that such richness in videos was important when patients captured moments of significance, for example, for personal reflection on well-being [96] and for storytelling in therapy sessions [63]. Likewise, such richness was important when patients shared health experiences on social media, which included not only information but also their emotions when dealing with the challenges of cancer [102,104] and mental health disorders [115]. This is not to say that patients cannot use a series of photos and captions to express rich narratives of health experiences on social media, for example, as illustrated by patients using photos to discuss mental health conditions on the web [112]. However, videos provide more opportunities for rich self-expression, for example, through nonverbal cues such as eye contact, facial expressions, and pausing; by involving other actors with their experiences; or by incorporating the physical and temporal contexts of their health experience [104,153].

Limitations

Our review is subject to several limitations. First, our inclusion criteria limited our review results to only English-language articles and published peer-reviewed literature from 2008 to January 2021.

Second, the articles included in this review comprise diverse study designs, target cohorts, and outcomes. A formal assessment of study quality was not undertaken because this was a scoping review [22], in which most published studies have been pilot or feasibility studies. The review did not find any randomized control trials, which is not surprising because photos and videos are often patient driven.

Third, synthesizing outcomes from a large collection of diverse studies across different contexts was challenging. Only a subset of papers reported health outcomes (reported under the functional value). Many papers presented formative research on the feasibility of introducing photos and videos into a particular context or on the experiences and value gained by patients and health care professionals. Hence, instead of outcomes, we framed the *Results* section more broadly around the various contexts, the value generated for patients and health care professionals, and their challenges. The value for patients and health care professionals was analyzed and collated based on an established framework on the value of PGHD [140]. The analysis of contexts and challenges was largely inductive because existing frameworks for PGHD (eg, the studies by West et al [27] and Abdolkhani et al [144]) did not cover the specifics of photos and videos such as challenges with the photo quality or emotional labor. To ensure consistency, the analysis was conducted independently by 3 members of the research team.

Finally, the broad scope of this review and the large number of articles did not allow for a comparison of effects. On the basis of this scoping review, future work is needed that focuses on specific health domains to critically assess and compare patient outcomes.

Practical Implications

This review shows that photos and videos provide a powerful way for patients to be actively engaged in their health care. For patients interested in their health, photos taken on smartphones are an accessible means of documenting, sharing, and reflecting on their health, particularly in areas that are easy to photograph, such as diet, skin, and everyday life experiences related to mental health. Patients can also use their phone to film themselves talking and reflecting upon personal experiences relevant to their health and well-being, which is often used to share knowledge on managing chronic conditions [104] or to reflect on experiences affecting their mental health [63,96,115]. Both photos and videos are powerful because they allow patients to share aspects of their health and lived experience, which they cannot easily describe through words alone [15]. Although photos require effort to take and examine, our review shows that such an effort can generate self-determination values where patients feel empowered [140] and that sharing photos can create emotional and social values.

Health care professionals interested in participatory health care [154] can empower patients by encouraging them to take relevant photos and discuss them during consultations. Photos and videos often document important details that a health care professional may not consider asking about [38,42,138]. We have seen that such dialogue about photos and videos can provide transactional value as well as functional value to better diagnose and treat conditions. On the basis of our review, such engagement can be effective when health care professionals are genuinely interested in the data and experiences of their patients to make shared decisions about treatments [155].

Health care professionals and patients must be aware of ethical challenges and professional standards to maintain privacy, confidentiality, and trust [156]. On the one hand, our review shows that health care professionals can build trust by taking the images provided by their patients seriously [38]. On the other hand, image sharing introduces privacy and confidentiality risks through a lack of secure transfer and storage [45], accidental access to other images on a patient's phone or social media account [31], and potential recordings of the health professional during a consultation [38]. Hence, health care professionals need to be sensitive to and respectful of any patient images to maintain professional relationships and confidentiality [156]. It is recommended that secure platforms be used, for example, by advising patients to bring images on their own phone instead of sharing them via social media [38]. Finally, clear communication is required to inform patients about privacy

protection in place and to establish expectations of how images are used [157].

There are two practical implications for health care services. First, health care provider support is crucial for harnessing the power of health data generated through patient photos and videos. Technology infrastructure, training, and policies are needed to safely transfer, store, access, and integrate patient-generated photos and videos with medical records [11]. In addition, health services need to create an environment where their staff has the time and support needed to review and analyze patient data [27,144]. Second, health care services that engage with patients to share photos and videos can gain crucial insights into the patient experience to help them improve their service delivery [136,137].

Future Research

Scoping reviews are often conducted to determine the value of undertaking a full systematic review [22]. On the basis of the prominent health areas identified in this review, we see value in conducting a narrower review to focus on photos related to skin diseases and to update existing reviews on photos used for dietary assessment [13] and melanoma detection [158]. For dietary assessment, our review identified a large number of recent feasibility studies of food photos with children [69,86] and adolescents [70,71,80,83-85], something that the original systematic review [13] had called for. In addition, our review highlighted that social media play an important role in sharing food photos with peers and gaining social support [88-92]. Similar to an expert review on melanoma detection [158], our results highlighted the importance of patient-generated photos for self-examination and education. In addition, our review also highlighted that patients share melanoma photos and experience videos on social media [101,106]. Finally, our review confirms observations from a professional review on surgical sites [159] that it is feasible for patients to take photos to keep track of wound healing [41,43,45] and that adoption by health professionals remains a challenge due to a lack of time [45].

The breadth of the health areas identified in this review suggests research opportunities to explore patient-generated photos and videos in new health areas. First, in social media contexts, photos and videos are widely used to communicate experiences with diseases that cannot be immediately photographed, such as infectious diseases [123-125,127], Alzheimer disease [95], and myocardial infarction [107]. This breadth suggests that photos and videos can also offer value with other health and well-being contexts that may be invisible to the camera but can be discussed, such as back pain, arthritis, and other musculoskeletal conditions. Second, in the context of education and health promotion, the study of vagina selfies, which allows women to explore their own intimate anatomy [135], as well as the use of videos to share experiences with mammography [121], suggests a broader potential of photos and videos to reflect on women's health, for example, with pregnancy and childbirth, osteoporosis, or breast cancer. Third, in the contexts of health care services, we see potential for using videos more widely to capture body movements for clinical purposes, similar to the presented studies on capturing toothbrushing skills [134] and an intermittently twitching hand [42]. For example, consultations with physiotherapists could benefit from patient-generated videos that capture rehabilitation exercises and activities of daily living at the patient's home.

The challenges faced by patients identified in this review highlight the need for further research on technological design. More work is needed to better understand accessibility needs, particularly when capturing videos in a health context. Collaboration between patients and caregivers is needed to ensure that technologies are usable and accessible. To encourage patients to take images, research into protocols and technology designs that train patients to take high-quality photos as well as provide relevant medical knowledge is needed [57,78,160]. To ensure high-quality images, newer smartphone cameras that offer higher sensitivity in low-light settings need to be harnessed together with research into designing visual aids and voice feedback to guide users in taking photos that capture the required content [28,161]. Finally, to retain engagement, patients benefit from technology designs that assist them in examining their images more effectively. This involves highlighting relevant information in photos, as well as integrating photos with other data that might be scattered across other devices, such as vital signs and lifestyle data from mobile and wearable devices, to explore connections and trends across different data sources [8].

The identified interpretability challenges highlight the need for further research to enhance the relevancy of photos and videos for clinicians. On the one hand, empirical research is needed to better understand the goals and priorities of clinicians [8]. On the other hand, sociotechnical studies are needed to explore how emerging technologies such as machine learning techniques can be harnessed to better manage the large number of photographic images. Our review included only 1 study that examined machine learning techniques to aid in melanoma detection in patient-generated photos [34], whereas in medical imaging, machine learning techniques are already used in clinical practice to aid in the diagnosis and prognosis of various health concerns [162,163]. However, even with sophisticated machine learning algorithms, effective integration into clinical practice remains an open question [164,165].

Finally, more research is needed to investigate privacy and misinformation on social media [166]. The privacy of PGHD is a complex issue across many forms of PGHD [1,7,11], which cannot be addressed simply through a more secure technology infrastructure or privacy policies. Inspired by Palen and Dourish [167], we see privacy as a dynamic practice in which patients negotiate access to personal information according to circumstances. More research is needed to investigate how patients manage their privacy under different circumstances: when they capture photos, manage them on their phones, share them in consultations, or post them on social media. Furthermore, our review identified that patient photos and videos shared on social media provide inaccurate and sometimes misleading information on vaccinations [123,124,126]. In light of current efforts to provide COVID-19 vaccines throughout the world, further research is needed to understand the dangers of misinformation on social media and their impact on public health advice on the COVID-19 pandemic and vaccinations, as

well as research to harness social media to improve the health literacy of patients [168].

Conclusions

This review showed that patient-generated photos and videos are used across a wide range of health care activities. Similar to other forms of PGHD, photos and videos provide critical information to aid in the diagnosis and treatment of various health conditions. However, going beyond textual and numerical PGHD, photos and videos are powerful media that facilitate rich and meaningful interactions, both in person and on social media. They connect fellow patients and facilitate the exchange of social and emotional support. Photos and videos are also powerful media for enriching transactions with health care professionals. Ultimately, they engage patients with their own health and well-being and empower them in their own care.

On the basis of this review, we present agenda for future research. On the one hand, this review highlighted opportunities to expand the use of photos and videos to other health and well-being areas and to better implement them in clinical practice. On the other hand, this review raised the need for more research to address key challenges such as accessibility for patients, relevancy and interpretability for clinicians, and privacy and misinformation on social media, to fully realize the potential of patient-generated photos and videos for health and well-being.

Acknowledgments

This review was supported by the Queensland University of Technology and the Prince Charles Hospital Foundation (grant INN2018-36).

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 , 63 KB - jmir_v24i4e28867_app1.docx]

Multimedia Appendix 2
Search strategy and keywords for all databases.
[DOCX File , 28 KB - jmir_v24i4e28867_app2.docx]

Multimedia Appendix 3
Data extraction spreadsheet.
[XLSX File (Microsoft Excel File), 70 KB - jmir_v24i4e28867_app3.xlsx]

Multimedia Appendix 4
Coding tree, exported from NVivo (version 12; QSR International).
[DOCX File , 31 KB - jmir_v24i4e28867_app4.docx]

References

1. Shapiro M, Johnston D, Wald J, Mon D. Patient-generated health data. Health IT. URL: https://www.healthit.gov/topic/scientific-initiatives/patient-generated-health-data#:~:text=Patient%2Dgenerated%20health%20data%20(PGHD,help%20address%20a%20health%20concern [accessed 2022-03-25]
2. Klasnja P, Pratt W. Healthcare in the pocket: mapping the space of mobile-phone health interventions. J Biomed Inform 2012 Feb;45(1):184-198 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.017] [Medline: 21925288]
3. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
4. 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]
5. Bragazzi NL. From P0 to P6 medicine, a model of highly participatory, narrative, interactive, and "augmented" medicine: some considerations on Salvatore Iaconesi's clinical story. Patient Prefer Adherence 2013;7:353-359 [FREE Full text] [doi: 10.2147/PPA.S38578] [Medline: 23650443]
6. Burns K. Engagement, empowerment and patient generated health data. Queensland University of Technology (QUT). URL: https://eprints.qut.edu.au/114076/ [accessed 2022-03-25]

7. Nittas V, Lun P, Ehrler F, Puhan MA, Mütsch M. Electronic patient-generated health data to facilitate disease prevention and health promotion: scoping review. J Med Internet Res 2019 Oct 14;21(10):e13320 [FREE Full text] [doi: 10.2196/13320] [Medline: 31613225]
8. Choe EK, Lee B, Andersen TO, Wilcox L, Fitzpatrick G. Harnessing the power of patient-generated data. IEEE Pervasive Comput 2018 Apr;17(2):50-56. [doi: 10.1109/mprv.2018.022511243]
9. Lavallee DC, Lee JR, Austin E, Bloch R, Lawrence SO, McCall D, et al. mHealth and patient generated health data: stakeholder perspectives on opportunities and barriers for transforming healthcare. Mhealth 2020;6:8 [FREE Full text] [doi: 10.21037/mhealth.2019.09.17] [Medline: 32190619]
10. Black N. Patient reported outcome measures could help transform healthcare. BMJ 2013 Jan 28;346:f167. [doi: 10.1136/bmj.f167] [Medline: 23358487]
11. Demiris G, Iribarren SJ, Sward K, Lee S, Yang R. Patient generated health data use in clinical practice: a systematic review. Nurs Outlook 2019 Jul;67(4):311-330 [FREE Full text] [doi: 10.1016/j.outlook.2019.04.005] [Medline: 31277895]
12. Burns K. Digital photography and the medical selfie. J Particip Med 2015 Feb 11;7:e3 [FREE Full text]
13. Gemming L, Utter J, Ni Mhurchu C. Image-assisted dietary assessment: a systematic review of the evidence. J Acad Nutr Diet 2015 Jan;115(1):64-77. [doi: 10.1016/j.jand.2014.09.015] [Medline: 25441955]
14. Berger J. Ways of Seeing. Harmondsworth, UK: Penguin; 1972.
15. Haines RJ, Oliffe JL, Bottorff JL, Poland BD. 'The missing picture': tobacco use through the eyes of smokers. Tob Control 2010 Jun 25;19(3):206-212. [doi: 10.1136/tc.2008.027565] [Medline: 20501493]
16. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010 Jun 14;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]
17. Liu L, Gu J, Shao F, Liang X, Yue L, Cheng Q, et al. Application and preliminary outcomes of remote diagnosis and treatment during the COVID-19 outbreak: retrospective cohort study. JMIR Mhealth Uhealth 2020 Jul 03;8(7):e19417 [FREE Full text] [doi: 10.2196/19417] [Medline: 32568722]
18. Boulton NE, Williams J, Jones RS. Could participant-produced photography augment therapeutic interventions for people with intellectual disabilities? A systematic review of the available evidence. J Intellect Disabil 2018 Mar 07;22(1):74-95. [doi: 10.1177/1744629516663027] [Medline: 27503432]
19. Mason RA, Davis HS, Boles MB, Goodwyn F. Efficacy of point-of-view video modeling. Remedial Special Educ 2013 May 16;34(6):333-345. [doi: 10.1177/0741932513486298]
20. 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]
21. 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]
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. Davis K, Drey N, Gould D. What are scoping studies? A review of the nursing literature. Int J Nurs Stud 2009 Oct;46(10):1386-1400. [doi: 10.1016/j.ijnurstu.2009.02.010] [Medline: 19328488]
24. 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]
25. Ray A, Scott A, Nikkhah D, Dheansa B. The medical selfie. BMJ 2015;351:h3145. [doi: 10.1136/bmj.h3145]
26. Figueiredo MC, Chen Y. Patient-generated health data: dimensions, challenges, and open questions. FNT Human Comput Interact 2020;13(3):165-297. [doi: 10.1561/1100000080]
27. West P, Kleek MV, Giordano R, Weal MJ, Shadbolt N. Common barriers to the use of patient-generated data across clinical settings. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18: CHI Conference on Human Factors in Computing Systems; Apr 21 - 26, 2018; Montreal QC Canada. [doi: 10.1145/3173574.3174058]
28. Brown R, Ploderer B, Seng L, Lazzarini P, Van Netten J. MyFootCare: a mobile self-tracking tool to promote self-care amongst people with diabetic foot ulcers. In: Proceedings of the 29th Australian Conference on Computer-Human Interaction. 2017 Presented at: OzCHI '17: 29th Australian Conference on Human-Computer Interaction; Nov 28- Dec 1, 2017; Brisbane Queensland Australia. [doi: 10.1145/3152771.3156158]
29. 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]
30. Boyce Z, Gilmore S, Xu C, Soyer H. The remote assessment of melanocytic skin lesions: a viable alternative to face-to-face consultation. Dermatology 2011;223(3):244-250. [doi: 10.1159/000333363] [Medline: 22095005]
31. Diethei D, Colley A, Kalving M, Salmela T, Häkkilä J, Schöning J. Medical selfies: emotional impacts and practical challenges. In: Proceedings of the 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services. 2020 Presented at: MobileHCI '20: 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services; Oct 5 - 8, 2020; Oldenburg Germany. [doi: 10.1145/3379503.3403555]

32. Kantor J. Skin self-photography for dysplastic nevus monitoring is associated with a decrease in the number of biopsies at follow-up: a retrospective analytical study. J Am Acad Dermatol 2015 Oct;73(4):704-705. [doi: 10.1016/j.jaad.2015.07.005] [Medline: 26369845]
33. Nijhawan R, Lee E, Nehal K. Biopsy site selfies--a quality improvement pilot study to assist with correct surgical site identification. Dermatol Surg 2015 Apr;41(4):499-504. [doi: 10.1097/DSS.0000000000000305] [Medline: 25760559]
34. Zhao J, Cheung N, Sosa R, Koh D. Design self-diagnosis applications for non-patients. In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems. 2015 Presented at: CHI '15: CHI Conference on Human Factors in Computing Systems; Apr 18 - 23, 2015; Seoul Republic of Korea. [doi: 10.1145/2702613.2732865]
35. Saleem M. Digital imaging by parents: an aid to the diagnosis of inguinal hernia in infants and children. Singapore Med J 2008 Feb;49(2):145-146 [FREE Full text] [Medline: 18301843]
36. Hartgers M, Jatoi A. E-mail and photographs: a case report of a patient-initiated diagnostic tool in the era of electronic communication. J Palliat Med 2010 Mar;13(3):335-337 [FREE Full text] [doi: 10.1089/jpm.2009.0215] [Medline: 20178438]
37. Ginting K, Stolfi A, Wright J, Omoloja A. Patient portal, patient-generated images, and medical decision-making in a pediatric ambulatory setting. Appl Clin Inform 2020 Oct 18;11(5):764-768 [FREE Full text] [doi: 10.1055/s-0040-1718754] [Medline: 33207384]
38. Tan L, Hu W, Brooker R. Patient-initiated camera phone images in general practice: a qualitative study of illustrated narratives. Br J Gen Pract 2014 Apr 26;64(622):e290-e294. [doi: 10.3399/bjgp14x679723]
39. Frühauf J, Schwantzer G, Ambros-Rudolph CM, Weger W, Ahlgrimm-Siess V, Salmhofer W, et al. Pilot study using teledermatology to manage high-need patients with psoriasis. Arch Dermatol 2010 Feb 01;146(2):200-201. [doi: 10.1001/archdermatol.2009.375] [Medline: 20157037]
40. Frühauf J, Schwantzer G, Ambros-Rudolph CM, Weger W, Ahlgrimm-Siess V, Salmhofer W, et al. Pilot study on the acceptance of mobile teledermatology for the home monitoring of high-need patients with psoriasis. Australas J Dermatol 2012 Feb;53(1):41-46. [doi: 10.1111/j.1440-0960.2011.00852.x] [Medline: 22309330]
41. Kummerow Broman K, Oyefule OO, Phillips SE, Baucom RB, Holzman MD, Sharp KW, et al. Postoperative care using a secure online patient portal: changing the (inter)face of general surgery. J Am Coll Surg 2015 Dec;221(6):1057-1066 [FREE Full text] [doi: 10.1016/j.jamcollsurg.2015.08.429] [Medline: 26453260]
42. Burns K, McBride CA, Patel B, FitzGerald G, Mathews S, Drennan J. Creating consumer-generated health data: interviews and a pilot trial exploring how and why patients engage. J Med Internet Res 2019 Jun 13;21(6):e12367 [FREE Full text] [doi: 10.2196/12367] [Medline: 31199312]
43. Gunter R, Fernandes-Taylor S, Mahnke A, Awoyinka L, Schroeder C, Wiseman J, et al. Evaluating patient usability of an image-based mobile health platform for postoperative wound monitoring. JMIR Mhealth Uhealth 2016 Sep 28;4(3):e113 [FREE Full text] [doi: 10.2196/mhealth.6023] [Medline: 27683059]
44. Michiels E, Ouellette L, Bush C, VanDePol E, Fleeger T, Jones JS. Camera phones for the follow-up of soft-tissue injuries in adult and pediatric ED patients: a feasibility study. Am J Emerg Med 2017 May;35(5):782-783. [doi: 10.1016/j.ajem.2016.10.033] [Medline: 28188056]
45. Miller M, Ross R, Voight C, Brouwer H, Karavite D, Gerber J, et al. Patient-generated digital images after pediatric ambulatory surgery. Appl Clin Inform 2017 Dec 19;07(03):646-652. [doi: 10.4338/aci-2015-12-cr-0168]
46. Mishra S, Miller A, Haldar S, Khelifi M, Eschler J, Elera R, et al. Supporting collaborative health tracking in the hospital: patients' perspectives. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems; Apr 21 - 26, 2018; Montreal QC Canada. [doi: 10.1145/3173574.3174224]
47. Pourdanesh F, Sayyedi A, Jamilian A, Yaghmaei M. Application of self-recorded photos using mobile phones in maxillofacial surgery. J MTM 2012 Sep 12;1(3):46-49. [doi: 10.7309/jmtm.22]
48. Walker TW, O'Connor N, Byrne S, McCann PJ, Kerin MJ. Electronic follow-up of facial lacerations in the emergency department. J Telemed Telecare 2011 Jan 26;17(3):133-136. [doi: 10.1258/jtt.2010.100307] [Medline: 21270048]
49. Hanu-Cernat LM, Hall E, Barnard NA. The diagnostic value of clinical photographs taken by patients. Br J Oral Maxillofac Surg 2009 Mar;47(2):167. [doi: 10.1016/j.bjoms.2008.07.191] [Medline: 18771825]
50. Dinsdale G, Moore T, Manning J, Murray A, Atkinson R, Ousey K, et al. Tracking digital ulcers in systemic sclerosis: a feasibility study assessing lesion area in patient-recorded smartphone photographs. Ann Rheum Dis 2018 Sep;77(9):1382-1384. [doi: 10.1136/annrheumdis-2017-212829] [Medline: 29431120]
51. Ashman AM, Collins CE, Brown LJ, Rae KM, Rollo ME. A brief tool to assess image-based dietary records and guide nutrition counselling among pregnant women: an evaluation. JMIR Mhealth Uhealth 2016 Nov 04;4(4):e123 [FREE Full text] [doi: 10.2196/mhealth.6469] [Medline: 27815234]
52. Carter MC, Burley VJ, Nykjaer C, Cade JE. 'My Meal Mate' (MMM): validation of the diet measures captured on a smartphone application to facilitate weight loss. Br J Nutr 2013 Feb 14;109(3):539-546. [doi: 10.1017/S0007114512001353] [Medline: 22717334]
53. Chung C, Wang Q, Schroeder J, Cole A, Zia J, Fogarty J, et al. Identifying and planning for individualized change: patient-provider collaboration using lightweight food diaries in healthy eating and irritable bowel syndrome. Proc ACM

Interact Mob Wearable Ubiquitous Technol 2019 Mar 29;3(1):1-27 [FREE Full text] [doi: 10.1145/3314394] [Medline: 31080941]
54. Ehrmann BJ, Anderson RM, Piatt GA, Funnell MM, Rashid H, Shedden K, et al. Digital photography as an educational food logging tool in obese patients with type 2 diabetes: lessons learned from a randomized, crossover pilot trial. Diabetes Educ 2014 Oct 29;40(1):89-99 [FREE Full text] [doi: 10.1177/0145721713508826] [Medline: 24168836]
55. Gemming L, Doherty A, Kelly P, Utter J, Ni Mhurchu C. Feasibility of a SenseCam-assisted 24-h recall to reduce under-reporting of energy intake. Eur J Clin Nutr 2013 Oct 4;67(10):1095-1099. [doi: 10.1038/ejcn.2013.156] [Medline: 24002044]
56. Higgins JA, LaSalle AL, Zhaoxing P, Kasten MY, Bing KN, Ridzon SE, et al. Validation of photographic food records in children: are pictures really worth a thousand words? Eur J Clin Nutr 2009 Aug;63(8):1025-1033. [doi: 10.1038/ejcn.2009.12] [Medline: 19259111]
57. Karway G, Grando MA, Grimm K, Groat D, Cook C, Thompson B. Self-management behaviors of patients with type 1 diabetes: comparing two sources of patient-generated data. Appl Clin Inform 2020 Jan 22;11(1):70-78 [FREE Full text] [doi: 10.1055/s-0039-1701002] [Medline: 31968384]
58. Mamykina L, Miller A, Mynatt E, Greenblatt D. Constructing identities through storytelling in diabetes management. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2010 Presented at: CHI '10: CHI Conference on Human Factors in Computing Systems; Apr 10 - 15, 2010; Atlanta Georgia USA. [doi: 10.1145/1753326.1753507]
59. O'Loughlin G, Cullen SJ, McGoldrick A, O'Connor S, Blain R, O'Malley S, et al. Using a wearable camera to increase the accuracy of dietary analysis. Am J Prev Med 2013 Mar;44(3):297-301. [doi: 10.1016/j.amepre.2012.11.007] [Medline: 23415128]
60. Rollo ME, Ash S, Lyons-Wall P, Russell A. Trial of a mobile phone method for recording dietary intake in adults with type 2 diabetes: evaluation and implications for future applications. J Telemed Telecare 2011 Aug 15;17(6):318-323. [doi: 10.1258/jtt.2011.100906] [Medline: 21844173]
61. Sun S, Belkin NJ. Managing personal information over the long-term, or not? Experiences by type 1 diabetes patients. Proc Assoc Info Sci Tech 2016 Dec 27;53(1):1-10. [doi: 10.1002/pra2.2016.14505301070]
62. Ptomey LT, Willis EA, Goetz JR, Lee J, Sullivan DK, Donnelly JE. Digital photography improves estimates of dietary intake in adolescents with intellectual and developmental disabilities. Disabil Health J 2015 Jan;8(1):146-150. [doi: 10.1016/j.dhjo.2014.08.011] [Medline: 25281035]
63. Matthews M, Doherty G. My mobile story: therapeutic storytelling for children. In: Proceedings of the CHI '11 Extended Abstracts on Human Factors in Computing Systems. 2011 Presented at: CHI '11: CHI Conference on Human Factors in Computing Systems; May 7 - 12, 2011; Vancouver BC Canada. [doi: 10.1145/1979742.1979860]
64. Quaglietti S. Creating a hope narrative for veterans in recovery using photography and written expression. J Creat Ment Health 2020 Jun 23;16(3):308-321. [doi: 10.1080/15401383.2020.1783415]
65. Sitvast JE, Abma TA, Widdershoven GA. Facades of suffering: clients' photo stories about mental illness. Arch Psychiatr Nurs 2010 Oct;24(5):349-361. [doi: 10.1016/j.apnu.2010.02.004] [Medline: 20851326]
66. Ulmer E, Hux K, Brown J, Nelms T, Reeder C. Using self-captured photographs to support the expressive communication of people with aphasia. Aphasiology 2016 Dec 30;31(10):1183-1204. [doi: 10.1080/02687038.2016.1274872]
67. Waycott J, Vetere F, Pedell S, Kulik L, Ozanne E, Gruner A, et al. Older adults as digital content producers. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: CHI '13: CHI Conference on Human Factors in Computing Systems; Apr 27 - May 2, 2013; Paris France. [doi: 10.1145/2470654.2470662]
68. Galloway G, Coyle J, Guillén JE, Flower K, Mendelson J. A simple, novel method for assessing medication adherence: capsule photographs taken with cellular telephones. J Addict Med 2011 Sep;5(3):170-174. [doi: 10.1097/ADM.0b013e3181fcb5fd] [Medline: 21844832]
69. Aflague TF, Boushey CJ, Guerrero RT, Ahmad Z, Kerr DA, Delp EJ. Feasibility and use of the mobile food record for capturing eating occasions among children ages 3-10 years in Guam. Nutrients 2015 Jun 02;7(6):4403-4415 [FREE Full text] [doi: 10.3390/nu7064403] [Medline: 26043037]
70. Boushey CJ, Harray AJ, Kerr DA, Schap TE, Paterson S, Aflague T, et al. How willing are adolescents to record their dietary intake? The mobile food record. JMIR Mhealth Uhealth 2015 May 29;3(2):e47 [FREE Full text] [doi: 10.2196/mhealth.4087] [Medline: 26024996]
71. Casperson SL, Sieling J, Moon J, Johnson L, Roemmich JN, Whigham L. A mobile phone food record app to digitally capture dietary intake for adolescents in a free-living environment: usability study. JMIR Mhealth Uhealth 2015 Mar 13;3(1):e30 [FREE Full text] [doi: 10.2196/mhealth.3324] [Medline: 25775506]
72. Coary S, Poor M. How consumer-generated images shape important consumption outcomes in the food domain. J Consum Market 2016 Jan 11;33(1):1-8. [doi: 10.1108/JCM-02-2015-1337]
73. Comulada WS, Swendeman D, Koussa MK, Mindry D, Medich M, Estrin D, et al. Adherence to self-monitoring healthy lifestyle behaviours through mobile phone-based ecological momentary assessments and photographic food records over 6 months in mostly ethnic minority mothers. Public Health Nutr 2018 Mar;21(4):679-688 [FREE Full text] [doi: 10.1017/S1368980017003044] [Medline: 29199630]

74. Cordeiro F, Bales E, Cherry E, Fogarty J. Rethinking the mobile food journal: exploring opportunities for lightweight photo-based capture. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015 Presented at: CHI '15: CHI Conference on Human Factors in Computing Systems; Apr 18 - 23, 2015; Seoul Republic of Korea. [doi: 10.1145/2702123.2702154]
75. Daugherty BL, Schap TE, Ettienne-Gittens R, Zhu FM, Bosch M, Delp EJ, et al. Novel technologies for assessing dietary intake: evaluating the usability of a mobile telephone food record among adults and adolescents. J Med Internet Res 2012 Apr 13;14(2):e58 [FREE Full text] [doi: 10.2196/jmir.1967] [Medline: 22504018]
76. Doumit R, Long J, Kazandjian C, Gharibeh N, Karam L, Song H, et al. Effects of recording food intake using cell phone camera pictures on energy intake and food choice. Worldviews Evid Based Nurs 2016 Jul;13(3):216-223. [doi: 10.1111/wvn.12123] [Medline: 26646555]
77. Dunn CG, Turner-McGrievy GM, Wilcox S, Hutto B. Dietary self-monitoring through calorie tracking but not through a digital photography app is associated with significant weight loss: the 2smart pilot study-a 6-month randomized trial. J Acad Nutr Diet 2019 Sep;119(9):1525-1532 [FREE Full text] [doi: 10.1016/j.jand.2019.03.013] [Medline: 31155474]
78. Fowler LA, Yingling LR, Brooks AT, Wallen GR, Peters-Lawrence M, McClurkin M, et al. Digital food records in community-based interventions: mixed-methods pilot study. JMIR Mhealth Uhealth 2018 Jul 17;6(7):e160 [FREE Full text] [doi: 10.2196/mhealth.9729] [Medline: 30021705]
79. Hongu N, Pope BT, Bilgiç P, Orr BJ, Suzuki A, Kim AS, et al. Usability of a smartphone food picture app for assisting 24-hour dietary recall: a pilot study. Nutr Res Pract 2015 Apr;9(2):207-212 [FREE Full text] [doi: 10.4162/nrp.2015.9.2.207] [Medline: 25861429]
80. Lee CD, Chae J, Schap TE, Kerr DA, Delp EJ, Ebert DS, et al. Comparison of known food weights with image-based portion-size automated estimation and adolescents' self-reported portion size. J Diabetes Sci Technol 2012 Mar 01;6(2):428-434 [FREE Full text] [doi: 10.1177/193229681200600231] [Medline: 22538157]
81. Martin CK, Correa JB, Han H, Allen HR, Rood JC, Champagne CM, et al. Validity of the Remote Food Photography Method (RFPM) for estimating energy and nutrient intake in near real-time. Obesity (Silver Spring) 2012 Apr;20(4):891-899 [FREE Full text] [doi: 10.1038/oby.2011.344] [Medline: 22134199]
82. Martin CK, Han H, Coulon SM, Allen HR, Champagne CM, Anton SD. A novel method to remotely measure food intake of free-living individuals in real time: the remote food photography method. Br J Nutr 2009 Feb;101(3):446-456 [FREE Full text] [doi: 10.1017/S0007114508027438] [Medline: 18616837]
83. Ptomey LT, Willis EA, Honas JJ, Mayo MS, Washburn RA, Herrmann SD, et al. Validity of energy intake estimated by digital photography plus recall in overweight and obese young adults. J Acad Nutr Diet 2015 Sep;115(9):1392-1399 [FREE Full text] [doi: 10.1016/j.jand.2015.05.006] [Medline: 26122282]
84. Segovia-Siapco G, Sabaté J. Using personal mobile phones to assess dietary intake in free-living adolescents: comparison of face-to-face versus telephone training. JMIR Mhealth Uhealth 2016 Jul 29;4(3):e91 [FREE Full text] [doi: 10.2196/mhealth.5418] [Medline: 27473291]
85. Six BL, Schap TE, Zhu FM, Mariappan A, Bosch M, Delp EJ, et al. Evidence-based development of a mobile telephone food record. J Am Diet Assoc 2010 Jan;110(1):74-79 [FREE Full text] [doi: 10.1016/j.jada.2009.10.010] [Medline: 20102830]
86. Small L, Sidora-Arcoleo K, Vaughan L, Creed-Capsel J, Chung K, Stevens C. Validity and reliability of photographic diet diaries for assessing dietary intake among young children. ICAN Infant Child Adolescent Nutrition 2009 Feb 01;1(1):27-36. [doi: 10.1177/1941406408330360]
87. Zepeda L, Deal D. Think before you eat: photographic food diaries as intervention tools to change dietary decision making and attitudes. Int J Consum Stud 2008;32(6):692-698. [doi: 10.1111/j.1470-6431.2008.00725.x]
88. Chung C, Agapie E, Schroeder J, Mishra S, Fogarty J, Munson S. When personal tracking becomes social: examining the use of Instagram for healthy eating. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI '17: CHI Conference on Human Factors in Computing Systems; May 6 - 11, 2017; Denver Colorado USA. [doi: 10.1145/3025453.3025747]
89. Epstein D, Cordeiro F, Fogarty J, Hsieh G, Munson S. Crumbs: lightweight daily food challenges to promote engagement and mindfulness. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 Presented at: CHI'16: CHI Conference on Human Factors in Computing Systems; May 7 - 12, 2016; San Jose California USA. [doi: 10.1145/2858036.2858044]
90. Helander E, Kaipainen K, Korhonen I, Wansink B. Factors related to sustained use of a free mobile app for dietary self-monitoring with photography and peer feedback: retrospective cohort study. J Med Internet Res 2014 Apr 15;16(4):e109 [FREE Full text] [doi: 10.2196/jmir.3084] [Medline: 24735567]
91. Linehan C, Doughty M, Lawson S, Kirman B, Olivier P, Moynihan P. Tagliatelle: social tagging to encourage healthier eating. In: Proceedings of the CHI '10 Extended Abstracts on Human Factors in Computing Systems. 2010 Presented at: CHI '10: CHI Conference on Human Factors in Computing Systems; Apr 10 - 15, 2010; Atlanta Georgia USA. [doi: 10.1145/1753846.1753980]
92. Turner-McGrievy GM, Helander EE, Kaipainen K, Perez-Macias JM, Korhonen I. The use of crowdsourcing for dietary self-monitoring: crowdsourced ratings of food pictures are comparable to ratings by trained observers. J Am Med Inform Assoc 2015 Apr;22(e1):e112-e119. [doi: 10.1136/amiajnl-2014-002636] [Medline: 25092793]

93. Nave C, Romão T, Correia N. Self-tracking emotional states through social media mobile photography. In: Proceedings of the 32nd International BCS Human Computer Interaction Conference (HCI). 2018 Presented at: 32nd International BCS Human Computer Interaction Conference (HCI); Jul 4 - 6, 2018; Belfast United Kingdom. [doi: 10.14236/ewic/hci2018.146]
94. Tsujita H, Rekimoto J. HappinessCounter: smile-encouraging appliance to increase positive mood. In: Proceedings of the CHI '11 Extended Abstracts on Human Factors in Computing Systems. 2011 Presented at: CHI '11: CHI Conference on Human Factors in Computing Systems; May 7 - 12, 2011; Vancouver BC Canada. [doi: 10.1145/1979742.1979608]
95. Crete-Nishihata M, Baecker R, Massimi M, Ptak D, Campigotto R, Kaufman L, et al. Reconstructing the past: personal memory technologies are not just personal and not just for memory. Human Comput Interact 2012;27(1-2):92-123. [doi: 10.1080/07370024.2012.656062]
96. Isaacs E, Konrad A, Walendowski A, Lennig T, Hollis V, Whittaker S. Echoes from the past: how technology mediated reflection improves well-being. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: CHI '13: CHI Conference on Human Factors in Computing Systems; Apr 27- May 2, 2013; Paris France. [doi: 10.1145/2470654.2466137]
97. Adams P, Baumer E, Gay G. Staccato social support in mobile health applications. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: CHI '14: CHI Conference on Human Factors in Computing Systems; Apr 26- May 1, 2014; Toronto Ontario Canada. [doi: 10.1145/2556288.2557297]
98. Haines-Saah RJ, Oliffe JL, White CF, Bottorff JL. "It is just not part of the culture here": young adults' photo-narratives about smoking, quitting, and healthy lifestyles in Vancouver, Canada. Health Place 2013 Jul;22:19-28 [FREE Full text] [doi: 10.1016/j.healthplace.2013.02.004] [Medline: 23542328]
99. Haines-Saah RJ, Kelly MT, Oliffe JL, Bottorff JL. Picture Me Smokefree: a qualitative study using social media and digital photography to engage young adults in tobacco reduction and cessation. J Med Internet Res 2015 Jan 26;17(1):e27 [FREE Full text] [doi: 10.2196/jmir.4061] [Medline: 25624064]
100. Oliffe J, Bottorff J, Kelly M, Halpin M. Analyzing participant produced photographs from an ethnographic study of fatherhood and smoking. Res Nurs Health 2008 Oct;31(5):529-539. [doi: 10.1002/nur.20269] [Medline: 18228606]
101. Cho H, Silver N, Na K, Adams D, Luong KT, Song C. Visual cancer communication on social media: an examination of content and effects of #Melanomasucks. J Med Internet Res 2018 Sep 05;20(9):e10501 [FREE Full text] [doi: 10.2196/10501] [Medline: 30185403]
102. Chou WS, Hunt Y, Folkers A, Augustson E. Cancer survivorship in the age of YouTube and social media: a narrative analysis. J Med Internet Res 2011 Jan 17;13(1):e7 [FREE Full text] [doi: 10.2196/jmir.1569] [Medline: 21247864]
103. Clerici C, Veneroni L, Bisogno G, Trapuzzano A, Ferrari A. Videos on rhabdomyosarcoma on YouTube: an example of the availability of information on pediatric tumors on the web. J Pediatr Hematol Oncol 2012 Nov;34(8):e329-e331. [doi: 10.1097/MPH.0b013e31825886f8] [Medline: 22858570]
104. Liu L, Huh J, Neogi T, Inkpen K, Pratt W. Health vlogger-viewer interaction in chronic illness management. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: CHI '13: CHI Conference on Human Factors in Computing Systems; Apr 27 - May 2, 2013; Paris France. [doi: 10.1145/2470654.2470663]
105. Tan ML, Kok K, Ganesh V, Thomas SS. Patient information on breast reconstruction in the era of the world wide web. A snapshot analysis of information available on youtube.com. Breast 2014 Feb;23(1):33-37. [doi: 10.1016/j.breast.2013.10.003] [Medline: 24215982]
106. Basch CH, Basch CE, Hillyer GC, Reeves R. YouTube videos related to skin cancer: a missed opportunity for cancer prevention and control. JMIR Cancer 2015 Mar 02;1(1):e1 [FREE Full text] [doi: 10.2196/cancer.4204] [Medline: 28410167]
107. Pant S, Deshmukh A, Murugiah K, Kumar G, Sachdeva R, Mehta JL. Assessing the credibility of the "YouTube approach" to health information on acute myocardial infarction. Clin Cardiol 2012 May;35(5):281-285 [FREE Full text] [doi: 10.1002/clc.21981] [Medline: 22487995]
108. Gómez-Zúñiga B, Fernandez-Luque L, Pousada M, Hernández-Encuentra E, Armayones M. ePatients on YouTube: analysis of four experiences from the patients' perspective. Med 2 0 2012;1(1):e1 [FREE Full text] [doi: 10.2196/med2.2039] [Medline: 25075229]
109. Sood A, Sarangi S, Pandey A, Murugiah K. YouTube as a source of information on kidney stone disease. Urology 2011 Mar;77(3):558-562. [doi: 10.1016/j.urology.2010.07.536] [Medline: 21131027]
110. Fernandez-Luque L, Elahi N, Grajales F. An analysis of personal medical information disclosed in YouTube videos created by patients with multiple sclerosis. In: Studies in Health Technology and Informatics. Amsterdam, Netherlands: IOS Press; 2009.
111. Andalibi N, Ozturk P, Forte A. Depression-related imagery on instagram. In: Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing. 2015 Presented at: CSCW '15: Computer Supported Cooperative Work and Social Computing; Mar 14 - 18, 2015; Vancouver BC Canada. [doi: 10.1145/2685553.2699014]
112. Andalibi N, Ozturk P, Forte A. Sensitive self-disclosures, responses, and social support on Instagram: the case of #Depression. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. 2017 Presented at: CSCW '17: Computer Supported Cooperative Work and Social Computing; Feb 25 - Mar 1, 2017; Portland Oregon USA. [doi: 10.1145/2998181.2998243]

113. Manikonda L, De Choudhury M. Modeling and understanding visual attributes of mental health disclosures in social media. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI '17: CHI Conference on Human Factors in Computing Systems; May 6 - 11, 2017; Denver Colorado USA. [doi: 10.1145/3025453.3025932]
114. Brown RC, Bendig E, Fischer T, Goldwich AD, Baumeister H, Plener PL. Can acute suicidality be predicted by Instagram data? Results from qualitative and quantitative language analyses. PLoS One 2019 Sep 10;14(9):e0220623 [FREE Full text] [doi: 10.1371/journal.pone.0220623] [Medline: 31504042]
115. Harrington C. Neo-liberal subjectivity, self-branding and 'my rape story' YouTube videos. Crit Sociol 2018 May 23;45(7-8):1181-1194. [doi: 10.1177/0896920518778107]
116. Seko Y. Picturesque wounds: a multimodal analysis of self-injury photographs on flickr. Forum Qual Soc Res 2013;14(2). [doi: 10.17169/fqs-14.2.1935]
117. Chancellor S, Lin Z, Goodman E, Zerwas S, De Choudhury M. Quantifying and predicting mental illness severity in online pro-eating disorder communities. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. 2016 Presented at: CSCW '16: Computer Supported Cooperative Work and Social Computing; Feb 27- Mar 2, 2016; San Francisco California USA. [doi: 10.1145/2818048.2819973]
118. Oksanen A, Garcia D, Sirola A, Näsi M, Kaakinen M, Keipi T, et al. Pro-anorexia and anti-pro-anorexia videos on YouTube: sentiment analysis of user responses. J Med Internet Res 2015 Nov 12;17(11):e256 [FREE Full text] [doi: 10.2196/jmir.5007] [Medline: 26563678]
119. 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]
120. Yom-Tov E, Fernandez-Luque L, Weber I, Crain SP. Pro-anorexia and pro-recovery photo sharing: a tale of two warring tribes. J Med Internet Res 2012 Nov 07;14(6):e151 [FREE Full text] [doi: 10.2196/jmir.2239] [Medline: 23134671]
121. Basch CH, Hillyer GC, MacDonald ZL, Reeves R, Basch CE. Characteristics of YouTube videos related to mammography. J Cancer Educ 2015 Dec;30(4):699-703. [doi: 10.1007/s13187-014-0769-9] [Medline: 25502853]
122. Basch CH, Hillyer GC, Reeves R, Basch CE. Analysis of YouTube videos related to bowel preparation for colonoscopy. World J Gastrointest Endosc 2014 Sep 16;6(9):432-435 [FREE Full text] [doi: 10.4253/wjge.v6.i9.432] [Medline: 25228945]
123. Ache KA, Wallace LS. Human papillomavirus vaccination coverage on YouTube. Am J Prev Med 2008 Oct;35(4):389-392. [doi: 10.1016/j.amepre.2008.06.029] [Medline: 18675530]
124. Briones R, Nan X, Madden K, Waks L. When vaccines go viral: an analysis of HPV vaccine coverage on YouTube. Health Commun 2012;27(5):478-485. [doi: 10.1080/10410236.2011.610258] [Medline: 22029723]
125. Chen T, Dredze M. Vaccine images on Twitter: analysis of what images are shared. J Med Internet Res 2018 Apr 03;20(4):e130 [FREE Full text] [doi: 10.2196/jmir.8221] [Medline: 29615386]
126. Guidry JP, Carlyle K, Messner M, Jin Y. On pins and needles: how vaccines are portrayed on Pinterest. Vaccine 2015 Sep 22;33(39):5051-5056. [doi: 10.1016/j.vaccine.2015.08.064] [Medline: 26319742]
127. Lama Y, Chen T, Dredze M, Jamison A, Quinn SC, Broniatowski DA. Discordance between human papillomavirus Twitter images and disparities in human papillomavirus risk and disease in the United States: mixed-methods analysis. J Med Internet Res 2018 Sep 14;20(9):e10244 [FREE Full text] [doi: 10.2196/10244] [Medline: 30217792]
128. Choe E, Lee N, Lee B, Pratt W, Kientz J. Understanding quantified-selfers' practices in collecting and exploring personal data. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: CHI '14: CHI Conference on Human Factors in Computing Systems; Apr 26- May 1, 2014; Toronto Ontario Canada. [doi: 10.1145/2556288.2557372]
129. Brinker T, Alfitian J, Seeger W, Groneberg D, von Kalle C, Enk A, et al. A face-aging smoking prevention/cessation intervention for nursery school students in Germany: an appearance-focused interventional study. Int J Environ Res Public Health 2018 Aug 04;15(8):1656 [FREE Full text] [doi: 10.3390/ijerph15081656] [Medline: 30081549]
130. Brinker TJ, Brieske CM, Esser S, Klode J, Mons U, Batra A, et al. A face-aging app for smoking cessation in a waiting room setting: pilot study in an HIV outpatient clinic. J Med Internet Res 2018 Aug 15;20(8):e10976 [FREE Full text] [doi: 10.2196/10976] [Medline: 30111525]
131. Brinker TJ, Brieske CM, Schaefer CM, Buslaff F, Gatzka M, Petri MP, et al. Photoaging mobile apps in school-based melanoma prevention: pilot study. J Med Internet Res 2017 Sep 08;19(9):e319 [FREE Full text] [doi: 10.2196/jmir.8661] [Medline: 28887295]
132. Brinker TJ, Heckl M, Gatzka M, Heppt MV, Resende Rodrigues H, Schneider S, et al. A skin cancer prevention facial-aging mobile app for secondary schools in Brazil: appearance-focused interventional study. JMIR Mhealth Uhealth 2018 Mar 09;6(3):e60 [FREE Full text] [doi: 10.2196/mhealth.9794] [Medline: 29523502]
133. Parker A, Kantroo V, Lee H, Osornio M, Sharma M, Grinter R. Health promotion as activism: building community capacity to effect social change. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2012 Presented at: CHI '12: CHI Conference on Human Factors in Computing Systems; May 5 - 10, 2012; Austin Texas USA. [doi: 10.1145/2207676.2207692]

134. Madan Kumar P, Mohandoss A, Walls T, Rooban T, Vernon L. Using smartphone video "selfies" to monitor change in toothbrushing behavior after a brief intervention: a pilot study. Indian J Dent Res 2016;27(3):268-277 [FREE Full text] [doi: 10.4103/0970-9290.186241] [Medline: 27411655]
135. Almeida T, Comber R, Wood G, Saraf D, Balaam M. On looking at the vagina through labella. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 Presented at: CHI'16: CHI Conference on Human Factors in Computing Systems; May 7 - 12, 2016; San Jose California USA. [doi: 10.1145/2858036.2858119]
136. Karisalmi N, Kaipio J, Lahdenne P. Improving patient experience in a children's hospital: new digital services for children and their families. In: Building Continents of Knowledge in Oceans of Data: The Future of Co-Created eHealth. Amsterdam, Netherlands: IOS Press; 2018.
137. Karisalmi N, Nieminen M. Selfies of sickness: the use of video diaries with chronically ill children. In: Building Capacity for Health Informatics in the Future. Amsterdam, Netherlands: IOS Press; 2017.
138. Pollack AH, Snyder J. Reflecting on patient-generated photographs of the pediatric renal transplant experience. Pediatr Transplant 2021 May 28;25(3):e13896. [doi: 10.1111/petr.13896] [Medline: 33111458]
139. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
140. Burns K, Tuzovic S. Chapter 22: Conceptualizing health consumer engagement: an extended framework of resource integration, co-creation and engagement. In: Handbook of Research on Customer Engagement. Cheltenham: Edward Elgar Publishing; 2019.
141. World Health Organization. International Statistical Classification of Diseases and Related Health Problems. 10th revision. Geneva, Switzerland: World Health Organization; 2010.
142. Porter M, Lee T. The strategy that will fix health care. Harvard Business Review. URL: https://hbr.org/2013/10/the-strategy-that-will-fix-health-care [accessed 2022-03-25]
143. Teisberg E, Wallace S, O'Hara S. Defining and implementing value-based health care. Acad Med 2020;95(5):682-685. [doi: 10.1097/acm.0000000000003122]
144. Abdolkhani R, Gray K, Borda A, DeSouza R. Patient-generated health data management and quality challenges in remote patient monitoring. JAMIA Open 2019 Dec;2(4):471-478 [FREE Full text] [doi: 10.1093/jamiaopen/ooz036] [Medline: 32025644]
145. Zimmerman M. Empowerment theory. In: Handbook of Community Psychology. Boston, MA: Springer; 2000.
146. Catalani C, Minkler M. Photovoice: a review of the literature in health and public health. Health Educ Behav 2010 Jun 01;37(3):424-451. [doi: 10.1177/1090198109342084] [Medline: 19797541]
147. Wang C, Burris MA. Photovoice: concept, methodology, and use for participatory needs assessment. Health Educ Behav 1997 Jun;24(3):369-387. [doi: 10.1177/109019819702400309] [Medline: 9158980]
148. Harper D. Talking about pictures: a case for photo elicitation. Visual Stud 2002 Jan;17(1):13-26. [doi: 10.1080/14725860220137345]
149. Schwartz D. Visual ethnography: using photography in qualitative research. Qual Sociol 1989;12(2):119-154. [doi: 10.1007/bf00988995]
150. boyd D. Why youth (heart) social network sites: the role of networked publics in teenage social life. In: MacArthur Foundation Series on Digital Learning – Youth, Identity, and Digital Media Volume. Cambridge, MA: MIT Press; 2007.
151. Daft RL, Lengel RH. Organizational information requirements, media richness and structural design. Manag Sci 1986 May;32(5):554-571. [doi: 10.1287/mnsc.32.5.554]
152. Ishii K, Lyons M, Carr S. Revisiting media richness theory for today and future. Human Behav Emerg Technol 2019 Apr 26;1(2):124-131 [FREE Full text] [doi: 10.1002/hbe2.138]
153. Zhang Y, Parker AG. Eat4Thought: a design of food journaling. In: Proceedings of the Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20: CHI Conference on Human Factors in Computing Systems; Apr 25 - 30, 2020; Honolulu HI USA. [doi: 10.1145/3334480.3383044]
154. Coughlin S, Roberts D, O'Neill K, Brooks P. Looking to tomorrow's healthcare today: a participatory health perspective. Intern Med J 2018 Jan 03;48(1):92-96. [doi: 10.1111/imj.13661] [Medline: 29314515]
155. Chakrabarti S. What's in a name? Compliance, adherence and concordance in chronic psychiatric disorders. World J Psychiatry 2014 Jun 22;4(2):30-36 [FREE Full text] [doi: 10.5498/wjp.v4.i2.30] [Medline: 25019054]
156. Denecke K, Bamidis P, Bond C, Gabarron E, Househ M, Lau AY, et al. Ethical issues of social media usage in healthcare. Yearb Med Inform 2015 Aug 13;10(1):137-147 [FREE Full text] [doi: 10.15265/IY-2015-001] [Medline: 26293861]
157. Chretien KC, Kind T. Social media and clinical care. Circulation 2013 Apr 02;127(13):1413-1421. [doi: 10.1161/circulationaha.112.128017]
158. Yagerman S, Marghoob A. Melanoma patient self-detection: a review of efficacy of the skin self-examination and patient-directed educational efforts. Expert Rev Anticancer Ther 2013 Dec 10;13(12):1423-1431. [doi: 10.1586/14737140.2013.856272] [Medline: 24236821]
159. Lober WB, Evans HL. Patient-generated health data in surgical site infection: changing clinical workflow and care delivery. Surg Infect (Larchmt) 2019 Oct 01;20(7):571-576. [doi: 10.1089/sur.2019.195] [Medline: 31397635]

160. Bowen AC, Burns K, Tong SY, Andrews RM, Liddle R, O'Meara IM, et al. Standardising and assessing digital images for use in clinical trials: a practical, reproducible method that blinds the assessor to treatment allocation. PLoS One 2014 Nov 6;9(11):e110395 [FREE Full text] [doi: 10.1371/journal.pone.0110395] [Medline: 25375169]
161. Yap MH, Chatwin KE, Ng C, Abbott CA, Bowling FL, Rajbhandari S, et al. A new mobile application for standardizing diabetic foot images. J Diabetes Sci Technol 2018 Jan;12(1):169-173 [FREE Full text] [doi: 10.1177/1932296817713761] [Medline: 28637356]
162. Shen D, Wu G, Suk H. Deep learning in medical image analysis. Annu Rev Biomed Eng 2017 Jun 21;19:221-248 [FREE Full text] [doi: 10.1146/annurev-bioeng-071516-044442] [Medline: 28301734]
163. Lundervold A, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. Z Med Phys 2019 May;29(2):102-127 [FREE Full text] [doi: 10.1016/j.zemedi.2018.11.002] [Medline: 30553609]
164. Yang Q, Steinfeld A, Zimmerman J. Unremarkable AI: fitting intelligent decision support into critical, clinical decision-making processes. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4 - 9, 2019; Glasgow Scotland Uk. [doi: 10.1145/3290605.3300468]
165. Liu X, Faes L, Kale A, Wagner S, Fu D, 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]
166. Dyer KA. Ethical challenges of medicine and health on the internet: a review. J Med Internet Res 2001 Jun 28;3(2):E23 [FREE Full text] [doi: 10.2196/jmir.3.2.e23] [Medline: 11720965]
167. Palen L, Dourish P. Unpacking "privacy" for a networked world. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2003 Presented at: CHI03: Human Factors in Computing Systems; Apr 5 - 10, 2003; Ft. Lauderdale Florida USA. [doi: 10.1145/642611.642635]
168. Health literacy-2021. Australian Medical Association. 2021. URL: https://ama.com.au/articles/health-literacy-2021 [accessed 2022-03-26]

Abbreviations

PGHD: patient-generated health data
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROM: patient-reported outcome measure

Edited by A Mavragani; submitted 17.03.21; peer-reviewed by Y Luo, J Chishtie, A Mahnke, N Slabyj; comments to author 27.05.21; revised version received 15.10.21; accepted 28.02.22; published 12.04.22.

Please cite as:
Ploderer B, Rezaei Aghdam A, Burns K
Patient-Generated Health Photos and Videos Across Health and Well-being Contexts: Scoping Review
J Med Internet Res 2022;24(4):e28867
URL: https://www.jmir.org/2022/4/e28867
doi:10.2196/28867
PMID:35412458

Review

The Willingness to Pay for Telemedicine Among Patients With Chronic Diseases: Systematic Review

Valerie Chua[1]; Jin Hean Koh[1]; Choon Huat Gerald Koh[1], PhD; Shilpa Tyagi[1], PhD

Office of Healthcare Transformation, Ministry of Health, Singapore, Singapore

Corresponding Author:
Valerie Chua
Office of Healthcare Transformation
Ministry of Health
1 Maritime Square
#12-10 Harborfront Centre
Singapore, 099253
Singapore
Phone: 65 6679 3147
Email: valerie.chua.ruisi@u.nus.edu

Abstract

Background: Telemedicine is increasingly being leveraged, as the need for remote access to health care has been driven by the rising chronic disease incidence and the COVID-19 pandemic. It is also important to understand patients' willingness to pay (WTP) for telemedicine and the factors contributing toward it, as this knowledge may inform health policy planning processes, such as resource allocation or the development of a pricing strategy for telemedicine services. Currently, most of the published literature is focused on cost-effectiveness analysis findings, which guide health care financing from the health system's perspective. However, there is limited exploration of the WTP from a patient's perspective, despite it being pertinent to the sustainability of telemedicine interventions.

Objective: To address this gap in research, this study aims to conduct a systematic review to describe the WTP for telemedicine interventions and to identify the factors influencing WTP among patients with chronic diseases in high-income settings.

Methods: We systematically searched 4 databases (PubMed, PsycINFO, Embase, and EconLit). A total of 2 authors were involved in the appraisal. Studies were included if they reported the WTP amounts or identified the factors associated with patients' WTP, involved patients aged ≥18 years who were diagnosed with chronic diseases, and were from high-income settings.

Results: A total of 11 studies from 7 countries met this study's inclusion criteria. The proportion of people willing to pay for telemedicine ranged from 19% to 70% across the studies, whereas the values for WTP amounts ranged from US $0.89 to US $821.25. We found a statistically significant correlation of age and distance to a preferred health facility with the WTP for telemedicine. Higher age was associated with a lower WTP, whereas longer travel distance was associated with a higher WTP.

Conclusions: On the basis of our findings, the following are recommendations that may enhance the WTP: exposure to the telemedicine intervention before assessing the WTP, the lowering of telemedicine costs, and the provision of patient education to raise awareness on telemedicine's benefits and address patients' concerns. In addition, we recommend that future research be directed at standardizing the reporting of WTP studies with the adoption of a common metric for WTP amounts, which may facilitate the generalization of findings and effect estimates.

(J Med Internet Res 2022;24(4):e33372) doi:10.2196/33372

KEYWORDS

willingness to pay; telemedicine; chronic disease; patients; systematic review; mobile phone

Introduction

Background

Telemedicine is described by the World Health Organization as "healthcare service delivery by healthcare professionals who use information and communication technologies to exchange valid information for disease prevention, diagnosis, treatment, research and evaluation, as well as for the education of healthcare providers." This is of interest, considering the need to advance individual- and community-level health [1]. It can be said that the 21st century, despite its challenges, represented

a bulk of opportunities for the transformation of health care as well as opportunities for leveraging telemedicine's potential in enhancing access to care, reducing travel and waiting times of patients, unnecessary emergency department visits, and the misuse of medication [2]. This is made possible by telemedicine as a platform for patient-practitioner contact with a multidisciplinary team, monitoring of vitals, and symptom assessment [3].

The advancement of telemedicine may be attributed to the rising rates of chronic diseases and increasing disease burdens, with chronic diseases such as chronic obstructive pulmonary disease, asthma, and diabetes representing 11.1 million disability-adjusted life years in high-income countries or 7% of the total disability-adjusted life years [4]. Telemedicine has since increasingly been adopted for supporting and integrating care processes in chronic disease management. This encompasses areas such as educating patients to enhance their self-management, enabling the transfer of information from home-based to clinical settings (ie, telemonitoring), and facilitating contact with health care professionals [5]. On March 11, 2020, the World Health Organization declared the COVID-19 pandemic, suggesting the need for prolonged social distancing [6]. The pandemic's onset has also greatly contributed to the growth of telemedicine use, which enabled distant health care visitations that helped mitigate potential transmissions [7].

While revolutionizing the health care landscape as a new model of care, telemedicine has also been lauded in view of the way it promotes and facilitates self-management [8]. Self-management is defined by the Institute of Medicine as "systematically providing education and supportive interventions by healthcare staff to patients in order to enhance patients' skills and confidence when it comes to managing their health conditions" [9]. Equipping patients for self-management also includes assessing patient progress and issues, partnering patients for goal setting, and offering problem-solving support [10]. Self-management in itself is deemed crucial because it primarily leads to changes in self-efficacy or the individual's confidence in managing their conditions. This potentially has positive and direct influences on health status and use [11]. Practical ways in which telemedicine may facilitate self-management are through areas such as enhancing patient-practitioner contact with a multidisciplinary team [3], educating patients, enabling monitoring and feedback provision, allowing for remote clinical reviews, supporting treatment adherence and lifestyle interventions, and intensive interventions [12].

Self-management has also increasingly been integrated into telemedicine-facilitated chronic disease management because patients who have more than one chronic illness require collaborative care and education in the self-management of chronic disease to ensure the best outcomes for patients [13]. The benefits of telemedicine-aided self-management are also apparent from a systematic review conducted by Hanlon et al [14], where telehealth interventions were found to be appropriate and effective in supporting chronic disease self-management. For example, diabetes and heart failure have the greatest evidence base for positive outcomes following telehealth-supported self-management for disease control and health care use. Although evidence on the impact of telemedicine-based self-management for other types of chronic diseases, especially cancers, is limited, overall, there were no adverse effects reported for any other chronic condition assessed in the study [14].

In addition to telemedicine's efficacy, it is important to also collect information on patients' willingness to pay (WTP) for telemedicine services, as WTP also serves as a surrogate for the demand and acceptability of such services. WTP is defined as the maximum quantity of resources that buyers are willing to forgo during a transaction in exchange for an object. In considering one's WTP, the buyer would evaluate if the trade is beneficial for themselves and would subsequently make the purchase when their WTP is greater than the cost of the object sold [15]. As a concept, WTP is increasingly being used to inform health care policy development [16] whereby developing an understanding of WTP is pertinent to aid decision-making. This occurs as knowing the demand for telemedicine services will inform processes such as planning for sustainable financing, resource allocation, or developing a pricing strategy for telemedicine services. It is also important to study the factors influencing patients' WTP, as this information would be helpful for designing interventions to further enhance WTP. Broadly, WTP may be assessed through the revealed WTP approach and stated preference approach [17]. Specifically for the stated preference approach, the main methods used are

the contingent valuation method (CVM) and discrete choice experiments (DCEs) [18]. CVM involves directly reporting one's WTP for a particular good through methods such as questionnaires and bidding games. It is the most commonly used method of assessing nonmeasurable economic benefits or costs for goods not present in the market at the point of assessment, thus allowing for the estimation of hypothetical goods' monetary value [19]. DCEs elicit preferences and are often used to understand the WTP of various attributes or characteristics of the product. This occurs when people are made to select between alternatives featuring the attributes [20].

Objectives

From our preliminary research, most studies evaluating telemedicine interventions focus on cost-effectiveness analysis [21-24] with limited to no focus on WTP for such interventions. Although cost-effectiveness analysis is important for a health system or policy makers to decide whether to provide financial support for cost-effectiveness interventions, it lacks the patients' perspective (ie, whether patients will adopt this intervention and would be willing to pay partially or fully if such intervention were to be made available). To address this gap, we aim to conduct a systematic review to describe the WTP for telemedicine interventions, and to identify the factors influencing WTP among patients with chronic diseases in high-income settings.

Methods

Information Sources

A systematic literature search was conducted from May to July 2021, abiding by the PRISMA (Preferred Reporting Items for

Systematic Reviews and Meta-Analyses) guidelines as presented in Multimedia Appendix 1 [25]. The databases reviewed were PubMed, PsycINFO, Embase, and EconLit.

Search Strategy

The search strategy was developed on PubMed before translating it to other databases by 1 author (VC) in consultation with the senior author (ST) and university librarian. The main concepts used in this study were *willingness to pay*, *telemedicine*, and *chronic diseases*. For chronic diseases, we included both indexed terms and free-text terms for *chronic diseases*, as well as the 20 chronic conditions included in the Singapore Ministry of Health's Chronic Disease Management Programme [26]. Although synonyms and variations in spelling for the keywords under each of the main concepts were combined in the search strategy using *OR*, each of the concepts were combined using *AND*. For the search, there were no limits with respect to the publication date, although only studies published in English were included in the systematic review. The detailed search strategy for PubMed is included in Multimedia Appendix 2.

Eligibility Criteria

Studies were included if they (1) reported the WTP values for telemedicine or identified factors associated with patients' WTP for telemedicine, (2) included patients aged ≥18 years with chronic diseases, and (3) were based in high-income settings. For our study, high-income settings refer to the countries included under the World Bank list of high-income countries [27]. Only observational studies (cohort, cross-sectional, and case-control studies) and interventional studies were included in the review. Studies were excluded if they were cost-effectiveness analysis studies, protocols, questionnaire validation studies, commentaries, qualitative studies, debates, editorials, newsletters, conference proceedings, letters, and policy reviews.

Study Selection

A preliminary screening of titles and abstracts to exclude articles that were clearly irrelevant to the eligibility criteria was conducted by 1 author (VC). Abstracts of the remaining articles were screened by 1 author (VC) and verified by a second author (ST) to identify potentially relevant articles. Subsequently, full text was retrieved for selected articles and independently assessed for eligibility by 2 authors (VC and ST). After agreement on the list of articles, the reference list of included articles were screened for additional relevant references.

Data Extraction and Synthesis for Study Characteristics, WTP for Telemedicine, and Factors Associated With WTP

Data were extracted from the finalized articles into a data extraction sheet. The extracted data included the aim of the study, the study design, the setting, recruitment, eligibility, the total sample size, the telemedicine intervention or service, the type of intervention, the independent variable, the measure of the independent variable, prior experience with the intervention, the outcome, how WTP was measured, the measure of the outcome, whether the data collected were qualitative or quantitative, the data collection method, the analysis, findings, effect estimates, population descriptions, strengths, and limitations.

This paper uses narrative synthesis as its analytical approach. Narrative synthesis descriptively analyzes and draws comparisons across the studies. This approach is appropriate for our systematic review, as the included studies were heterogeneous, limiting the possibility of statistically pooling effect estimates.

Quality Assessment

The quality assessment for the finalized articles was conducted by 2 authors (VC and JHK) with the help of the Centre for Evidence-based Management guidelines, which may be used to assess the value, relevance, and trustworthiness of cross-sectional studies [28]. Each publication was graded twice independently by VC and JHK. The results in the table were agreed upon by both the authors. There were few disagreements between gradings, and they were resolved in consultation between the authors and a senior coauthor.

Results

Study Selection

Our database search yielded 195 results, of which 184 (94.4%) records were excluded after the title and abstract screening, as well as in consideration of our study's eligibility criteria. A total of 11 articles were included in the final analysis. The study by Losiouk et al [29] was included in our review despite the inclusion criteria stating only to consider patients >18 years with chronic diseases, as the intervention of telemonitoring children with diabetes was meant for parental use.

Data Extraction and Synthesis for Study Characteristics, WTP for Telemedicine, and Factors Associated With WTP

Study Characteristics

The characteristics of the studies included in our systematic review are summarized in Table 1. Of the 11 articles, 3 (27%) were from the United States [30-32], 2 (18%) were from Italy [29,33], 2 (18%) were from Australia [34,35], 1 (9%) was from the United Kingdom [36], 1 (9%) was from South Korea [37], 1 (9%) was from Norway [38], and 1 (9%) was from Belgium [39]. The sample size of the included studies ranged from 23 to 350. The studies' telemedicine interventions can be largely classified into the diagnosis and management of chronic diseases such as cardiovascular diseases [30,32,33,36], diabetes [29,31,37], and skin cancer [34,35]. For the study by Bergmo and Wangberg [38], it was not specified which chronic disease the telemedicine intervention targeted. To estimate WTP, of the 11 studies, 2 (18%) used the CVM, 4 (36%) used a DCE, 1 (9%) used conjoint analysis, 2 (18%) used surveys, and 2 (18%) used questionnaires.

Table 1. Summary of included studies (N=11).

Study	Country	Year	Eligibility	Sample size, n	Intervention	Measurement method for WTP[a]	Patients willing to pay	WTP	Standardized WTP (US $ in 2021)
Bradford et al [30]	United States	HTN[b] clinical study (1999-2000), CHF[c] clinical study (2000-2001)	Patients were recruited from a HTN and CHF study; eligibility criteria was not stated	34	Telemedicine for HTN: peripherals send information on blood pressure, temperature, weight, heart function, and so on. Telemedicine under CHF: weight scale, blood pressure monitor, pulse oximeter, stethoscope, handheld ECG[d], and a base PC platform	CVM[e]	At US $29.96, 32% of the population with HTN would pay out of pocket to access telemedicine. At US $29.96, >45% of the population with CHF would be willing to pay out of pocket for telemedicine access	The dollar amount when randomly varied among patients had a normal distribution with a US $20 mean (per visit)	US $29.96 per visit
Bettiga et al [33]	Italy	N/A[f,g] (paper's year of publication will be used as a reference for currency standardization)	Healthy patients without HTN	350	Mobile health technologies that are connected to the internet and made accessible via smartphones	Survey	N/A	N/A	N/A
Fletcher et al [36]	United Kingdom	June 3-20, 2016	Patients with self-reported HTN who were aged ≥18 years	212	Telemedicine for HTN management	DCE[h]	N/A	A total of €374.74 (US $414.76), €398.98 (US $441.59), and €673.45 (US $745.37) for a 10%, 15%, and 25% reduction in 5-year cardiovascular disease risk, respectively	US $456.99, US $486. 55, and US $821.25 for a 10%, 15%, and 25% reduction in 5-year cardiovascular disease risk, respectively
Losiouk et al [29]	Italy	A clinical trial conducted in 2015	Participants in the baseline and poststudy questionnaire were parents of children with diabetes	167	Web-based telemonitoring service that allowed parents to oversee their child	Questionnaire	N/A	Median WTP of €200 (US $265.68) annually	Median of US $246.40 annually

Study	Country	Year	Eligibility	Sample size, n	Intervention	Measurement method for WTP[a]	Patients willing to pay	WTP	Standardized WTP (US $ in 2021)
Park et al [37]	South Korea	Patients were surveyed from October to November 2009. Physicians were surveyed in January 2010.	Patients surveyed visited outpatient clinics at 2 tertiary care hospitals for diabetes	41	Telemedicine for diabetes management	Conjoint analysis	N/A	Marginal WTP for comprehensiveness of service is ₩16,957 (US $15.26) monthly. WTP for mobile phone over internet-based medical services is ₩15,899 (US $14.31) monthly. WTP for general hospital over physician-based services is ₩15,143 (US $13.63) monthly	US $18.43 monthly for service comprehensiveness. US $17.27 monthly for mobile phone over internet-based services. US $16.46 monthly for general hospital over physician-based services
Snoswell et al [34]	Australia	N/A[i]	Voluntary participants from the SKIN[j] Research Project RCT[k] were included if they owned or could access an iPhone compatible with the study's dermoscopic attachments. Participants were excluded if in the last 5 years, they were diagnosed with melanoma	118	Direct-to-consumer teledermoscopy, which allows patients to interact directly with their dermatologists	DCE	N/A	Marginal WTP of Aus $1.18 (US $0.88) to switch from a GP[l] visitation to mobile teledermoscopy; WTP of Aus $43 (US $32.14) to switch from a GP to a dermatologist; WTP of Aus $117 (US $87.46) to switch to an increased chance of melanoma detection	Marginal WTP of US $0.89 to switch from a GP visitation to mobile teledermoscopy; WTP of US $32.25 to switch from a GP to a dermatologist; WTP of US $87.75 to switch to an increased chance of melanoma detection

Study	Country	Year	Eligibility	Sample size, n	Intervention	Measurement method for WTP[a]	Patients willing to pay	WTP	Standardized WTP (US $ in 2021)
Bergmo and Wangberg [38]	Norway	The RCT was conducted from 2002 to 2003	The study's participants were aged ≥18 years who had internet access and were keen on communicating with their GP electronically	151	Intervention groups were given access to an electronic communication system for communication with their GP	Questionnaire	Of participants, 51% expressed a positive WTP, 21% expressed a WTP of 0, and 28% declined to answer	The mean WTP for the intervention group is €4.52 (US $5.11), whereas that of the control group is €6.78 (US $7.66). WTP are expressed per web-based consultation session	The intervention group has a mean WTP of US $7.36, and the control group has a mean WTP of US $11.04
Scherrenberg et al [39]	Belgium	July to August 2020	Patients From Jessa Hospital	93	Remote cardiac rehabilitation exposure via telephone, video consultations. or live exercise	DCE	Of patients, 70% were willing to pay as much for telerehabilitation as center-based CR[m]	N/A	N/A
Ramchandran et al [31]	United States	2017	Participants had diabetes, had to be cognitively and medically fit to be interviewed or participate in the focus group held in English. Participants had to have a dilated eye examination, be assessed through teleophthalmology, or did not visit an eye physician in the past 2 years	23	Teleophthalmology, which utilizes a camera-based retinopathy exam in noneye care settings for remote image assessment	Survey	Of patients, >50% indicated their WTP to be US $32.38 or US $43.18	WTP was the amount patients usually copay (not stated)	Of patients, >50% indicated their WTP to be US $32.38 or US $43.18

Study	Country	Year	Eligibility	Sample size, n	Intervention	Measurement method for WTP[a]	Patients willing to pay	WTP	Standardized WTP (US $ in 2021)
Spinks et al [35]	Australia	N/A[n]	To be included, participants had to be aged 50-64 years, reside in Queensland, and have moderate or high melanoma risk	35	Teledermoscopy images for review by teledermatologists	DCE	N/A	Participants had a WTP of Aus $110 (US $101.20) to move from choosing between SSE[o], skin cancer clinic, and GP screening to a scenario where teledermoscopy and dermatologists are offered	Participants had a WTP of US $89.70 to move from choosing between SSE, skin cancer, clinic and GP screening to a scenario where teledermoscopy and dermatologists are offered
Bradford et al [32]	United States	Clinical trial conducted from 2000 to 2001.	Patients with CHF discharged from CHF-relevant inpatient stays	126	A PC-dependent system that collected clinical data for care and monitoring of patients with CHF	DBDC[p] CVM	Of patients, 55% had a WTP of US $29.96 for telemedicine rather than in-person care at the physician's office. Of patients, 19% had a WTP of US $59.91 for telemedicine rather than in-person care at the physician's office	WTP of US $20 and US $40 per visit	WTP of US $29.96 and US $59.91 per visit

[a]WTP: willingness to pay.

[b]HTN: hypertension.

[c]CHF: chronic heart failure.

[d]ECG: electrocardiogram.

[e]CVM: contingent valuation method.

[f]N/A: not applicable.

[g]The paper's year of publication in 2020 will be cited as the year in which the study is conducted, as the time frame for when the intervention was conducted was not provided.

[h]DCE: discrete choice experiment.

[i]The paper's year of publication in 2018 will be cited as the year in which the study was conducted, as the year of study was not reported in the paper.

[j]SKIN: Skin Innovation.

[k]RCT: randomized control trial.

[l]GP: general practitioner.

[m]CR: cardiac rehabilitation.

[n]The paper's year of publication in 2016 will be cited as the year in which the study was conducted, as the year of study was not reported in the paper.

[o]SSE: skin self-examination.

[p]DBDC: double-bounded dichotomous choice.

WTP for Telemedicine

The results for the WTP for telemedicine were expressed in terms of the proportion of patients who were willing to pay for the intervention, as well as the specific WTP amounts highlighted across the 11 studies (Table 1). Regarding the proportion of patients willing to pay for telemedicine, WTP percentages across the 11 studies ranged from 19% to 70%. The study by Bradford et al [32] reported 19% of patients were willing to pay for telemedicine rather than in-person care in the physician's office when the price was raised from US $20 to US $40. In contrast, the study by Scherrenberg et al [39] expressed that 70% of patients were willing to pay as much for telerehabilitation as center-based cardiac rehabilitation.

Regarding the WTP amount, the monetary values highlighted from all the studies were standardized to a baseline currency of US $ in 2021 for comparison between studies. This was done by converting the WTP values into US $ for the same year that the study was conducted. The converted WTP values in US $ were then converted into US $ in 2021 by considering inflation. The WTP values provided by the studies ranged from US $0.89 to US $821.25. The WTP of US $0.89 was the marginal WTP arising from switching from a general practitioner (GP) visitation to mobile teledermoscopy, whereas the US $821.25 was the annual WTP for a 25% reduction in 5-year cardiovascular disease risk. Different metrics for the WTP values were provided across the studies, for example, WTP per month, WTP per year, and WTP per session.

Factors Associated With the WTP

Across the studies, the association between different sociodemographic, socioeconomic, and health service variables with WTP was reported. For the purpose of this review, we have summarized variables that were reported by at least two included studies, namely, age, income, gender, travel time, marital status, and ethnicity (Table 2) The correlations between age and distance from health facilities with the WTP for telemedicine were negative (ie, the higher the age, the lower the WTP for telemedicine) and positive (ie, the longer the traveling distance, the higher the WTP for telemedicine), respectively. Both correlations were statistically significant. Whereas being female and having higher income were associated with higher WTP, being married was associated with a lower WTP for telemedicine. However, these correlations are not statistically significant. For ethnicity, although there was conflicting evidence from 2 separate studies, both of these estimates were not statistically significant.

Table 2. Factors associated with WTP[a] in the included studies.

Study	Demographics				Socioeconomic income	Health service distance to preferred health facility
	Gender (female)	Age	Married	Ethnicity		
Bradford et al [30]	+[b]	−−[c]	−[d]	−	+	++[e]
Bettiga et al [33]	N/A[f]	N/A	N/A	N/A	N/A	N/A
Fletcher et al [36]	N/A	N/A	N/A	N/A	N/A	N/A
Losiouk et al [29]	N/A	N/A	N/A	N/A	N/A	N/A
Park et al [37]	N/A	−−	N/A	N/A	N/A	N/A
Snoswell et al [34]	N/A	N/A	N/A	N/A	N/A	N/A
Bergmo and Wangberg [38]	N/A	++	N/A	N/A	+	N/A
Scherrenberg et al [39]	N/A	N/A	N/A	N/A	N/A	N/A
Ramchandran et al [31]	N/A	N/A	N/A	N/A	N/A	N/A
Spinks et al [35]	N/A	N/A	N/A	N/A	N/A	N/A
Bradford et al [32]	+	−−	−	+	+	++

[a]WTP: willingness to pay.

[b]The effect of the variable is positive and nonsignificant.

[c]The effect of the variable is negative and significant.

[d]The effect of the variable is negative and nonsignificant.

[e]The effect of the variable is positive and significant.

[f]N/A: not applicable; the effect of the variable is not applicable to this study.

Quality Appraisal

The details of the quality assessment conducted for the 11 included studies are presented in Table 3. While the total number of *Yes* responses under each study's quality assessment ranged from 4 to 10 (range 0-12), the total number of *No* responses ranged from 2 to 6 (range 0-12). The total number of *Can't tell* responses yielded by each study's quality assessment ranged from 0 to 3 (range 0-12). Across the 11 studies, 3 (27%) studies [34,35,37] received the highest rating of 10 *Yes* responses, whereas 2 (18%) studies [30,39] received the lowest rating of 5 *Yes* responses. All the studies except 1 [33] received a *Yes* response to the question on whether the study has a clearly focused question that helps provide relevant context to the reader

in understanding the study. All the studies except 1 [35] received a *No* response to the question on whether there were prestudy considerations of statistical power. Across the board, this may not be strictly followed or widely adopted, potentially owing to the difficulty in sampling sufficient patients who met the studies' selection criteria. All studies except 1 [39] received a *Yes* response for whether statistical significance was assessed. All studies received a *Yes* response for the applicability of results to the reviewer's organization as findings were found to be relevant to the Singapore context. This is in consideration of the chronic disease prevalence which accounted for 29.3% of deaths in 2019, highlighting the importance of exploring the WTP for telemedicine interventions in the context of chronic diseases [40].

Table 3. Quality appraisal of the included studies.

Question	Bradford et al [30]	Bettiga et al [33]	Fletcher et al [36]	Losiouk et al [29]	Park et al [37]	Snoswell et al [34]	Bergmo and Wangberg [38]	Scherrenberg et al [39]	Ramchandran et al [31]	Spinks et al [35]	Bradford et al [32]
1. Did the study address a clearly focused question or issue?	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2. Is the research method (study design) appropriate for answering the research question?	Yes	Yes	Yes	Yes	Yes	Yes	Cannot tell	Yes	Cannot tell	Yes	Yes
3. Is the method for the selection of the participants (employees, teams, divisions, and organizations) clearly described?	Cannot tell	Cannot tell	Yes	Cannot tell	Yes	Yes	Yes	Yes	Yes	Yes	Yes
4. Could the way the sample was obtained introduce (selection) bias	Cannot tell	Cannot tell	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
5. Was the sample of participants representative with regard to the population to which the findings will be referred?	Cannot tell	Cannot tell	Cannot tell	No	Yes	Yes	Yes	Cannot tell	Yes	Yes	Yes
6. Was the sample size based on prestudy considerations of statistical power?	No	No	No	No	No	No	No	No	No	Yes	No
7. Was a satisfactory response rate achieved?	Cannot tell	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No
8. Are the measurements (questionnaires) likely to be valid and reliable?	Yes	Yes	Yes	No	Yes	Yes	No	No	Yes	No	Yes
9. Was the statistical significance assessed?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
10. Are CIs given for the main results?	No	No	Yes	No	No	Yes	Yes	No	No	Yes	No
11. Could there be confounding factors that have not been accounted for?	No	Yes	No	Yes	Yes	No	Cannot tell	No	Yes	Yes	Yes
12. Can the results be applied to your organization?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Discussion

Principal Findings

To our knowledge, this study is the first systematic review summarizing the WTP in patients with chronic diseases for telemedicine in high-income settings, and the first systematic review summarizing the factors associated with WTP. We reported the proportion willing to pay for a telemedicine intervention to vary between 19% and 70%, with the WTP amount ranging from US $0.89 to US $821.25. In addition, we found age and distance to preferred health facility as the only reported factors to be significantly associated with the WTP for telemedicine interventions. Although gender, marital status, ethnicity, and income were reported to be associated with the WTP for telemedicine intervention, this association was not statistically significant.

For age, whereas 3 studies [30,37,38] reported a statistically significant negative correlation with the WTP for telemedicine intervention, 1 [38] reported a statistically significant positive correlation. The former is consistent with the literature, as adults

aged >65 years are reported to be less inclined to use technology-based interventions than their younger counterparts [41]. This may be attributed to how pensioners tend to fear incurring high costs that come with purchasing electronics or investing in home-based health monitoring systems required by some telemedicine interventions [42]. This may also be attributed to the resistance of older patients toward new technologies, which may explain lower rates of acceptance for such technology-enabled interventions. Older patients also tend to prefer personal contact with health care providers, perceiving distance-based services such as telemedicine to be insufficient to meet this need [43], thus potentially lowering their WTP for telemedicine. On the contrary, Bergmo and Wangberg [38] reported older age to be significantly associated with a higher WTP for telemedicine services. However, on further reviewing their sample, we found that most included participants were high daily internet users, which may not be representative of the general population, including the older subgroup. Furthermore, more than half of the sample was college educated, which is reported in the literature to increase the acceptance of technology [44]. Hence, for reporting the association of age with the WTP for telemedicine, this study was identified as a potential outlier, and our study concludes a negative and statistically significant correlation following the consistency in results reported by other included studies [30,32,37]. For the variable distance to preferred health facility, 2 studies [30,32] reported a statistically significant positive correlation where further distance is associated with a higher WTP for telemedicine intervention. This WTP may be explained by the convenience afforded by telemedicine services, which may be conducted remotely in patients' homes. Convenience was also reported as the main determinant for WTP for people aged ≥18 years looking to have consultations with their physicians over video conferencing [45]. This WTP may also be explained by how telemedicine was found to reduce associated financial burdens by reducing traveling costs, as telemedicine is conducted remotely [46], thus reducing costs to productivity [47], which patients are likely to be willing to pay to mitigate.

Moving forward with recommendations on how to enhance the WTP for telemedicine, there are also several other trends in patients' WTP for telemedicine we could identify as we drew comparisons across the 11 studies. These pertain to the influence of prior telemedicine exposure, the impact of telemedicine costs, and the importance of patient education in contributing to WTP.

Across 3 studies [29,38,39], participants were exposed to the telemedicine intervention before being asked about their WTP for telemedicine. With the availability of this information, we can review the implications of having prior exposure on WTP. Losiouk et al [29] reported a median WTP of US $246.40 annually by parents of children with diabetes who were enrolled into a clinical trial for telemonitoring. In the second study [38], the control group communicated with their GPs as per usual (office visitations, and telephones), whereas the intervention group had access to a messaging system for a period of 1 year. Approximately 51% (77/151) had expressed a WTP with a positive value for the telemedicine intervention. In the third study [39], out of all the patients who underwent cardiac telerehabilitation sessions, approximately 70% (highest proportion across all 11 reviewed studies) were willing to pay as much for telerehabilitation as compared with a center-based session for cardiac rehabilitation. Considering that the proportion of patients willing to pay for telemedicine in this review ranged from 19% to 70%, the proportion of people who were willing to pay provided by studies with prior exposure to telemedicine intervention were relatively on the higher side. A similar trend was observed for the WTP values, which ranged from US $0.89 to US $821.25 in this review. The values provided by the study with prior exposure to telemedicine intervention were on the higher side as well. The findings from our systematic review therefore suggest that experience with the telemedicine technology in question may contribute to a greater WTP by patients. Further research may therefore explore the extent to which exposure to a telemedicine intervention may influence the WTP for telemedicine, as well as the differential impact on patients of different demographics (age, gender, etc).

Telemedicine costs were cited as a major barrier to access by 36% (4/11) of the studies. According to 1 study [37], the attribute ranked most highly in importance was cost, which had a relative importance estimated at 29%. According to another study [34], the preference weight for the telemedicine cost attribute was extremely significant at P=.001. According to Ramchandran et al [31], almost all participants shared that their limited insurance coverage for medical care, having fixed income and limited budget were barriers for obtaining dilated eye examinations. Many patients expressed their wish to know the cost of the examination before deciding and would be more keen on participating if they knew that insurance would pay for the service. A few stated that they would pay for eye exams because they knew the value of it if they were able to afford it. In view of the overwhelming evidence in favor of lowering telemedicine costs and prices, this spotlights the need to supply provisions that promote telemedicine access and mitigate the risks of falling into income poverty, which is no exception even for high-income countries [48]. In particular, the effect of reimbursement as a provision had been demonstrated in how a notable barrier in telehealth adoption in the United States is the shortage of significant Medicare, Medicaid, and commercial insurance reimbursements. Cost reductions in general have also been found to contribute to higher telehealth implementation [7]. In view of equity concerns, interventions that allow for cost alleviation should also be more directed to groups of people whose WTP for health services is more greatly influenced by their financial ability to pay. These groups of people who are more cost-sensitive are the low-income earners, those without university education, those who are older, and those with poor health status [49].

Patient education also plays a crucial role in eliciting patient demand, acceptance, and thus the WTP for telemedicine services. It is important to raise awareness about the efficacy of telemedicine services while simultaneously addressing patient concerns, especially because telemedicine is a relatively new model of care for many, even within high-income settings. The commonalities in findings consolidated from across the 11 articles were able to provide directions for educating patients. First, it is important to highlight the potential risk reductions brought about by the ease and regularity of telemedicine-based

services (ie, telemonitoring and teledermoscopy) as compared with inpatient clinical visits, which may be more infrequent or less accessible. The WTP for risk reductions had been reported by one of the studies [36], in which participants were willing to pay for larger reductions in heart disease risk, with a WTP of US $456.99, US $486.55, and US $821.25 for a 10%, 15%, and 25% reduction in 5-year cardiovascular disease risk, respectively. Another study [34] also highlighted how consumers demonstrated the largest WTP within the study of US $87.75 to switch to telemedicine alternatives when there was an increased chance of melanoma detection. Second, it is important to address the concerns that patients may have toward telemedicine that potentially limit their use. This was evidenced by Bergmo and Wangberg [38], where 48% of patients surveyed were unwilling to pay for electronic contact with GP potentially because of there being fewer than expected benefits, a hypothetical bias, or a simple preference for face-to-face consultations. Ramchandran et al [31] also found that older participants had a strong preference for seeing an eye physician personally because of the value they place on the relationship with the health care provider, as well as the perceived thoroughness and expertise of the examination. When educating patients, emphasis should therefore be placed on the reliability and quality of teleophthalmology because web-based care, which is frequently paired with remote patient monitoring, can improve physical assessments even without the physical presence of the practitioner [50]. Furthermore, there are also studies that show a trend of higher satisfaction for physician interpersonal skills and patient-centered communications during telemedicine [51], which may be meaningful to consider. Third, the additional benefits of telemedicine that patients themselves may not immediately foresee or intuitively consider may be brought to their awareness. An example of such additional benefits that can be highlighted to patients during patient education interventions would be the reduction in parental fatigue, which is deemed the worst effect of diabetes management yet the most improvable item in daily telemonitoring when measured during the postintervention survey.

Strengths

The following are the strengths of our review. First, to our knowledge, we are the first to conduct a systematic review on the WTP for telemedicine among patients with chronic diseases in high-income settings, thus providing a precedence for which this topic may be further explored. Second, our study used multiple databases, namely, PubMed, PsycINFO, Embase, and EconLit. This allowed for the comprehensive capture of resources. Third, as our review was conducted in accordance with the PRISMA guidelines, it is systematic and transparent. Fourth, the risk of bias was conducted for the 11 included studies following the Centre for Evidence-based Management guidelines [28]. Having 2 authors review the quality of included studies further helped to strengthen the reliability of the risk of bias assessment conducted. Finally, we did not place limits on the date of publication when it came to the selection of studies for inclusion in our review.

Limitations

The following are our study limitations. The main limitation was the heterogeneity across the studies included in this review, which made it hard to draw conclusions beyond descriptive comparisons. First, there was no consistency in reporting the outcomes of interest, as some studies did not report the proportion of patients willing to pay for telemedicine [29,33-37], whereas some studies did not report the WTP values [33,39]. Because of this missing information, the general conclusions on WTP values or proportion of patients willing to pay for telemedicine may not necessarily be applicable to all the studies included in our systematic review. Second, with reference to Table 2, only 36% (4/11) of the studies [30,32,37,38] provided information on how specific variables correlated with the WTP for telemedicine. This limited the amount of information available to draw conclusions on factors associated with WTP. Third, the diseases represented in the selected studies may be broadly classified into diabetes, cardiovascular diseases, and skin cancer, thus only accounting for a limited number of diseases. Fourth, different methods were used in eliciting WTP across the studies, each with their own limitations as well with potential implications on the validity of the WTP values collected. Another limitation was the inclusion of articles published only in English and the exclusion of gray literature, which may have implications on the comprehensiveness of the reviewed evidence. There are also inherent differences in health care systems, financing of health care, and socioeconomic context across included studies from different countries. Overall, we acknowledge the large heterogeneity in patient populations, interventions, health care settings, interventions, and outcome measures across the 11 studies, which may potentially limit the generalizability of our findings. However, as the literature on the WTP for telemedicine is sparse at the moment, we were unable to focus on selecting papers with specific interventions, patient populations, settings, and outcome measures in this systematic review.

We also agree and acknowledge that it would be meaningful to delve more into the diverse and emerging forms of telehealth. Telemedicine technologies have developed to include a range of services such as synchronous and asynchronous consultations [52], which are of interest as they represent the latest advances in telecommunication technology, and networks in health care and promise in service delivery [53]. Artificial intelligence chatbots have also been found to provide a personalized medical consultation experience, provide immediate access to medical information, provide diagnosis recommendations, and connect patients to health care providers beyond their immediate communities [54]. However, WTP studies involving novel telemedicine technologies such as synchronous and asynchronous consultations and artificial intelligence chatbots are not yet well established, resulting in a limited range of telemedicine services represented in our paper despite the emergence of such forms of telemedicine over the years. Hence, we would recommend that further WTP research for telemedicine can compile their findings for the more novel telemedicine forms such as the ones we have identified. This would be a good starting point before empirical evidence may be synthesized in subsequent systematic reviews.

Conclusions

On the basis of the findings of this systematic review on the WTP for telemedicine among patients with chronic health conditions, we conclude that the WTP for telemedicine varied considerably across the literature, ranging from 19% to 70%. In addition, we found age and distance to preferred health facility as the only reported factors to be significantly associated with the WTP for telemedicine interventions. The following are the practical recommendations based on our findings, which may be used to guide future interventions to boost the WTP for telemedicine. Prior exposure to the telemedicine intervention, as well as mitigating telemedicine costs, where possible, may enhance the acceptability and WTP for such interventions. Patient education was also found to be important for raising awareness on telemedicine's benefits and addressing patients' concerns. In view of the heterogeneity in the existing literature on WTP, future research efforts should focus on standardizing the conduct and reporting of such WTP studies so as to facilitate generalization of findings, pooling of effect estimates and generating of actionable insights to enhance the outreach of telemedicine interventions to patients with chronic diseases.

Conflicts of Interest

CHGK is on the board of a telerehabilitation company called T-Rehab Pte Ltd.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) study flowchart.
[DOCX File , 65 KB - jmir_v24i4e33372_app1.docx]

Multimedia Appendix 2
Search strategy for the PubMed database.
[DOCX File , 29 KB - jmir_v24i4e33372_app2.docx]

References

1. World Health Organization. A Health Telematics Policy in Support of WHO'S Health-For-All Strategy for Global Development. 1998. URL: https://apps.who.int/iris/bitstream/handle/10665/63857/WHO_DGO_98.1.pdf [accessed 2022-03-17]
2. Rutledge C, Kott K, Schweickert P, Poston R, Fowler C, Haney T. Telehealth and eHealth in nurse practitioner training: current perspectives. Adv Med Educ Pract 2017;8:399-409 [FREE Full text] [doi: 10.2147/AMEP.S116071] [Medline: 28721113]
3. Penedo FJ, Oswald LB, Kronenfeld JP, Garcia SF, Cella D, Yanez B. The increasing value of eHealth in the delivery of patient-centred cancer care. Lancet Oncol 2020 May;21(5):e240-e251 [FREE Full text] [doi: 10.1016/S1470-2045(20)30021-8] [Medline: 32359500]
4. Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJ. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. Lancet 2006 May 27;367(9524):1747-1757. [doi: 10.1016/S0140-6736(06)68770-9] [Medline: 16731270]
5. Wootton R. Twenty years of telemedicine in chronic disease management--an evidence synthesis. J Telemed Telecare 2012 Jun;18(4):211-220 [FREE Full text] [doi: 10.1258/jtt.2012.120219] [Medline: 22674020]
6. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. Science 2020 May 22;368(6493):860-868 [FREE Full text] [doi: 10.1126/science.abb5793] [Medline: 32291278]
7. Gajarawala SN, Pelkowski JN. Telehealth benefits and barriers. J Nurse Pract 2021 Feb;17(2):218-221 [FREE Full text] [doi: 10.1016/j.nurpra.2020.09.013] [Medline: 33106751]
8. Steventon A, Bardsley M, Billings J, Dixon J, Doll H, Hirani S, Whole System Demonstrator Evaluation Team. Effect of telehealth on use of secondary care and mortality: findings from the Whole System Demonstrator cluster randomised trial. BMJ 2012 Jun 21;344:e3874 [FREE Full text] [doi: 10.1136/bmj.e3874] [Medline: 22723612]
9. Practice facilitation handbook. Agency for Healthcare Research and Quality. URL: https://www.ahrq.gov/ncepcr/tools/pf-handbook/mod21.html#pearson [accessed 2022-03-07]
10. -. IOM: healthcare delivery, education need improvement. RN 2003;66(5):14.
11. Lorig K, Seleznick M, Lubeck D, Ung E, Chastain RL, Holman HR. The beneficial outcomes of the arthritis self-management course are not adequately explained by behavior change. Arthritis Rheum 1989 Jan;32(1):91-95. [doi: 10.1002/anr.1780320116] [Medline: 2912467]
12. Inglis S, Clark R, Dierckx R, Prieto-Merino D, Cleland J. Structured telephone support or non-invasive telemonitoring for patients with heart failure. Cochrane Database Syst Rev 2015 Oct 31;2015(10):CD007228 [FREE Full text] [doi: 10.1002/14651858.CD007228.pub3] [Medline: 26517969]

13. Allegrante JP, Wells MT, Peterson JC. Interventions to support behavioral self-management of chronic diseases. Annu Rev Public Health 2019 Apr 01;40:127-146 [FREE Full text] [doi: 10.1146/annurev-publhealth-040218-044008] [Medline: 30601717]
14. Hanlon P, Daines L, Campbell C, McKinstry B, Weller D, Pinnock H. Telehealth interventions to support self-management of long-term conditions: a systematic metareview of diabetes, heart failure, asthma, chronic obstructive pulmonary disease, and cancer. J Med Internet Res 2017 May 17;19(5):e172 [FREE Full text] [doi: 10.2196/jmir.6688] [Medline: 28526671]
15. Plassmann H, O'Doherty J, Rangel A. Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. J Neuroscience 2007 Sep 12;27(37):9984-9988. [doi: 10.1523/jneurosci.2131-07.2007]
16. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Econ 2012 Feb;21(2):145-172. [doi: 10.1002/hec.1697] [Medline: 22223558]
17. Breidert H, Hahsler M, Reutterer T. A review of methods for measuring willingness-to-pay. Innov Marketing 2015;1(4). [doi: 10.1007/978-3-8350-9244-0_4]
18. A review of the discrete choice experiment - with emphasis on its application in health care. Syddansk Universitet. URL: https://portal.findresearcher.sdu.dk/en/publications/a-review-of-the-discrete-choice-experiment-with-emphasis-on-its-a [accessed 2022-03-07]
19. Chen Y, Ying Y, Chang K, Hsieh Y. Study of patients' willingness to pay for a cure of chronic obstructive pulmonary disease in Taiwan. Int J Environ Res Public Health 2016 Mar 01;13(3):273. [doi: 10.3390/ijerph13030273] [Medline: 26938547]
20. Siyaranamual M, Amalia M, Yusuf A, Alisjahbana A. Consumers' willingness to pay for electricity service attributes: a discrete choice experiment in urban Indonesia. Energy Reports 2020 Nov;6:562-571. [doi: 10.1016/j.egyr.2020.02.018]
21. Akiyama M, Yoo B. A systematic review of the economic evaluation of telemedicine in Japan. J Prev Med Public Health 2016 Jul;49(4):183-196 [FREE Full text] [doi: 10.3961/jpmph.16.043] [Medline: 27499161]
22. Henderson C, Knapp M, Fernández JL, Beecham J, Hirani SP, Cartwright M, Whole System Demonstrator evaluation team. Cost effectiveness of telehealth for patients with long term conditions (Whole Systems Demonstrator telehealth questionnaire study): nested economic evaluation in a pragmatic, cluster randomised controlled trial. BMJ 2013 Mar 20;346:f1035 [FREE Full text] [doi: 10.1136/bmj.f1035] [Medline: 23520339]
23. Nordyke RJ, Appelbaum K, Berman MA. Estimating the impact of novel digital therapeutics in type 2 diabetes and hypertension: health economic analysis. J Med Internet Res 2019 Oct 09;21(10):e15814 [FREE Full text] [doi: 10.2196/15814] [Medline: 31599740]
24. Oksman E, Linna M, Hörhammer I, Lammintakanen J, Talja M. Cost-effectiveness analysis for a tele-based health coaching program for chronic disease in primary care. BMC Health Serv Res 2017 Feb 15;17(1):138 [FREE Full text] [doi: 10.1186/s12913-017-2088-4] [Medline: 28202032]
25. 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]
26. Chronic Disease Management Programme (CDMP). Ministry of Health Singapore. URL: https://www.moh.gov.sg/policies-and-legislation/chronic-disease-management-programme-(cdmp) [accessed 2022-02-18]
27. High income. The World Bank. URL: https://data.worldbank.org/country/XD [accessed 2022-02-18]
28. Critical appraisal of a cross-sectional study (survey). CEBMa. URL: https://cebma.org/wp-content/uploads/Critical-Appraisal-Questions-for-a-Cross-Sectional-Study-July-2014-1.pdf [accessed 2022-02-18]
29. Losiouk E, Lanzola G, Del Favero S, Boscari F, Messori M, Rabbone I, et al. Parental evaluation of a telemonitoring service for children with Type 1 Diabetes. J Telemed Telecare 2017 Mar 26;24(3):230-237. [doi: 10.1177/1357633x17695172]
30. Bradford WD, Kleit A, Krousel-Wood M, Re RM. Comparing willingness to pay for telemedicine across a chronic heart failure and hypertension population. Telemed J E Health 2005 Aug;11(4):430-438. [doi: 10.1089/tmj.2005.11.430] [Medline: 16149888]
31. Ramchandran RS, Yilmaz S, Greaux E, Dozier A. Patient perceived value of teleophthalmology in an urban, low income US population with diabetes. PLoS One 2020;15(1):e0225300 [FREE Full text] [doi: 10.1371/journal.pone.0225300] [Medline: 31917793]
32. Bradford WD, Kleit AN, Krousel-Wood MA, Re RM. Willingness to pay for telemedicine assessed by the double-bounded dichotomous choice method. J Telemed Telecare 2004;10(6):325-330. [doi: 10.1258/1357633042601991] [Medline: 15603629]
33. Bettiga D, Lamberti L, Lettieri E. Individuals' adoption of smart technologies for preventive health care: a structural equation modeling approach. Health Care Manag Sci 2020 Jun;23(2):203-214. [doi: 10.1007/s10729-019-09468-2] [Medline: 30684067]
34. Snoswell CL, Whitty JA, Caffery LJ, Loescher LJ, Gillespie N, Janda M. Direct-to-consumer mobile teledermoscopy for skin cancer screening: preliminary results demonstrating willingness-to-pay in Australia. J Telemed Telecare 2018 Oct 22;24(10):683-689. [doi: 10.1177/1357633x18799582]
35. Spinks J, Janda M, Soyer HP, Whitty JA. Consumer preferences for teledermoscopy screening to detect melanoma early. J Telemed Telecare 2015 May 29;22(1):39-46. [doi: 10.1177/1357633x15586701]

36. Fletcher B, Hinton L, McManus R, Rivero-Arias O. Patient preferences for management of high blood pressure in the UK: a discrete choice experiment. Br J Gen Pract 2019 Aug 12;69(686):e629-e637. [doi: 10.3399/bjgp19x705101]
37. Park H, Chon Y, Lee J, Choi I, Yoon K. Service design attributes affecting diabetic patient preferences of telemedicine in South Korea. Telemed J E Health 2011;17(6):442-451 [FREE Full text] [doi: 10.1089/tmj.2010.0201] [Medline: 21631382]
38. Bergmo TS, Wangberg SC. Patients' willingness to pay for electronic communication with their general practitioner. Eur J Health Econ 2007 Jun;8(2):105-110. [doi: 10.1007/s10198-006-0014-5] [Medline: 17186205]
39. Scherrenberg M, Falter M, Dendale P. Patient experiences and willingness-to-pay for cardiac telerehabilitation during the first surge of the COVID-19 pandemic: single-centre experience. Acta Cardiol 2021 Apr;76(2):151-157. [doi: 10.1080/00015385.2020.1846920] [Medline: 33183162]
40. Heart statistics. Singapore Heart Foundation. URL: https://www.myheart.org.sg/my-heart/heart-statistics/singapore-statistics/ [accessed 2022-02-18]
41. The web at 25 in the U.S. Pew Research Center. URL: https://www.pewresearch.org/internet/2014/02/27/the-web-at-25-in-the-u-s/ [accessed 2022-02-18]
42. Overcoming the psychological barriers to telemedicine: empowering older Americans to use remote health monitoring services. Doc Player. URL: https://docplayer.net/13773466-Overcoming-the-psychological-barriers-to-telemedicine-empowering-older-americans-to-use-remote-health-monitoring-services.html [accessed 2022-02-18]
43. Bujnowska-Fedak M, Grata-Borkowska U. Use of telemedicine-based care for the aging and elderly: promises and pitfalls. Smart Healthcare Technol Telehealth 2015 May;2015:91. [doi: 10.2147/shtt.s59498]
44. Chou C, Chang C, Lee T, Chou H, Mills M. Technology acceptance and quality of life of the elderly in a telecare program. Comput Inform Nurs 2013 Jul;31(7):335-342. [doi: 10.1097/NXN.0b013e318295e5ce] [Medline: 23728446]
45. Roettl J, Bidmon S, Terlutter R. What predicts patients' willingness to undergo online treatment and pay for online treatment? Results from a web-based survey to investigate the changing patient-physician relationship. J Med Internet Res 2016 Feb 04;18(2):e32 [FREE Full text] [doi: 10.2196/jmir.5244] [Medline: 26846162]
46. Bator EX, Gleason JM, Lorenzo AJ, Kanaroglou N, Farhat WA, Bägli DJ, et al. The burden of attending a pediatric surgical clinic and family preferences toward telemedicine. J Pediatr Surg 2015 Oct;50(10):1776-1782. [doi: 10.1016/j.jpedsurg.2015.06.005] [Medline: 26195452]
47. Telehealth benefits and adoption: connecting people and providers (full). Canada Health Infoway. URL: https://www.infoway-inforoute.ca/en/component/edocman/resources/reports/333-telehealth-benefits-and-adoption-connecting-people-and-providers-full [accessed 2022-02-18]
48. Callander EJ, Fox H, Lindsay D. Out-of-pocket healthcare expenditure in Australia: trends, inequalities and the impact on household living standards in a high-income country with a universal health care system. Health Econ Rev 2019 Mar 11;9(1):10 [FREE Full text] [doi: 10.1186/s13561-019-0227-9] [Medline: 30859357]
49. Tambor M, Pavlova M, Rechel B, Golinowska S, Sowada C, Groot W. Willingness to pay for publicly financed health care services in Central and Eastern Europe: evidence from six countries based on a contingent valuation method. Soc Sci Med 2014 Sep;116:193-201. [doi: 10.1016/j.socscimed.2014.07.009] [Medline: 25016327]
50. Gardenier D, Ballard-Hernandez J, O'Donnell M. Is telehealth here to stay? J Nurse Practitioners 2021 Mar;17(3):265-266. [doi: 10.1016/j.nurpra.2020.12.008]
51. Agha Z, Schapira RM, Laud PW, McNutt G, Roter DL. Patient satisfaction with physician-patient communication during telemedicine. Telemed J E Health 2009 Nov;15(9):830-839. [doi: 10.1089/tmj.2009.0030] [Medline: 19919189]
52. Serrano CI, Shah V, Abràmoff MD. Use of expectation disconfirmation theory to test patient satisfaction with asynchronous telemedicine for diabetic retinopathy detection. Int J Telemed Appl 2018 Oct 11;2018:7015272-7015214 [FREE Full text] [doi: 10.1155/2018/7015272] [Medline: 30405712]
53. Dario C, Toffanin R, Calcaterra F, Saccavini C, Stafylas P, Mancin S, et al. Telemonitoring of type 2 diabetes mellitus in Italy. Telemed J E Health 2017 Feb;23(2):143-152. [doi: 10.1089/tmj.2015.0224] [Medline: 27379995]
54. Palanica A, Flaschner P, Thommandram A, Li M, Fossat Y. Physicians' perceptions of chatbots in health care: cross-sectional web-based survey. J Med Internet Res 2019 Apr 05;21(4):e12887 [FREE Full text] [doi: 10.2196/12887] [Medline: 30950796]

Abbreviations

CVM: contingent valuation method
DCE: discrete choice experiment
GP: general practitioner
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
WTP: willingness to pay

Edited by T Leung; submitted 04.09.21; peer-reviewed by D Gunasekeran, J Multmeier; comments to author 07.10.21; revised version received 16.11.21; accepted 18.01.22; published 13.04.22.

Please cite as:
Chua V, Koh JH, Koh CHG, Tyagi S
The Willingness to Pay for Telemedicine Among Patients With Chronic Diseases: Systematic Review
J Med Internet Res 2022;24(4):e33372
URL: https://www.jmir.org/2022/4/e33372
doi:10.2196/33372
PMID:35416779

Review

Accuracy of Wearable Transdermal Alcohol Sensors: Systematic Review

Eileen Brobbin[1], BSc, MSc; Paolo Deluca[1], PhD; Sofia Hemrage[1], BSc, MSc; Colin Drummond[1], MB, ChB, MD

Department of Addictions, King's College London, London, United Kingdom

Corresponding Author:
Eileen Brobbin, BSc, MSc
Department of Addictions
King's College London
Addiction Science Building
4 Windsor Walk
London, SE5 8BB
United Kingdom
Phone: 44 0207 836 5454
Email: eileen.brobbin@kcl.ac.uk

Abstract

Background: There are a range of wearable transdermal alcohol sensors that are available and are being developed. These devices have the potential to monitor alcohol consumption continuously over extended periods in an objective manner, overcoming some of the limitations of other alcohol measurement methods (blood, breath, and urine).

Objective: The objective of our systematic review was to assess wearable transdermal alcohol sensor accuracy.

Methods: A systematic search of the CINAHL, Embase, Google Scholar, MEDLINE, PsycINFO, PubMed, and Scopus bibliographic databases was conducted in February 2021. In total, 2 team members (EB and SH) independently screened studies for inclusion, extracted data, and assessed the risk of bias. The methodological quality of each study was appraised using the Mixed Methods Appraisal Tool. The primary outcome was transdermal alcohol sensor accuracy. The data were presented as a narrative synthesis.

Results: We identified and analyzed 32 studies. Study designs included laboratory, ambulatory, and mixed designs, as well as randomized controlled trials; the length of time for which the device was worn ranged from days to weeks; and the analyzed sample sizes ranged from 1 to 250. The results for transdermal alcohol concentration data from various transdermal alcohol sensors were generally found to positively correlate with breath alcohol concentration, blood alcohol concentration, and self-report (moderate to large correlations). However, there were some discrepancies between study reports; for example, WrisTAS sensitivity ranged from 24% to 85.6%, and specificity ranged from 67.5% to 92.94%. Higher malfunctions were reported with the BACtrack prototype (16%-38%) and WrisTAS (8%) than with SCRAM (2%); however, the former devices also reported a reduced time lag for peak transdermal alcohol concentration values when compared with SCRAM. It was also found that many companies were developing new models of wearable transdermal alcohol sensors.

Conclusions: As shown, there is a lack of consistency in the studies on wearable transdermal alcohol sensor accuracy regarding study procedures and analyses of findings, thus making it difficult to draw direct comparisons between them. This needs to be considered in future research, and there needs to be an increase in studies directly comparing different transdermal alcohol sensors. There is also a lack of research investigating the accuracy of transdermal alcohol sensors as a tool for monitoring alcohol consumption in clinical populations and use over extended periods. Although there is some preliminary evidence suggesting the accuracy of these devices, this needs to be further investigated in clinical populations.

Trial Registration: PROSPERO CRD42021231027; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=231027

(J Med Internet Res 2022;24(4):e35178) doi:10.2196/35178

KEYWORDS

alcohol consumption; alcohol detection; alcohol monitoring; alcohol treatment; digital technology; ecologic momentary assessment; transdermal alcohol sensors; wearables; mobile phone

Introduction

Background

Current alcohol treatment largely relies on self-report data, which can be convenient and low-cost; however, there are limitations [1-3]. Reporting alcohol consumption over long periods may lead to recall bias [4]. Self-report can be made even more challenging when working with an alcohol-dependent population [5]. The current *gold standard* of alcohol assessments and alcohol research are self-report instruments, such as the Timeline Follow-Back [6] and the Alcohol Use Disorders Identification Test [7]. In recent years, various wearable transdermal alcohol sensor (TAS) devices have been developed. These devices measure alcohol consumption from alcohol vapors excreted through the skin via sweat, known as transdermal alcohol concentration (TAC), and can be worn on the wrist or ankle. Wearing a device all day for long periods allows for regular, repeated measurements and data capture in real time [8,9]. Thus, the TAS device is less likely to miss episodes of alcohol consumption than breath alcohol measurement. It is noninvasive, objective, and low-maintenance and allows behavior to be captured in real-world contexts [10]. These advantages address some of the limitations of other methods currently used, such as breathalyzers and blood and urine tests. Another advantage is that this technology has the potential to communicate with a smartphone [10-12]. Data captured can be uploaded over a mobile network and delivered to the patient, clinician, and researcher in near real time. This means reduced time and resources required by staff in addition to potentially more comprehensive and accurate data collected.

There is a range of TAS brands in various stages of development and validation, some available directly to the public. The 2 devices that have been most widely validated and reported on are SCRAM and WrisTAS [10,13-18]. More recently, Skyn from BACtrack, ION Wearable (rebranded as Proof) from Milo Sensors, and BARE and ORBIS from Smart Start have been developed. BACtrack uses a fuel cell alcohol sensor, whereas ION uses an enzymatic electrochemical biosensor cartridge [10,11]. BACtrack can be used as a stand-alone wearable or integrated into the band of a smartwatch [10,12] and, although it uses the same technology as WrisTAS, BACtrack uses a newer generation with improvements in performance and size [10]. At the beginning of 2021, Smart Start announced the development of ORBIS, a TAS with GPS monitoring technology designed to look like a smartwatch. Instead of fuel cell technology, this device uses a transdermal sensor that allows for continuous monitoring and has cellular communication embedded, removing the need to pair with a smartphone. This device is currently being tested through independent research and clinical and pilot trials [19]. These changes in TAS design and use—for example, linking to a smartphone app to display real-time data—address some of the limitations of the older devices. However, these new generations are in earlier stages of development, and researchers have only just begun exploring their validity, reliability, and usability [10-12].

The accuracy of these devices is evaluated by comparing TAC output with self-reported alcohol consumption via Timeline Follow-Back or drinking diaries, blood tests (blood alcohol concentration [BAC]), or breath alcohol concentration (BrAC). When comparing with self-report, the participant typically wears the device and notes how much alcohol they consume and when. This can allow for drinking in the participant's own environment and for typically greater, self-dosing drinking to be recorded. However, there can be limits to self-report data, including accuracy of recall, adherence to recording, and social acceptability reporting bias, particularly in clinical populations [1-5]. Sensitivity, also known as true positive rate, measures the number of positive drinking events as indicated by TAC that are true positive drinking events when compared with self-report, BAC, or BrAC. Specificity, also known as true negative rate, measures the number of events where alcohol consumption was not detected and there was no alcohol consumed when compared with self-report, BAC, or BrAC [20,21]. Most studies report correlation or sensitivity and specificity data of the TAS device.

The data collected and calculated from TAC (peaks of use, time to peak, and area under the curve [AUC]) can be compared with the measurements collected via breathalyzer or blood tests (BrAC or BAC). Although TAS devices can automatically take readings at predefined time points, owing to the need for frequent administration of breathalyzer readings or the need for blood tests, studies using these comparisons typically require laboratory settings. This means that there are typically fixed-dose amounts of alcohol given by a research team, and the data are taken during a limited period (a few hours). TAC, BrAC, and BAC data are then statistically analyzed to determine the correlation. It would be optimal for TAS devices to perform with high accuracy in both laboratory and natural, real-world drinking situations.

van Egmond et al [15] conducted a systematic review in 2020 exploring the validity of wearable TAS devices. However, their review only included papers published since 2013 that used TAS devices and validation measures obtained from the devices and provided correlation or detection rate measures. This meant that papers identified by van Egmond that included Milo ION, MOX sensors, and a study on BACtrack were not included in their review [15] because of lack of validation measures. For our review, we decided to include all papers that met our inclusion criteria for investigating the use of wearable TAS devices, with no time constraints. We aimed to explore a broad overview of TAS technology within the field and demonstrate the growth in the range and development of devices.

Objective

This systematic review aims to investigate the current knowledge by systematically identifying and evaluating the existing literature on the use of TAS devices in clinical and nonclinical populations, alone or in conjunction with a psychosocial intervention. The primary objective is to assess the level of accuracy of TAS devices. There is a linked review paper investigating the acceptability and feasibility of the devices (Brobbin et al, unpublished data, 2022).

Methods

Overview

This systematic review was conducted according to the PRISMA-P (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols) guidelines [22]. This protocol has been registered in PROSPERO (CRD42021231027). On review of the results, it was decided that the findings of the systematic review should be reported in 2 papers: one focusing on the accuracy outcomes and a second one focusing on the acceptability and feasibility outcomes.

Inclusion Criteria

Studies meeting all criteria were included: full-text original studies published in peer-reviewed journals, written in English, and using a wearable transdermal sensor device reporting accuracy outcomes. For the purpose of this review, a wearable TAS is defined as a wearable device that can measure alcohol consumption from alcohol vapors excreted via the skin. There were no restrictions on publication year or participant clinical and demographic characteristics. Data based on conference abstracts, dissertations, and gray literature were not included.

Information Sources

Bibliographic databases included CINAHL, Embase, Google Scholar, MEDLINE, PsycINFO, PubMed, and Scopus. Searches were carried out between February 1, 2021, and February 8, 2021 (Multimedia Appendix 1 [10,11,13,23-51] and Multimedia Appendix 2). The searches were supplemented by cross-checking the reference lists of key publications, related systematic reviews, and included papers.

All identified titles and abstracts were screened in Covidence (Veritas Innovation Ltd) to identify studies that potentially met the inclusion criteria. From this list, the full text was retrieved and assessed by a reviewer (EB); any doubts were discussed with a second reviewer (SH). Any disagreement was discussed with a third reviewer (PD). A data extraction form was created and pilot-tested with the first 5 included studies and refined as necessary (Multimedia Appendix 3). EB extracted the data independently, and the second reviewer (SH) completed the entry check for accuracy. Any discrepancies were resolved through discussion with the third reviewer (PD).

Outcomes

All outcome measures reported, both objective and self-reported, were extracted. The primary outcome was the accuracy of wearable TAS devices for measuring alcohol consumption compared with other methods (self-report, BAC, and BrAC; definitions of these measures are reported in Multimedia Appendix 4 [14,23,25-28,32,37,52]). This included data on correlations, sensitivity and specificity, percentage or amount of unsuccessful data points collected, and any time lag differences for the sensor to reach peak TAC compared with peak BrAC and delay in peak time from drink consumption.

Quality Assessment

We used the Mixed Methods Appraisal Tool (MMAT) as it is designed for appraisal in reviews that include a range of designs (qualitative, quantitative, and mixed methods) [53]. EB completed this independently, and any queries were discussed with a second reviewer (SH). Any disagreements were resolved through discussion (PD).

Data Synthesis and Analyses

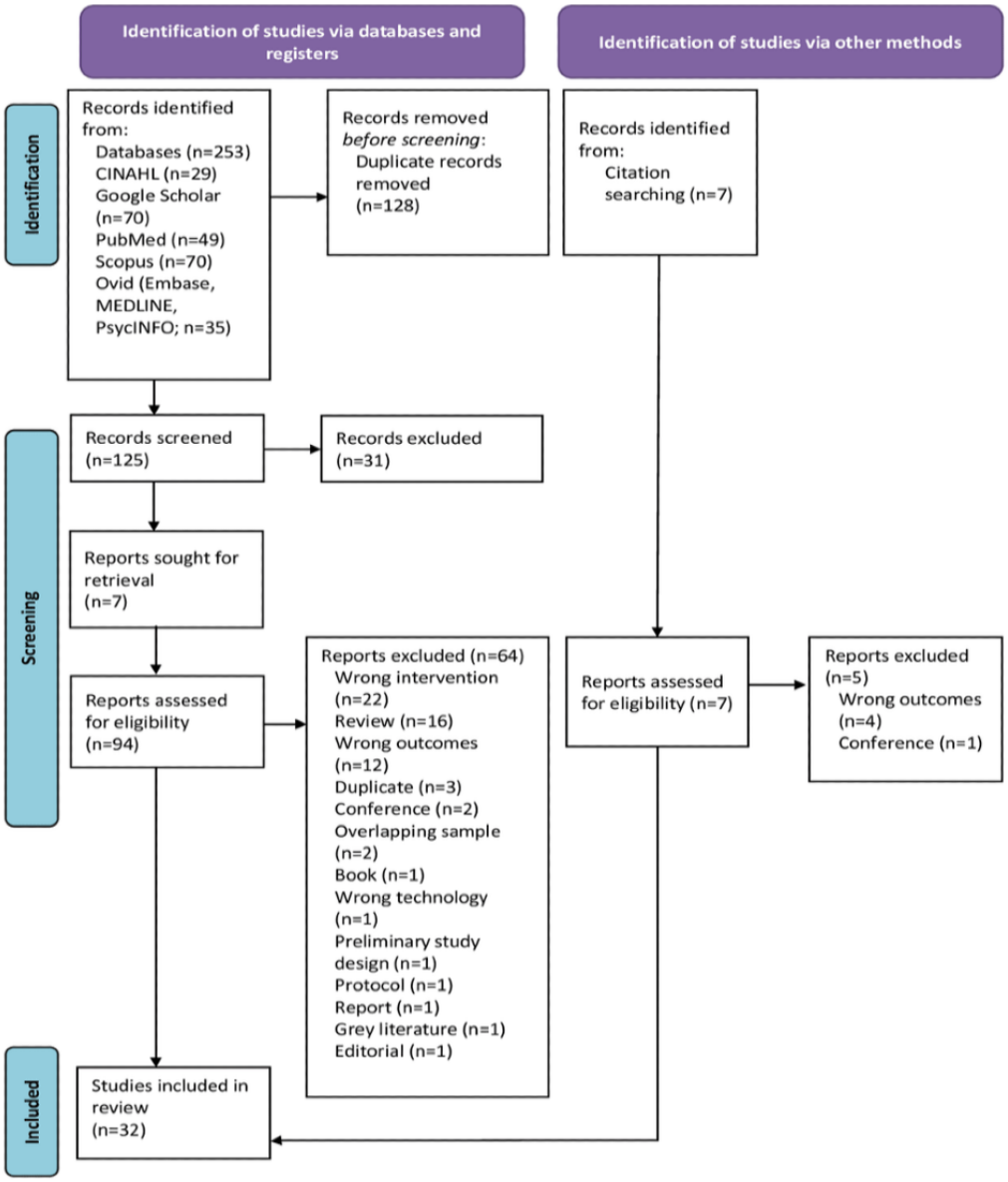
To draw conclusions about the included studies, we developed a synthesis of study characteristics. The data are summarized using a structured narrative description for accuracy measures, and we report the available data reported for correlations, sensitivity and specificity, failure rates, and time lag on SCRAM, WrisTAS, and BACtrack, with a separate section for other TAS model studies (with only 1 study on each of these). Not all studies reported accuracy data on all these measures. A meta-analysis was not possible because of the methodological heterogeneity.

Results

Overview

After removing duplicates, a total of 125 papers were screened; 31 (24.8%) were excluded at the title and abstract screening, and 94 (75.2%) full-text papers were assessed. Of those 94 papers, a total of 64 (68%) were then excluded (the reasons are provided in Multimedia Appendix 1). There were 7 additional papers identified through citation searching, of which 2 (29%) were included (Multimedia Appendix 1). The final sample included 32 publications (Figure 1).

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



Study Characteristics

Of the 32 studies, 19 (59%) used a SCRAM device [13,23-40], 7 (22%) used a WrisTAS device [34,41-46], 5 (16%) used BACtrack Skyn [10,29,30,40,47], 1 (3%) used an ION Milo sensor [11], 1 (3%) used Quantac Tally [10], 1 (3%) used a wearable Internet of Things (IoT) sensor [48], 1 (3%) used an MOX sensor [49], 1 (3%) used a proton-exchange membrane (PEM) fuel cell sensor [50], and 1 (3%) did not name the TAS used [51]. Some studies (4/32, 13%) used more than one version of the device (eg, SCRAMx and SCRAM-II), and 16% (5/32) used more than one brand of device (Table 1).

A large proportion of studies (29/32, 91%) were conducted with adults in good health. Only 9% (3/32) of the studies included participants who were diagnosed alcohol-dependent clinical populations. Most of this research was conducted in the United States, with 94% (30/32) of the studies located there. The earliest paper included was from 1992, but most studies (21/32, 66%) were published as of 2015 (Table 1).

There were 1228 participants enrolled in total in the included studies, with 1147 included in the procedure or analysis. Therefore, 81 participants who were enrolled were not included in the results (for reasons such as withdrawing or missing data). In total, 1 paper (1/32, 3%) was still in the early stages of data collection for one of their studies and so these participant numbers were unknown [10]. Not all studies included detailed information on the participants' age, sex, and ethnicity. From the information provided, it could be seen that the participants' ages ranged from 18 to 58 years, most studies included women and men (27/32, 84%; 2/32, 6% included men only; 1/32, 3% included women only; and 2/32, 6% were unknown) and, for most studies (19/32, 59%), White participants represented a high proportion of the sample (11/32, 34% were unknown; Table 1).

Table 1. Characteristics of the included studies (N=32).

Study and year	Design	Aim	Participants enrolled (N=1228) vs participants included (n=1147), n[a]	Population	Device	MMAT[b] score
Alessi et al, 2019 [23]	Ambulatory	Assess how we can measure alcohol consumption with this technology	66 (63)	Clinical: alcohol outpatient	SCRAMx	40%
Ayala et al, 2009 [24]	Laboratory	Assessing nonalcoholic energy drinks with TAM[c]	15 (15)	Nonclinical: good health	SCRAM-II	80%
Barnett et al, 2017 [25]	RCT[d], ambulatory	Effectiveness of TAM in implementing CM[e] for alcohol reduction treatment in various population groups and evaluating the efficacy of CM reduction in alcohol use	30 (30)	Nonclinical: heavy drinkers	SCRAM-II and SCRAMx	80%
Barnett et al, 2014 [26]	Ambulatory	Assess how we can measure alcohol consumption with this technology	66 (66)	Nonclinical: heavy drinkers	SCRAM-II and SCRAMx	100%
Barnett et al, 2011 [27]	Ambulatory	Effectiveness of TAM in implementing CM for alcohol reduction treatment in various population groups and evaluating the efficacy of CM reduction in alcohol use	20 (13)	Nonclinical: heavy drinkers	SCRAM	80%
Bond et al, 2014 [41]	Ambulatory	Assess how we can measure alcohol consumption with this technology	250 (250)	Nonclinical: good health	WrisTAS (5, 6, and 7)	60%
Croff et al, 2020 [42]	Ambulatory	Assess acceptability, adherence, and feasibility with this technology	59 (57)	Nonclinical: good health	WrisTAS-7	80%
Davidson et al, 1997 [51]	Laboratory	Assess how we can measure alcohol consumption with this technology	15 (12)	Nonclinical: social drinkers	Not named	80%
Dougherty et al, 2012 [28]	Laboratory	Assess how we can measure alcohol consumption with this technology	22 (21)	Nonclinical: good health	SCRAM-II	100%
Fairbairn and Kang, 2019 [29]	Laboratory	Assess how we can measure alcohol consumption with this technology	50 (30)	Nonclinical: social drinkers	BACtrack prototype and SCRAM	60%
Fairbairn et al, 2020 [30]	Laboratory	Assess how we can measure alcohol consumption with this technology	110 (73)	Nonclinical: good health	BACtrack prototype and SCRAM	60%

Study and year	Design	Aim	Participants enrolled (N=1228) vs participants included (n=1147), n[a]	Population	Device	MMAT[b] score
Fairbairn et al, 2019 [13]	Mixed design	Estimating BrAC[f] from TAC[g]	48 (48)	Nonclinical: social drinkers	SCRAM	100%
Hill-Kapturczak et al, 2014 [31]	Laboratory	Assess how we can measure alcohol consumption with this technology	22 (19)	Nonclinical: good health	SCRAM-II	80%
Jalal et al, 2020 [50]	Laboratory	Assess how we can measure alcohol consumption with this technology	8 (8)	Nonclinical: good health	PEM[h] fuel cell sensor	60%
Karns-Wright et al, 2018 [32]	Ambulatory	Assess how we can measure alcohol consumption with this technology	32 (30)	Nonclinical: good health	SCRAM-CAM	80%
Karns-Wright et al, 2017 [33]	Laboratory	Estimating BrAC from TAC	61 (61)	Nonclinical: good health	SCRAM	80%
Lansdorp et al, 2019 [11]	Ambulatory	Assess how we can measure alcohol consumption with this technology	1 (1)	Nonclinical: good health	Milo sensor	60%
Lawson et al, 2019 [49]	Laboratory	Assess how we can measure alcohol consumption with this technology	6 (6)	Nonclinical: good health	MOX sensor	60%
Li et al, 2020 [48]	Laboratory	Assess how we can measure alcohol consumption with this technology	2 (2)	Nonclinical: good health	Wearable IoT[i] alcohol sensor	80%
Luczak et al, 2015 [43]	Mixed design	Assess how we can measure alcohol consumption with this technology	32 (32)	Nonclinical: good health	WrisTAS-7	80%
Marques and McKnight, 2009 [34]	Ambulatory and laboratory	Assess how we can measure alcohol consumption with this technology	22 (22)	Nonclinical: good health	SCRAM and WrisTAS-5	60%
Norman et al, 2020 [35]	Ambulatory	Assess acceptability, adherence, and feasibility with this technology and how we can measure alcohol consumption with this technology	14 (14)	Nonclinical: good health	SCRAM	60%
Rash et al, 2019 [36]	Ambulatory	Assess how we can measure alcohol consumption with this technology	22 (19)	Nonclinical: heavy drinking	SCRAMx	100%
Roache et al, 2015 [37]	Laboratory	Assess how we can measure alcohol consumption with this technology	61 (61)	Nonclinical: good health	SCRAM-II (study 1) and SCRAMx (studies 2 and 3)	80%
Roache et al, 2019 [38]	Ambulatory	Assess how we can measure alcohol consumption with this technology	30 (30)	Nonclinical: good health	SCRAM-CAM	80%

Study and year	Design	Aim	Participants enrolled (N=1228) vs participants included (n=1147), n[a]	Population	Device	MMAT[b] score
Rosenberg et al, 2021 [47]	Ambulatory	Assess acceptability, adherence, and feasibility with this technology	5 (5)	Nonclinical: good health	BACtrack	80%
Sakai et al, 2006 [39]	Ambulatory and laboratory	Assess how we can measure alcohol consumption with this technology	44 (44)	Alcohol-dependent and non–alcohol-dependent	SCRAM	100%
Simons et al, 2015 [44]	Ambulatory	Assess how we can measure alcohol consumption with this technology	60 (60)	Nonclinical: good health	WrisTAS-7	80%
Swift et al, 1992 [45]	Laboratory	Assess how we can measure alcohol consumption with this technology	15 (15)	Nonclinical: good health and alcohol-dependent	WrisTAS	80%
Wang et al, 2019 [10]	Ambulatory and laboratory	Assess how we can measure alcohol consumption with this technology	Still recruiting	Nonclinical: good health	Quantac Tally and BACtrack	20%
Wang et al, 2021 [40]	Ambulatory and laboratory	Assess how we can measure alcohol consumption with this technology	25 (25)	Nonclinical: good health	BACtrack and SCRAM-CAM	80%
Webster and Gabler, 2008 [46]	Laboratory	Assess how we can measure alcohol consumption with this technology	15 (15)	Nonclinical: good health	WrisTAS	80%

[a]The numbers in parentheses in this column are the number of participants that were included in each study after drop outs/withdrawals.

[b]MMAT: Mixed Methods Appraisal Tool.

[c]TAM: transdermal alcohol monitoring.

[d]RCT: randomized controlled trial.

[e]CM: contingency management.

[f]BrAC: breath alcohol concentration.

[g]TAC: transdermal alcohol concentration.

[h]PEM: proton-exchange membrane.

[i]IoT: Internet of Things.

Quality Assessment

All studies apart from 2 (30/32, 94%) met a minimum of 3 out of 5 criteria [10,23] (Table 1). This is due to Alessi et al [23] not providing details about randomization and participant information, and the study by Wang et al [10] was difficult to score owing to incomplete data collection as their study was still ongoing at the time of publication. With the MMAT, exclusion of low–methodological-quality studies is discouraged [53]. The MMAT scores for each study are provided in Multimedia Appendix 5 [10,11,13,23-51]. All the records selected for data extraction were considered to be at low risk of bias. Owing to the nature of many of the studies included, blinding of participants and personnel was not possible; in some studies (3/32, 9%), there were clear differences in demographics or amount of alcohol provided between groups [30,41,44] or incentives provided within contingency management studies [25] where personnel were required to know participant allocation. Similarly, in many studies (17/32, 53%), there was only 1 group of participants all completing the same task, so randomization or blinding was not required or possible [10,13,24,28,31-34,40,42,43,45,46,48-51], or there was only 1 participant [11]. Some studies (16/32, 50%) did not provide clear information on participant data, randomization, incomplete outcome data, and selective reporting, resulting in potential bias because of limited information in the paper [10,11,24,25,29,30,34,40,41,43-46,48-50].

Accuracy Measures

Overview

Of the 32 studies, 19 (59%) explored the accuracy of SCRAM in laboratory or ambulatory settings [13,23-40], 7 (22%) explored the accuracy of WrisTAS in laboratory or ambulatory

settings [34,41-46], and 5 (16%) explored the accuracy of a BACtrack device or prototype in laboratory or ambulatory settings [10,29,30,40,47]. Finally, there were six (6/32, 19%) other TAS devices within the included papers: a PEM fuel cell–based wearable alcohol sensor [50], Milo [11], MOX [49], Quantac Tally [10], a wearable IoT sensor [48], and 1 study (1/32, 3%) that did not name the device but simply described it as a TAS throughout [51].

Correlations

Overview

Approximately 6% (2/32) of the studies reported discrepancies between SCRAM TAC and self-report [23,39], whereas another study (4/32, 3%) reported moderate to high correlations (r=0.79 and P<.001 [25]; r=0.79-0.94 and P<.01 [27]; r=0.68 and P<.001 [36]). Karns-Wright et al [32] found that, when the concordance with any self-reported drinking was examined as a whole, the concordance rate was significantly higher for moderate and heavy TAC (mean 75.91%, SD 15.06%) than for only heavy TAC (mean 73.69%, SD 15.23%; t_{29}=2.05; P=.04). Most studies that reported on the correlation between TAC and BAC or BrAC data found moderate to strong correlations (weighted correlations between TAC and estimated BAC (eBAC) [26]: r=0.54 and P<.001; Pearson correlation coefficient range for peak TAC and BrAC [28]: 0.700-0.997; peak TAC and peak estimated BrAC: r=0.56 and P=.001; TAC and BrAC [29]: r=0.60 and P<.001; correlation between peak TAC and peak BrAC [31]: $F_{1,73}$=160.03 and P<.001). However, Sakai et al [39] found varying disagreements between peak TAC and peak BrAC, and Bland-Altman analyses for laboratory and community participants showed varying disagreement between peak and AUC BrAC and TAC data.

WrisTAS

Bond et al [41] found that the correlation between the AUC and the unadjusted number of drinks was 0.62, whereas the correlation between the AUC and the adjusted number of drinks was 0.73 (P=.04 for the difference between correlation coefficient estimates). A low correlation was reported between WrisTAS TAC and BAC (0.20-0.24; correlation between TAC and eBAC in adolescents only: 0.083-0.10; correlation between TAC and eBAC in young adults only: 0.37-0.39; significantly different by age group: P<.001) [42]. However, Swift et al [45] found that the TAS signal was similar in amplitude and time course to the BAC curve (curve peak amplitude: r=0.61 and P<.02; AUCs of BAC and TAS: r=0.91 and P<.001) [45], whereas another study (1/32, 3%) found that TAC overestimated BAC levels in the self-dose situation by 86%. By contrast, in the laboratory situation, WrisTAS logged TAC peaks just 0.019 g/dL lower than the mean peak BAC [34].

BACtrack

An overall high correlation between self-report and TAC was found (correlation for drinking start time: r=0.90 and P<.001; AUC for drinking event: r=0.7 and P=.008) [47]. BACtrack was able to distinguish low and high alcohol doses. There was a high correlation between TAC and BrAC data (r=0.77; P<.001) [29]. For the study by Wang et al [40], data collection was ongoing at the time of publication, but initial data suggested consistency between TAC and BrAC. Fairbairn et al [30] used SCRAM and BACtrack and found that models for estimating real-time BrAC measurements built with data from the Skyn sensor outperformed similar models built with data from SCRAM. Differences between estimated BrAC and BrAC were 60% higher for models based on data from SCRAM versus BACtrack.

Sensitivity and Specificity

SCRAM

Studies found that the ability of SCRAM to predict or distinguish levels of alcohol consumption had a significant positive relationship with the amount consumed [23,31,33,37,39]. Dougherty et al [28] plotted receiver operating characteristic curves using the peak TAC to predict the number of drinks consumed. A peak TAC value cutoff of ≥0.011 g/dL classified participants as having drank at least one beer with 97.9% accuracy (AUC=0.99, sensitivity 98.6%, and specificity 95%). A peak TAC value cutoff of ≥0.024 g/dL classified participants as having drank 1-2 beers or >2 beers with 85.1% accuracy (AUC=0.93, sensitivity 92.3%, and specificity 76.2%).

WrisTAS

Bond et al [41] found that the sensitivity for WrisTAS compared with self-report was 85.6% and the specificity compared with self-report was 67.5% (percentage of days during which diaries and devices indicated no drinking event). Croff et al [42] found that the sensitivity was 40% and the specificity was 87.9%. Marques and McKnight [34] found that the sensitivity was 24%. This low rate was explained because of erratic output and unreliable data recording during 67% of episodes recorded. WrisTAS more accurately estimated laboratory dosing levels than self-dosing. Simons et al [44] found that WrisTAS correctly identified 85.74% of self-reported drinking (sensitivity 72.35% and specificity 92.94%).

Failure Rates

The failure rate for SCRAM was reported to be low at 2% [29] and, in the study by Fairbairn et al [30], 6 files were found to be missing because of procedural issues associated with SCRAM assignment, and 1 was missing because of device malfunction.

Croff et al [42] found that, of the 471,625 data points collected via the WrisTAS, 35,803 data points were missing or corrupted, which equated to 186 whole or partial days of missing data. This was more common for adolescent (11.78% of daily data collected) than for young adult participants (8.59%; χ^2=−18.4; P<.001). WrisTAS data files were often found to have spikes despite the participants not reporting a drinking event. For some of these, it was most likely an environmental alcohol (perfume or hand sanitizer). However, other spikes may have been due to unreported alcohol consumption based on the data interpretation [43].

Marques and McKnight [34] found a low sensitivity, the reason being erratic output and not recording or missing data during drinking episodes (this was reported as a chipset rather than a sensor problem). It was also reported that it was harder to interpret the data because of *noisy* patterns. However, Swift et al [45] reported occasional noise, usually because of defects in

the interface cable, but transient noise was easily distinguished from a drinking event.

The BACtrack prototype was found to have a higher failure rate between 16% and 38% [29,30]. Missing data were due to device malfunction and device and user issues (described as being lost by the researchers as they learned how to use the prototypes). It was found that BACtrack Skyn data showed a lot of *noise* thought to be due to the rapid increase or decrease of the TAC signal [40].

Time Lag

The peak TAC measured by SCRAM was 120 minutes after the peak BrAC [29]. Across all measured portions of the BrAC curve, SCRAM lagged by 69 minutes (P<.001). Approximately 9% (3/32) of the studies found a delay in SCRAM data behind alcohol consumption and BrAC data of approximately 2-3 hours [13,35,39] or even longer and mean TAC peak delays of 4.5 (SD 2.9) hours relative to BAC peaks [34].

The mean time to peak for WrisTAS was 71 (SE 7) minutes, and the mean time for the BAC curve was 107 (SE 12) minutes; for the TAS curve, the mean difference in onset times was significantly different (P<.02). In the controlled consumption experiment, the peak values of the TAS concentration-time curves lagged by approximately 30 minutes behind the breathalyzer time curves. In intoxicated participants, the peak TAS signal lagged up to 120 minutes behind the BAC.

BACtrack measurements lagged the peak BrAC by 24 minutes (P<.001) [29]. Rosenberg et al [47] found a time lag of 2 and 3 hours for peak TAC compared with peak BrAC in low- and high-dose groups, respectively. Initial results found a time lag for Skyn of approximately 135 minutes after drinking for peak TAC compared with 60 minutes for peak BrAC [10].

Other TAS Devices

A PEM fuel cell–based wearable alcohol-sensing device was used in a human volunteer pilot study [50]. The measurements from the device showed a significant correlation with the calculated theoretical values. The device provided continuous BAC data, which were processed and fitted into a principal component regression model to determine the accurate transcutaneous alcohol content. Breathalyzer measurements showed greater variation than sensor data.

Lansdorp et al [11] used the Milo sensor to measure data continuously over 2 days. After a state of baseline data was recorded, a solution of 0.05 mol/L ethanol in 1x phosphate-buffered saline was flowed over the diffusion-limiting membrane. The mean sensor response time (time to reach the current 50% of the maximal plateau after the addition of a known concentration of ethanol) under laboratory conditions was 36 (SD 6) minutes with 12 sensors. A linear sensor range between 0 and 0.05 mol/L of ethanol was found.

Investigation of a MOX sensor found that the TAC curve was right-shifted from the BAC and BrAC curves and there was a time delay for the peak of approximately 80 minutes. The 2 different concentrations of ethanol (0.5 g/L and 0.8 g/L) absorbed by the participants could be discriminated [49].

The Quantac Tally was explored before Quantac Co. ceased its business operations. TAC measurements peaked on average 115 minutes after drinking onset, with a gradual increase to peak concentration [40].

The use of a wearable IoT sensor [48] found that the mean values could be distinguished between different alcohol doses. They also compared wearing the sensor on different body parts (left upper arm and left ankle); these results found that the different body parts found different TAC gas (TACg) data. To further explore how perspiration affects the data, the same participant wore the device as normal versus wearing 3 jumpers or jackets to induce sweating. The comparison between BAC and TACg for AUC ratio was 0.98 for BAC and 0.52 for TACg. Spikes in TACg data from the jumper experiment were correlated with spikes in humidity [48].

Davidson et al [51] explored whether a TAS could accurately measure low BAC. The results showed that the device was able to detect TAC at the lowest dose (10 mg/dL), which was not measured by BrAC or BAC. For doses 2 (20 mg/dL) and 3 (40 mg/dL), TAC, BrAC, and BAC were in general agreement. However, BrAC and BAC were in stronger agreement (dose 2: r=0.092 and P<.001; dose 3: r=0.88 and P<.001) than TAC compared with BrAC (dose 2: r=0.52 and P>.05; dose 3: r=0.7 and P>.05) and BAC (dose 2: r=0.60 and P<.05; dose 3: r=0.52 and P>.05). There was also a time lag for TAC output, which was affected by *noise*.

Discussion

Principal Findings

The aim of this review was to assess the current knowledge on the accuracy of TAS devices. We identified 32 papers, few of which reported the use of this technology in clinical populations. Of the 32 papers, we identified only 3 (9%) studies that used alcohol-dependent participants. Most studies used either SCRAM (19/32, 59%) or WrisTAS (7/32, 22%). However, there were some studies that investigated BACtrack Skyn (5/32, 16%) and others (6/32, 19%). This is a growing field, and the production and investigation of additional TAS devices supports this. This review found that wearable TAS devices could detect alcohol consumption with moderate to strong accuracy over various periods. However, factors such as the amount of alcohol consumed, the environment (laboratory and self-dose real-world setting), age, and where the device is worn must be considered. The findings differed across the TAS brands included, and studies on each brand reported different limitations; for example, time lag compared with BrAC or BAC, data file errors, or device failure.

The SCRAM, WrisTAS, and BACtrack studies showed reasonably high data capture rates but demonstrate that this is <100% [26,27,32,34,41,42,44]. There were also reports of device failure, malfunctions, noise, and tampering, all of which reduced the amount of successful data capture [29,30,43]. It was noted over time that SCRAM lost accuracy, most likely because of water accumulation [34]. The main feature of using these devices compared with other tests, such as breathalyzers or urine tests, is the capacity for continuous measurements, so

this is an important benefit of TAS devices. There has been some investigation into factors such as sex and BMI on episode detection, with no conclusive relationship determined [26,28].

SCRAM devices had the lowest amount of recorded failure rates of the included devices [29,30,34,42], but the time lag to peak alcohol concentration was considerably longer than that of the other brands of TAS devices [10,13,29,34,40,45,47]. Although SCRAM, WrisTAS, and BACtrack showed positive correlations with self-report, BAC, and BrAC, the accuracy of determining the amount of alcohol appeared to be stronger in BACtrack than in SCRAM in some studies [29] but not in others [40]. WrisTAS showed the weakest support for accuracy [34,42]. There were fewer studies on BACtrack compared with SCRAM and WrisTAS, and the studies included that used BACtrack mostly used a device prototype. However, the preliminary results appear promising. Similar to WrisTAS, there seems to be a high failure rate and noise in the data files, but there is support for a shorter time lag compared with SCRAM as well as a strong positive correlation with BAC or BrAC, and it maintains the ability to distinguish between high and low doses [29,40,47].

Approximately 3% (1/32) of the studies compared SCRAM and WrisTAS in laboratory and ambulatory settings. When the participants self-dosed, the true positive rates for both devices increased and, the higher the peak BAC level, the higher the rate of true positive TAC [34]. Although SCRAM more accurately estimated self-dosing and WrisTAS more accurately estimated laboratory dosing, neither was accurately estimated in both settings. This difference between devices is something that should be investigated further and could affect preference for brand of device for different uses in different settings [34].

Our findings relating to SCRAM, WrisTAS, and BACtrack were largely in line with the review findings of van Egmond [15]. Data from these TAS devices were found to moderately to highly positively correlate with self-report and BrAC, with WrisTAS and BACtrack devices showing higher malfunction, failure rates, and noise within data files compared with SCRAM. Van Egmond [15] did not include papers exploring other TAS devices.

There was 1 individual paper (1/32, 3%) on each of the other 6 wearable TAS devices [10,11,48-51]. The development of other devices suggests that this is a rapidly evolving technology. The unnamed device in the study by Davidson et al [51] demonstrates the ability of this technology to accurately measure low BAC and the general agreement between TAC, BAC, and BrAC data. Li et al [48] is the only study that compares data taken from the same device but worn on different body parts. The data found suggest that the location of the device on the user does affect the measurements captured, which is important to consider.

Limitations

This review highlights the small but growing number of studies investigating the accuracy of wearable TAS devices. With only a small number of studies testing more than one brand or even the same setting, intervention, or population, it is difficult to draw direct comparisons. Most studies (29/32, 91%) included healthy adults, and the length of time studied ranged from a day to a few months. If the ultimate objective is for these devices to be used within the clinical population or the criminal justice system for extended periods, this manner of use needs to be explored further. No included study was conducted in a criminal justice setting with offenders. However, there is a growing appeal for this technology to be used in this environment; for example, the use of SCRAM in the UK criminal justice system for alcohol-related offenses [54].

A mix of study designs was included in both the ambulatory and laboratory settings. Although it is useful to conduct studies in controlled laboratory environments to investigate validity, reliability, and accuracy, the target use is not necessarily a laboratory environment. Hence, how the devices perform in vivo with populations other than those studied must be investigated. This would then be able to inform the devices' long-term effectiveness in future treatment, interventions, research, and policy.

Conclusions

There are currently a small number of studies in this area, but research on the use of this technology is growing and, owing to technological advancements, the accuracy and ability of these devices is improving considerably. What is needed is for research to expand into other populations, such as clinical populations and offenders within the criminal justice system, to examine their accuracy and reliability in the intended target populations and contexts. Although the accuracy outcomes for this technology are promising, there is a limit to this research because of the mostly laboratory and short-duration study design.

The use of wearable TAS devices is becoming more accepted and appealing to society, as evidenced by their increasing implementation in the criminal justice system and increasing research. The implications of this review are that we need to investigate the engagement in real-world settings, where transdermal sensors are intended to be implemented, with the target populations. This can then inform clinical interventions, treatment, research, and use.

Acknowledgments

This is a summary of independent research funded by the National Institute for Health Research (NIHR) Applied Research Collaboration South London. The views expressed are those of the authors and not necessarily those of the National Health Service (NHS), the NIHR, or the Department of Health and Social Care. CD and PD were supported by the NIHR Specialist Biomedical Research Centre for Mental Health at South London and Maudsley NHS Foundation Trust and King's College London. They were also supported by the NIHR Collaboration for Leadership in Applied Health Research and Care at King's

College Hospital NHS Foundation Trust and the NIHR Applied Research Collaboration South London at King's College Hospital NHS Foundation Trust. CD is supported by an NIHR Senior Investigator Award.

Authors' Contributions

EB conceived the work, designed the protocol, completed the systematic review and data extraction, and drafted and revised the manuscript. PD conceived the work, edited the protocol, was the third reviewer for the data extraction, and revised the manuscript. SH was the second reviewer for the data extraction and revised the manuscript. CD conceived the work, edited the protocol, and revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Database search, screening, and inclusion of studies for transdermal alcohol sensor accuracy systematic review.
[DOCX File , 35 KB - jmir_v24i4e35178_app1.docx]

Multimedia Appendix 2
Database search terms.
[DOCX File , 17 KB - jmir_v24i4e35178_app2.docx]

Multimedia Appendix 3
Data extraction form.
[DOCX File , 14 KB - jmir_v24i4e35178_app3.docx]

Multimedia Appendix 4
Transdermal alcohol concentration and blood alcohol concentration measurement definition.
[DOCX File , 21 KB - jmir_v24i4e35178_app4.docx]

Multimedia Appendix 5
Risk of bias Mixed Methods Appraisal Tool scores for transdermal alcohol sensor accuracy systematic review.
[DOCX File , 24 KB - jmir_v24i4e35178_app5.docx]

References

1. McPhail S, Haines T. Response shift, recall bias and their effect on measuring change in health-related quality of life amongst older hospital patients. Health Qual Life Outcomes 2010 Jul 10;8:65 [FREE Full text] [doi: 10.1186/1477-7525-8-65] [Medline: 20618978]
2. Schmier JK, Halpern MT. Patient recall and recall bias of health state and health status. Expert Rev Pharmacoecon Outcomes Res 2004 Apr;4(2):159-163. [doi: 10.1586/14737167.4.2.159] [Medline: 19807519]
3. Rosenman R, Tennekoon V, Hill LG. Measuring bias in self-reported data. Int J Behav Healthc Res 2011 Oct;2(4):320-332 [FREE Full text] [doi: 10.1504/IJBHR.2011.043414] [Medline: 25383095]
4. Piasecki TM. Assessment of alcohol use in the natural environment. Alcohol Clin Exp Res 2019 Apr;43(4):564-577 [FREE Full text] [doi: 10.1111/acer.13975] [Medline: 30748019]
5. Del Boca FK, Darkes J. The validity of self-reports of alcohol consumption: state of the science and challenges for research. Addiction 2003 Dec;98 Suppl 2:1-12. [doi: 10.1046/j.1359-6357.2003.00586.x] [Medline: 14984237]
6. Sobell LC, Sobell MB. Timeline follow-back: a technique for assessing self-reported alcohol consumption. In: Litten RZ, Allen JP, editors. Measuring alcohol consumption: psychosocial and biological methods. Totowa, NJ, USA: Humana Press; 1992:41-72.
7. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on early detection of persons with harmful alcohol consumption--II. Addiction 1993 Jun;88(6):791-804. [doi: 10.1111/j.1360-0443.1993.tb02093.x] [Medline: 8329970]
8. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
9. Bertz JW, Epstein DH, Preston KL. Combining ecological momentary assessment with objective, ambulatory measures of behavior and physiology in substance-use research. Addict Behav 2018 Aug;83:5-17 [FREE Full text] [doi: 10.1016/j.addbeh.2017.11.027] [Medline: 29174666]
10. Wang Y, Fridberg DJ, Leeman RF, Cook RL, Porges EC. Wrist-worn alcohol biosensors: strengths, limitations, and future directions. Alcohol 2019 Dec;81:83-92 [FREE Full text] [doi: 10.1016/j.alcohol.2018.08.013] [Medline: 30179709]

11. Lansdorp B, Ramsay W, Hamidand R, Strenk E. Wearable enzymatic alcohol biosensor. Sensors (Basel) 2019 May 24;19(10):2380 [FREE Full text] [doi: 10.3390/s19102380] [Medline: 31137611]
12. Campbell AS, Kim J, Wang J. Wearable electrochemical alcohol biosensors. Curr Opin Electrochem 2018 Aug;10:126-135 [FREE Full text] [doi: 10.1016/j.coelec.2018.05.014] [Medline: 30859141]
13. Fairbairn CE, Rosen IG, Luczak SE, Venerable WJ. Estimating the quantity and time course of alcohol consumption from transdermal alcohol sensor data: a combined laboratory-ambulatory study. Alcohol 2019 Dec;81:111-116 [FREE Full text] [doi: 10.1016/j.alcohol.2018.08.015] [Medline: 30179707]
14. Dougherty DM, Karns TE, Mullen J, Liang Y, Lake SL, Roache JD, et al. Transdermal alcohol concentration data collected during a contingency management program to reduce at-risk drinking. Drug Alcohol Depend 2015 Mar 01;148:77-84 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.12.021] [Medline: 25582388]
15. van Egmond K, Wright CJ, Livingston M, Kuntsche E. Wearable transdermal alcohol monitors: a systematic review of detection validity, and relationship between transdermal and breath alcohol concentration and influencing factors. Alcohol Clin Exp Res 2020 Oct;44(10):1918-1932. [doi: 10.1111/acer.14432] [Medline: 32767791]
16. Greenfield TK, Bond J, Kerr WC. Biomonitoring for improving alcohol consumption surveys: the new gold standard? Alcohol Res 2014;36(1):39-45 [FREE Full text] [Medline: 26258999]
17. Leffingwell TR, Cooney NJ, Murphy JG, Luczak S, Rosen G, Dougherty DM, et al. Continuous objective monitoring of alcohol use: twenty-first century measurement using transdermal sensors. Alcohol Clin Exp Res 2013 Jan;37(1):16-22 [FREE Full text] [doi: 10.1111/j.1530-0277.2012.01869.x] [Medline: 22823467]
18. Barnett NP. Alcohol sensors and their potential for improving clinical care. Addiction 2015 Jan;110(1):1-3. [doi: 10.1111/add.12764] [Medline: 25515824]
19. Ogle R. Smart Start revolutionizes transdermal alcohol and electronic monitoring with a single solution. Smart Start. 2021. URL: https://ww1.prweb.com/prfiles/2021/01/27/17691004/27JAN2021_Transdermal_PR_with_photo.pdf [accessed 2021-06-18]
20. Mandrekar JN. Simple statistical measures for diagnostic accuracy assessment. J Thorac Oncol 2010 Jun;5(6):763-764 [FREE Full text] [doi: 10.1097/JTO.0b013e3181dab122] [Medline: 20502268]
21. Altman DG, Bland JM. Diagnostic tests. 1: sensitivity and specificity. BMJ 1994 Jun 11;308(6943):1552 [FREE Full text] [doi: 10.1136/bmj.308.6943.1552] [Medline: 8019315]
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. Alessi SM, Barnett NP, Petry NM. Objective continuous monitoring of alcohol consumption for three months among alcohol use disorder treatment outpatients. Alcohol 2019 Dec;81:131-138 [FREE Full text] [doi: 10.1016/j.alcohol.2019.01.008] [Medline: 30710610]
24. Ayala J, Simons K, Kerrigan S. Quantitative determination of caffeine and alcohol in energy drinks and the potential to produce positive transdermal alcohol concentrations in human subjects. J Anal Toxicol 2009;33(1):27-33. [doi: 10.1093/jat/33.1.27] [Medline: 19161666]
25. Barnett NP, Celio MA, Tidey JW, Murphy JG, Colby SM, Swift RM. A preliminary randomized controlled trial of contingency management for alcohol use reduction using a transdermal alcohol sensor. Addiction 2017 Jun;112(6):1025-1035 [FREE Full text] [doi: 10.1111/add.13767] [Medline: 28107772]
26. Barnett NP, Meade EB, Glynn TR. Predictors of detection of alcohol use episodes using a transdermal alcohol sensor. Exp Clin Psychopharmacol 2014 Feb;22(1):86-96 [FREE Full text] [doi: 10.1037/a0034821] [Medline: 24490713]
27. Barnett NP, Tidey J, Murphy JG, Swift R, Colby SM. Contingency management for alcohol use reduction: a pilot study using a transdermal alcohol sensor. Drug Alcohol Depend 2011 Nov 01;118(2-3):391-399 [FREE Full text] [doi: 10.1016/j.drugalcdep.2011.04.023] [Medline: 21665385]
28. Dougherty DM, Charles NE, Acheson A, John S, Furr RM, Hill-Kapturczak N. Comparing the detection of transdermal and breath alcohol concentrations during periods of alcohol consumption ranging from moderate drinking to binge drinking. Exp Clin Psychopharmacol 2012 Oct;20(5):373-381 [FREE Full text] [doi: 10.1037/a0029021] [Medline: 22708608]
29. Fairbairn CE, Kang D. Temporal dynamics of transdermal alcohol concentration measured via new-generation wrist-worn biosensor. Alcohol Clin Exp Res 2019 Oct;43(10):2060-2069 [FREE Full text] [doi: 10.1111/acer.14172] [Medline: 31469451]
30. Fairbairn CE, Kang D, Bosch N. Using machine learning for real-time BAC estimation from a new-generation transdermal biosensor in the laboratory. Drug Alcohol Depend 2020 Nov 01;216:108205 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.108205] [Medline: 32853998]
31. Hill-Kapturczak N, Lake SL, Roache JD, Cates SE, Liang Y, Dougherty DM. Do variable rates of alcohol drinking alter the ability to use transdermal alcohol monitors to estimate peak breath alcohol and total number of drinks? Alcohol Clin Exp Res 2014 Oct;38(10):2517-2522 [FREE Full text] [doi: 10.1111/acer.12528] [Medline: 25335857]
32. Karns-Wright TE, Dougherty DM, Hill-Kapturczak N, Mathias CW, Roache JD. The correspondence between transdermal alcohol monitoring and daily self-reported alcohol consumption. Addict Behav 2018 Oct;85:147-152 [FREE Full text] [doi: 10.1016/j.addbeh.2018.06.006] [Medline: 29910035]

33. Karns-Wright TE, Roache JD, Hill-Kapturczak N, Liang Y, Mullen J, Dougherty DM. Time delays in transdermal alcohol concentrations relative to breath alcohol concentrations. Alcohol Alcohol 2017 Jan;52(1):35-41 [FREE Full text] [doi: 10.1093/alcalc/agw058] [Medline: 27522029]
34. Marques PR, McKnight AS. Field and laboratory alcohol detection with 2 types of transdermal devices. Alcohol Clin Exp Res 2009 Apr;33(4):703-711. [doi: 10.1111/j.1530-0277.2008.00887.x] [Medline: 19170663]
35. Norman T, Peacock A, Ferguson SG, Kuntsche E, Bruno R. Combining transdermal and breath alcohol assessments, real-time drink logs and retrospective self-reports to measure alcohol consumption and intoxication across a multi-day music festival. Drug Alcohol Rev 2021 Nov;40(7):1112-1121. [doi: 10.1111/dar.13215] [Medline: 33174260]
36. Rash CJ, Petry NM, Alessi SM, Barnett NP. Monitoring alcohol use in heavy drinking soup kitchen attendees. Alcohol 2019 Dec;81:139-147 [FREE Full text] [doi: 10.1016/j.alcohol.2018.10.001] [Medline: 30308286]
37. Roache JD, Karns TE, Hill-Kapturczak N, Mullen J, Liang Y, Lamb RJ, et al. Using transdermal alcohol monitoring to detect low-level drinking. Alcohol Clin Exp Res 2015 Jul;39(7):1120-1127 [FREE Full text] [doi: 10.1111/acer.12750] [Medline: 25988708]
38. Roache JD, Karns-Wright TE, Goros M, Hill-Kapturczak N, Mathias CW, Dougherty DM. Processing transdermal alcohol concentration (TAC) data to detect low-level drinking. Alcohol 2019 Dec;81:101-110 [FREE Full text] [doi: 10.1016/j.alcohol.2018.08.014] [Medline: 30179708]
39. Sakai JT, Mikulich-Gilbertson SK, Long RJ, Crowley TJ. Validity of transdermal alcohol monitoring: fixed and self-regulated dosing. Alcohol Clin Exp Res 2006 Jan;30(1):26-33. [doi: 10.1111/j.1530-0277.2006.00004.x] [Medline: 16433729]
40. Wang Y, Fridberg DJ, Shortell DD, Leeman RF, Barnett NP, Cook RL, et al. Wrist-worn alcohol biosensors: applications and usability in behavioral research. Alcohol 2021 May;92:25-34. [doi: 10.1016/j.alcohol.2021.01.007] [Medline: 33609635]
41. Bond JC, Greenfield TK, Patterson D, Kerr WC. Adjustments for drink size and ethanol content: new results from a self-report diary and transdermal sensor validation study. Alcohol Clin Exp Res 2014 Dec;38(12):3060-3067 [FREE Full text] [doi: 10.1111/acer.12589] [Medline: 25581661]
42. Croff JM, Hartwell ML, Chiaf AL, Crockett EK, Washburn IJ. Feasibility and reliability of continuously monitoring alcohol use among female adolescents and young adults. Drug Alcohol Rev 2021 Nov;40(7):1143-1154. [doi: 10.1111/dar.13045] [Medline: 32185847]
43. Luczak SE, Rosen IG, Wall TL. Development of a real-time repeated-measures assessment protocol to capture change over the course of a drinking episode. Alcohol Alcohol 2015 Mar;50(2):180-187 [FREE Full text] [doi: 10.1093/alcalc/agu100] [Medline: 25568142]
44. Simons JS, Wills TA, Emery NN, Marks RM. Quantifying alcohol consumption: self-report, transdermal assessment, and prediction of dependence symptoms. Addict Behav 2015 Nov;50:205-212 [FREE Full text] [doi: 10.1016/j.addbeh.2015.06.042] [Medline: 26160523]
45. Swift RM, Martin CS, Swette L, LaConti A, Kackley N. Studies on a wearable, electronic, transdermal alcohol sensor. Alcohol Clin Exp Res 1992 Aug;16(4):721-725. [doi: 10.1111/j.1530-0277.1992.tb00668.x] [Medline: 1530135]
46. Webster GD, Gabler HC. Modeling of transdermal transport of alcohol effect of body mass and gender. Biomed Sci Instrum 2008;44:361-366. [Medline: 19141942]
47. Rosenberg M, Ludema C, Kianersi S, Luetke M, Jozkowski K, Guerra-Reyes L, et al. Wearable alcohol monitors for alcohol use data collection among college students: feasibility and acceptability in a pilot study. medRxiv (forthcoming) 2021. [doi: 10.1101/2021.02.17.21251959]
48. Li B, Downen RS, Dong Q, Tran N, LeSaux M, Meltzer AC, et al. A discreet wearable IoT sensor for continuous transdermal alcohol monitoring - challenges and opportunities. IEEE Sens J 2021 Feb 15;21(4):5322-5330 [FREE Full text] [doi: 10.1109/jsen.2020.3030254] [Medline: 33746626]
49. Lawson B, Aguir K, Fiorido T, Martini-Laithier V, Bouchakour R, Burtey S, et al. Skin alcohol perspiration measurements using MOX sensors. Sens Actuators B Chem 2019 Feb;280:306-312. [doi: 10.1016/j.snb.2018.09.082]
50. Jalal AH, Umasankar Y, Pretto Jr EA, Bhansali S. A wearable micro-fuel cell sensor for the determination of blood alcohol content (BAC): a multivariate regression model approach. ISSS J Micro Smart Syst 2020 Sep 17;9(2):131-142. [doi: 10.1007/s41683-020-00059-w]
51. Davidson D, Camara P, Swift R. Behavioral effects and pharmacokinetics of low-dose intravenous alcohol in humans. Alcohol Clin Exp Res 1997 Oct;21(7):1294-1299. [Medline: 9347092]
52. NIAAA Council approves definition of binge drinking. National Institute on Alcohol Abuse and Alcoholism. 2004. URL: https://pubs.niaaa.nih.gov/publications/Newsletter/winter2004/Newsletter_Number3.pdf [accessed 2022-04-07]
53. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The mixed methods appraisal tool (MMAT) version 2018 for information professionals and researchers. Educ Inf 2018 Dec 18;34(4):285-291. [doi: 10.3233/efi-180221]
54. Offenders to be banned from drinking to cut alcohol-fuelled crime. Government of United Kingdom. 2021. URL: https://www.gov.uk/government/news/offenders-to-be-banned-from-drinking-to-cut-alcohol-fuelled-crime [accessed 2021-11-24]

Abbreviations

AUC: area under the curve

BAC: blood alcohol concentration
BrAC: breath alcohol concentration
eBAC: estimated blood alcohol concentration
IoT: Internet of Things
MMAT: Mixed Methods Appraisal Tool
NHS: National Health Service
NIHR: National Institute for Health Research
PEM: proton-exchange membrane
PRISMA-P: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols
TAC: transdermal alcohol concentration
TACg: transdermal alcohol concentration gas
TAS: transdermal alcohol sensor

Edited by A Mavragani; submitted 24.11.21; peer-reviewed by M Akhlaghdoust, F Velayati, JC Shin; comments to author 19.01.22; revised version received 03.02.22; accepted 19.02.22; published 14.04.22.

Please cite as:
Brobbin E, Deluca P, Hemrage S, Drummond C
Accuracy of Wearable Transdermal Alcohol Sensors: Systematic Review
J Med Internet Res 2022;24(4):e35178
URL: https://www.jmir.org/2022/4/e35178
doi:10.2196/35178
PMID:35436239

Review

Social Media–Based Interventions for Health Behavior Change in Low- and Middle-Income Countries: Systematic Review

Jessie Seiler[1], BA, MPH; Tanya E Libby[1], BS, MPH; Emahlea Jackson[1], BA, MPH, MS; JR Lingappa[2], MD, PhD; WD Evans[3], PhD

[1]Department of Epidemiology, University of Washington, Seattle, WA, United States

[2]Departments of Global Health, Medicine, and Pediatrics, University of Washington, Seattle, WA, United States

[3]Departments of Prevention and Community Health & Global Health, George Washington University, Washington, DC, United States

Corresponding Author:
Jessie Seiler, BA, MPH
Department of Epidemiology
University of Washington
3980 15th Ave NE
Seattle, WA, 98195
United States
Phone: 1 206 543 1065
Email: jseiler@uw.edu

Abstract

Background: Despite the wealth of evidence regarding effective health behavior change techniques using digital interventions to focus on residents of high-income countries, there is limited information of a similar nature for low- and middle-income countries.

Objective: The aim of this review is to identify and describe the available literature on effective social media–based behavior change interventions within low- and middle-income countries.

Methods: This systematic review was conducted in accordance with the 2009 PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. We searched PubMed, Embase, Elsevier, CINAHL, PsycInfo, and Global Index Medicus, and the final search was conducted on April 6, 2021. We excluded studies published before 2000 because of the subject matter. We included studies that evaluated interventions conducted at least partly on a social media platform.

Results: We identified 1832 studies, of which 108 (5.89%) passed title-abstract review and were evaluated by full-text review. In all, 30.6% (33/108) were included in the final analysis. Although 22 studies concluded that the social media intervention was effective, only 13 quantified the level of social media engagement, of which, few used theory (n=8) or a conceptual model (n=5) of behavior change.

Conclusions: We identified gaps in the settings of interventions, types and sectors of interventions, length of follow-up, evaluation techniques, use of theoretical and conceptual models, and discussions of the privacy implications of social media use.

Trial Registration: PROSPERO CRD42020223572; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=223572

(J Med Internet Res 2022;24(4):e31889) doi:10.2196/31889

KEYWORDS

social media; behavior change; low- and middle-income countries; mobile phone

Introduction

Background

Social media platforms are ubiquitous in modern world. Social media has become a widely used platform for advertisers, news media, and government agencies to reach billions of users as a form of connection and a means of communication with relatives, friends, businesses, colleagues, media figures, and acquaintances. For example, Meta (formerly known as Facebook) is a major conglomerate of social media networks, and it is estimated that 2.7 billion active users engaged with their flagship Facebook social media service in mid-2020 [1]. With such a wide reach and worldwide omnipresence, social media platforms—including Instagram, Twitter, Reddit, WeChat, and others—have been of increasing interest in the

implementation of behavior change interventions and public health campaigns.

In a 2018 systematic review, Elaheebocus et al [2] determined that among peer-based social media interventions focused on tobacco smoking, nutrition, physical activity, or alcohol consumption, those with a sharing-enabled feature were most likely to elicit positive intervention outcomes. Results from multiple systematic reviews suggest that among adolescents and children, social media interventions—in comparison with in-person interventions—are underused; however, they may be effective tools for health promotion and behavior change [3-6].

Despite their potential utility, social media platforms have a checkered history of privacy and data theft. Of note is the mismatch in the intention versus actualization of privacy behavior and data sharing on social media platforms [7,8]. Although many people cite the desire to protect their privacy and limit the amount of information gathered by said platforms, average social media users do not exercise caution when granting access to third-party software or websites to use their data [7]. Given the diminished barrier to data collection as well as the high financial value and ease of access of said data, researchers have called for increased protection against malicious data exploits (eg, theft) and the development of ethical frameworks that encourage cautious behavior in both data collection and social media use [9-11].

Behavior change strategies implemented using digital technology include training, coaching, and text messages, all of which are potentially more effective with increased frequency, intensity, and follow-up [12]. Overall, the most effective health behavior change interventions use a combination of both digital and face-to-face components, lending credence to the importance of classical social behavior change modalities, including human interaction and in-person accountability [5,13-15]. The most commonly cited research gaps include multiple, noncomparable measures (eg, engagement and reach) to evaluate digital media–related behavior change campaigns [5,16,17]. Other areas highlighted for improvement include clarification of dose, intensity of intervention delivery, and measurement of long-term outcomes [17,18].

The preponderance of evidence characterizing effective behavior change techniques using digital interventions has been collected by focusing on residents of high-income countries. There are limited data of a similar nature for low- and middle-income countries (LMICs). A recent Pew Research study explored internet and smartphone use in LMICs and found a median of 67% of respondents reported having access to and using the internet, with 42% reporting access to smartphones in 2017 [19]. In the context of technology-based interventions for HIV prevention and care delivery, Maloney et al [20] found that, compared with high-income countries, the distinguishing characteristic of a successful intervention in LMICs appeared to be related to how well the intervention was tailored to serve the unique needs of a given community, village, or region, which was highly dependent on the culture within that group. Given the demonstrated need for these interventions to be crafted specifically for the setting of interest and the growing availability of technology in LMICs, a focus on behavior change interventions delivered over social media in LMICs is justified.

Objectives

The goal of this review is to identify and describe the available literature regarding effective social media–based health behavior change interventions within LMICs.

Methods

Overview

We conducted a systematic review in concordance with the 2009 PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to understand what behavioral interventions have been implemented using social media in LMICs and to characterize the evidence of their effectiveness [21]. One of our goals was to gather data on where and how these interventions are being implemented, including what subject areas have used social media interventions and for which outcomes. We also aimed to assess evaluation patterns, both in terms of the type of evaluations being carried out and whether social media interventions were deemed effective. Finally, we intended to collect information on funding sources, cost-effectiveness, and the use of theoretical and conceptual models. The review was registered with PROSPERO, where the review protocol can be viewed (registration number CRD42020223572).

Search Strategy and Selection Criteria

We searched PubMed, Embase, Elsevier, CINAHL, PsycInfo, and Global Index Medicus for studies that detailed behavior change interventions with some component conducted on a social media platform. Our search terms, which were developed with the help of both subject matter experts and a librarian with expertise in conducting systematic reviews, encompassed both social media and health behavior change concepts and included terms to describe LMICs, as well as the names of all countries categorized as LMICs by the World Bank (Multimedia Appendix 1 provides full search threads). The final search for publications included in this review was conducted on April 6, 2021.

In recognition of the fact that social media is a relatively new format, we used search string limitations to exclude studies published before the year 2000. We also excluded studies for which we could not find a full-text version in English, including conference abstracts.

Studies were selected for review if they presented original evaluation data (formative, process or implementation, outcome or effectiveness, or impact related) for a behavior change intervention that was at least partly conducted over a social media platform and used the social components of the medium [22]. We included studies that examined changes in both behavior and health knowledge.

We excluded studies that did not describe a purposeful, planned intervention; accidental changes in service delivery that resulted in natural experiments were not of interest. We also excluded studies that included only 1-way communication, such as reminder text messages, as opposed to a multidirectional

exchange of information, ideas, or opinions. For example, studies that described interventions that involved automated daily reminders for a certain behavior were excluded, as were studies that involved promulgating advertisements on a social media platform without evaluating engagement.

For our purposes, we defined social media relatively broadly and included search terms that would identify studies that used specific platforms that connect a network of individuals together for behavior change purposes, including both those tailor-made for the purposes of the intervention as well as the more established social networks that exist for commercial purposes (ie, Twitter, Facebook, and WhatsApp). However, we excluded studies that used social media to implement one-to-one conversations, such as conversations between health care providers and patients, as these conversations did not use the networking component of the apps.

In all, 2 independent reviewers used Covidence (Covidence systematic review software, Veritas Health Innovation) to screen each title and abstract to identify potentially eligible records (JS, TEL, or EJ). During screening, disagreements among the reviewers were settled through team consensus. If the disagreement could not be settled with information available in the title and abstract, the study was passed on to a full-text review. Full-text versions of potentially eligible records were reviewed and data were independently extracted by 2 reviewers, with discrepancies resolved through discussion and consensus (JS, TEL, or EJ).

Data Extraction

We extracted information on each identified program, including its setting, intended audience, and intended behavior change; the method of measuring both exposure and outcome; the strength of the descriptions of the intervention and its social media components; the social media platform used and its role; observed outcomes; how they were evaluated; cost information; and whether their design and implementation were guided by the use of a theoretical framework or conceptual model. We also looked for information on how social media's role in each intervention was described and evaluated. Data extraction was conducted by a single reviewer for each study, with 10% of the manuscripts and extracted data selected randomly for quality control checking by a second reviewer (JS, TEL, or EJ).

The evaluation-specific information extracted for each study varied depending on the stage of evaluation assessed in the publication. For the process or implementation evaluations, we looked for information on the focus of the interventions and barriers to and facilitators of implementation. For outcome evaluations, we looked for the same indicators as we did for process or implementation evaluations along with data on changed behaviors. For impact evaluations, we collected all the indicators already mentioned as well as data on changes in health outcomes among participants.

For all studies, we searched for information on funding and cost-effectiveness associated with the program. We also extracted information on whether and how each intervention used a theoretical framework in its design and implementation.

Risk of Bias Assessment

For each study, we conducted a qualitative assessment of the potential for bias based on the available descriptions of methods for inclusion in the intervention and analysis of the results, when available. We also noted the potential for bias, as recorded in the *Limitations* section.

Results

General Characteristics of Included Studies

A total of 1832 studies were identified based on the search strategy, of which 108 (5.89%) passed title-abstract review and were evaluated by full-text review (Figure 1). At full-text review, 75 studies were excluded: 22 (29%) for using a social media platform without using the networking capabilities of these platforms (ie, for using only one-to-one communications), 19 (25%) because no version of the full text of the study could be found in English, 16 (21%) because no intervention was described, 12 (16%) because the intervention did not take place in an LMIC, and 6 (8%) because the intervention did not use a social media platform at all. Finally, 33 studies were included in this review [23-55]. The key study characteristics are summarized in Tables 1 and 2.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. LMIC: low- and middle-income country.

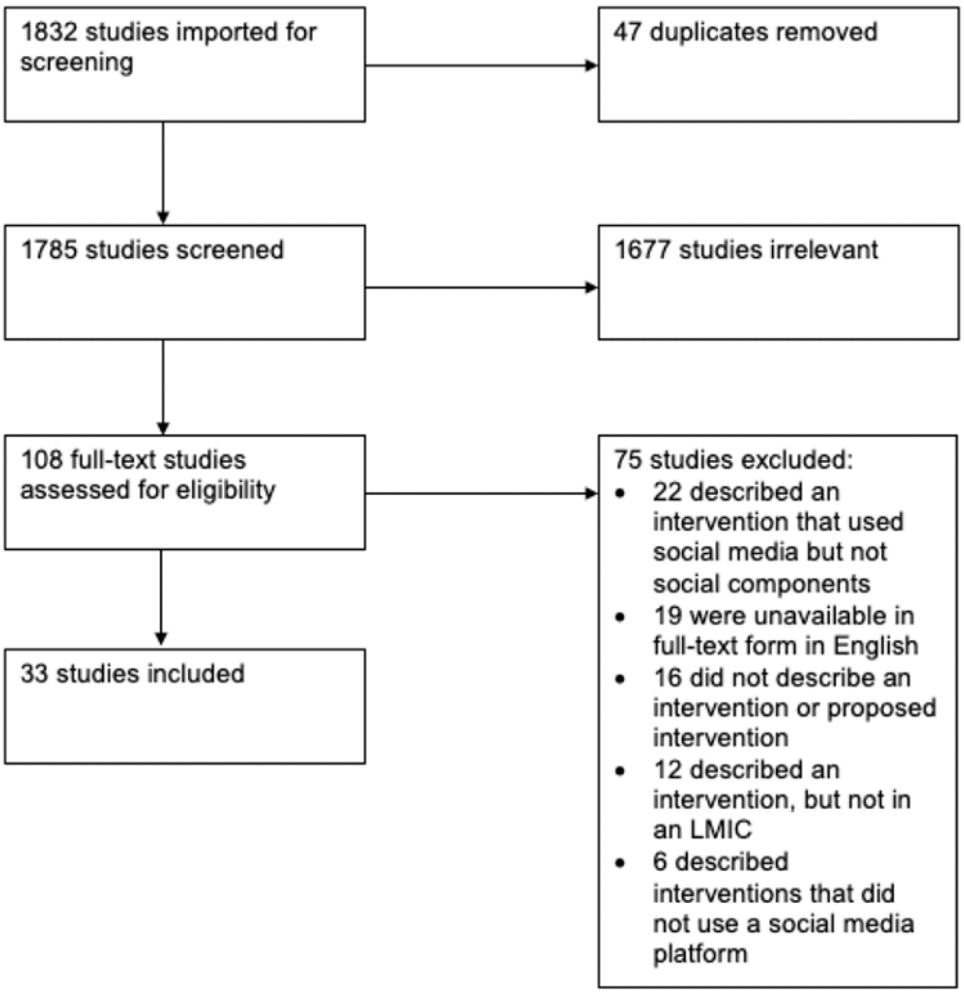


Table 1. Characteristics of the included studies.

Study	Setting	Target audience of behavior	Desired behavior change	Social media platform used	Length of follow-up
Garett et al [23]	Lima metro area, Peru	MSM[a] adults at high risk of HIV	HIV testing	Facebook	1 year
Young et al [24]	Lima metro area, Peru	MSM adults at high risk of HIV	HIV testing	Facebook	1 year
Harding et al [25]	Ghana	People who breastfeed or might be in a position to support or promote breastfeeding	Breastfeeding and supporting breastfeeding	Facebook and Twitter	No information available
Cao et al [26]	Guangdong and Shandong provinces, China	MSM aged >16 years	HIV testing	WhatsApp	3 months
Sap et al [27]	Cameroon	Adolescents and young adults living with diabetes	Diabetes knowledge and glycemic control	Not specified	2 months
Cole et al [28]	Kerala, India	People living in the area of a Nipah virus outbreak	Nipah virus knowledge	Reddit	No information available
Hutchinson et al [29]	Kenya	Adolescents and young adults	Increased knowledge and changed behaviors around family planning and income generation	Facebook, Twitter, Instagram, YouTube, and WhatsApp	1 year
Goldenhersch et al [30]	Buenos Aires, Argentina	Smokers aged between 26 and 65 years	Tobacco cessation	Created for the intervention	90 days
Cool et al [31]	The Philippines	General population	Emergency response to typhoon	Facebook, Twitter, and Instagram	No information available
Hamill et al [32]	Alexandria, Egypt	General population	Build awareness and support for an upcoming smoking ban	Facebook	1 month
Jiebing et al [33]	Tibet	Pregnant women	Increase number of prenatal care visits	WeChat	Length of pregnancy
Mo et al [34]	Shanghai, China	Undergraduate students	Increased physical activity	WeChat	7 weeks
Wu et al [35]	Guangzhou and Shenzhen, China	MSM	Increase STI[b] testing and other health-seeking activities	Created for the intervention	No information available
Ahmad et al [36]	Selangor, Malaysia	Parents and their primary school-going children (aged 8-11 years) who were overweight or obese	Physical activity, healthy diet, and reduced screen time	WhatsApp and Facebook	6 months
Cavalcanti et al [37]	João Pessoa city, Brazil	Mothers who recently gave birth	Breastfeeding	Facebook	6 months
Chen et al [38]	China	Chinese men aged 25-44 years who smoke tobacco	Smoking cessation	WeChat	6 weeks
Todorovic et al [39]	Belgrade, Serbia	First and fifth-year medical students	Physical activity	Facebook	1 month
Chai et al [40]	Zhongshan City, Guangdong Province, China	Employees of labor-intensive manufacturing factories aged >16 years	Smoking cessation	WeChat	1 month
Lwin et al [41]	Sri Lanka	General population	Dengue knowledge and prevention strategies	Created for the intervention	No information available
Pereira et al [42]	Brazil	General population	Uptake of HPV[c] vaccine	Facebook	No information available
Gamboa et al [43]	La Vega, San Francisco, and Puerto Plata; Dominican Republic	Youth aged 14-18 years in communities at high risk for arboviruses	Zika prevention behaviors	Facebook	No information available

Study	Setting	Target audience of behavior	Desired behavior change	Social media platform used	Length of follow-up
Januraga et al [44]	Indonesia	Urban, unmarried adolescent girls aged 16-19 years	Healthy diet	Facebook, Instagram, YouTube, and another app created for the intervention	No information available
Thammasarn and Banchonhattakit [45]	Nakhon Ratchasima Province, Thailand	Senior-primary school students	Physical activity and healthy diet	Not specified	No information available
He et al [46]	China	General population	Weight loss	WeChat	No information available
Souza et al [47]	Brazil	Community leaders	Civic engagement in general public health work	WhatsApp	No information available
Chiu et al [48]	Peru	MSM	HIV prevention and testing	Facebook	No information available
Chiu et al [49]	Peru	MSM	HIV prevention and testing	Facebook	No information available
Purdy [50]	Turkey	General population	Condom use	Facebook	No information available
Parsapure et al [51]	Kermanshah, Iran	General population	Vaginal health	Not specified	6 months
Wu et al [52]	Huzhu County, Qinghai Province, China	Breastfeeding mothers aged >17 years	Breastfeeding	WeChat	5 months
Hutchinson et al [53]	Ghana	Adolescent girls	Refusal to smoke tobacco	Facebook, Instagram, and YouTube	No information available
Diamond-Smith [54]	India	Women aged between 18 and 49 years	Anemia-related knowledge	Facebook	5 months
Chang et al [55]	Zhejiang Province, China	General population	Physical activity and healthy diet	WeChat	No information available

[a]MSM: men who have sex with men.

[b]STI: sexually transmitted infection.

[c]HPV: human papillomavirus.

Table 2. Methodologic quality, cost, and funding of included studies.

Study	Social media's role in this intervention or outcome clearly reported	This role quantified[a]	Social media described as effective	Theoretical model application	Conceptual model application	Cost of intervention reported	Funding source reported	Sources of bias reported
Garett et al [23]	Yes	No	Yes	No	No	No	Yes	Recall
Young et al [24]	Yes	Yes	Yes	No	No	No	Yes	Self-reported
Harding et al [25]	Yes	Yes	Yes	No	No	No	Yes	Nonrepresentative of general population owing to internet or social media access
Cao et al [26]	Yes	No	Yes	No	No	No	No	None
Sap et al [27]	Yes	No	Yes	No	No	No	No	Nonrepresentative of general population owing to internet or social media access; self-reported
Cole et al [28]	Yes	Yes	Yes	No	No	No	Yes	None
Hutchinson et al [29]	No	No	Yes	Transtheoretical model	No	No	Yes	Nonrepresentative of general population owing to internet or social media access
Goldenhersch et al [30]	Yes	Yes	Yes	Contemplation ladder	No	No	Yes	Short follow-up; self-reported
Cool et al, 2015 [31]	Yes	Yes	Yes	No	No	No	Yes	None
Hamill et al, 2015 [32]	Yes	Yes	Yes	No	No	Yes	Yes	None
Jiebing et al [33]	Yes	No	Yes	No	No	No	Yes	None
Mo et al [34]	Yes	No	Yes	Theory of planned behavior	No	No	Yes	Self-reported
Wu et al [35]	No	No	No	No	No	No	Yes	Nonrepresentative of general population owing to internet or social media access
Ahmad et al [36]	Yes	Yes	Yes	Social cognitive theory	Yes	No	Yes	Nonrepresentative of general population owing to internet or social media access; other selection bias
Cavalcanti et al [37]	Yes	Yes	Yes	No	No	No	No	Research team was not blinded to randomization
Chen et al [38]	No	No	No	COM-B[b] and Behavior Change Wheel framework	Yes	No	Yes	Participants and researchers were not blinded
Todorovic et al [39]	Yes	No	Yes	No	No	No	Yes	No randomization
Chai et al [40]	No	No	No	No	Yes	No	Yes	Lost to follow-up
Lwin et al [41]	No	No	No	Protection motivation theory	Yes	No	Yes	Nonrepresentative of general population owing to internet or social media access

Study	Social media's role in this intervention or outcome clearly reported	This role quantified[a]	Social media described as effective	Theoretical model application	Conceptual model application	Cost of intervention reported	Funding source reported	Sources of bias reported
Pereira et al [42]	Yes	Yes	No	No	No	No	No	Nonrepresentative of general population owing to internet or social media access
Gamboa et al [43]	Yes	Yes	Yes	Social cognitive theory	No	No	Yes	Self-reported
Januraga et al [44]	Yes	No	No	Technology acceptance model	No	No	Yes	Nonrepresentative of general population owing to internet or social media access
Thammasarn and Banchonhattakit [45]	No	No	No	No	No	No	Yes	None
He et al [46]	Yes	No	Yes	No	No	No	Yes	None
Souza et al [47]	No	No	No	No	No	No	No	None
Chiu et al [48]	Yes	No	Yes	No	No	No	No	Self-reported; lost to follow-up; nonrepresentative of general population owing to internet or social media access
Chiu et al [49]	Yes	No	No	No	No	No	No	Self-reported; lost to follow-up; nonrepresentative of general population owing to internet or social media access
Purdy [50]	Yes	Yes	Yes	No	No	No	Yes	None
Parsapure et al [51]	No	No	Yes	No	No	No	Yes	None
Wu et al [52]	Yes	No	Yes	No	No	No	Yes	Lost to follow-up
Hutchinson et al [53]	No	Yes	No	No	Yes	No	Yes	Self-reported; interviewer bias
Diamond-Smith [54]	Yes	Yes	Yes	No	No	No	Yes	Selection bias
Chang et al [55]	Yes	Yes	No	No	No	No	No	None

[a]Quantified through clicks, shares, comments, or other method of engagement.

[b]COM-B: Capability Opportunity Motivation Behavior.

Geographic and Methodological Characteristics of Included Studies

The studies included for review were conducted in geographically diverse LMICs, with China (8/33, 24%), Peru (4/33, 12%, all from the same program), and Brazil (3/33, 9%) being the sites with the most social media–based interventions (Table 3). Other countries included Argentina, Cameroon, the Dominican Republic, Egypt, Ghana, India, Indonesia, Iran, Kenya, Malaysia, Serbia, Sri Lanka, Thailand, the Philippines, and Turkey (Table 1). Of these, 33% (11/33) are upper-middle–income countries and 21% (7/33) are lower-middle–income countries.

Of the 33 studies included in the review, 23 (70%) were limited to a particular subnational locality, typically either a large city or a particular region, and 10 (30%) were designed to be national in scope.

Common study designs included randomized trials (12/23, 52%) and observational studies (11/23, 48%). Only 13% (3/23) of studies were qualitative, and none used mixed methods.

Common focus populations included a particular age group of interest (11/23, 49%), the general population (8/23, 35%), and men who have sex with men (6/23, 26%). The desired behavior change component varied widely, but studies most frequently aimed to change HIV testing and knowledge (5/23, 22%), increase physical activity and weight loss (5/23, 22%), and smoking cessation (5/23, 22%).

Studies have frequently combined types of evaluations, with 6% (2/33) of studies including components related to formative

evaluations, 52% (17/33) with process evaluation components, 76% (25/33) with outcome indicators, and 18% (6/33) with impact measurements. In all, 9% (3/33) of studies were described with insufficient detail to clarify the specific type of evaluation conducted.

Although most studies described the overall interventions thoroughly enough that they could be understood (24/33, 73%), few reported social media use with sufficient clarity (19/33, 58%). In 42% (14/33) of studies, the role of social media in greater intervention was quantified using some measure of engagement, including clicks, likes, comments, shares, and retweets. However, the authors concluded that the use of social media was effective (22/33, 67%). Studies, especially those with data on outcomes or impacts, frequently paired data on social media use with a measurement of effectiveness captured outside of social media use, such as changes in anthropometric measurements, knowledge of a specific health topic, or smoking status (22/33, 67%).

Table 3. Characteristics of included studies (N=33).

Study characteristics	Studies, n (%)
Country	
China	8 (24)
Peru	4 (12)
Brazil	3 (9)
Other[a]	19 (58)
Social media platform[b]	
Facebook	17 (52)
WeChat	8 (24)
Instagram	4 (12)
Platform created for intervention	4 (12)
YouTube	3 (9)
WhatsApp	3 (9)
Twitter	3 (9)
Reddit	1 (3)
Multiple platforms	6 (18)
Methodologic quality	
The role of social media in this intervention or outcome was clearly reported	24 (73)
This role was quantified through clicks, shares, comments, or other method of engagement	14 (42)
Social media was described as effective	22 (67)
A theoretical or conceptual model was used	10 (30)

[a]Argentina, Cameroon, the Dominican Republic, Egypt, Ghana, India, Indonesia, Iran, Kenya, Malaysia, Serbia, Sri Lanka, Thailand, the Philippines, and Turkey.

[b]Categories are not mutually exclusive.

Facebook was the most common social media platform used in the studies (17/33, 52%), followed by WeChat (8/33, 24%), Instagram (4/33, 12%), and WhatsApp (3/33, 9%; Table 3). A few studies (4/33, 12%) designed their own apps for use on mobile phones with built-in social media components and 18% (6/33) of studies spread their efforts across several platforms (eg, Facebook, Twitter, and Instagram).

Few studies included a description of either a theoretical model (8/33, 24%) or developed a conceptual model (5/33, 15%) that guided researchers' efforts in the design or evaluation stages. Although clear information on the cost of the interventions was rare (1/33, 3%), most (24/33, 73%) studies included information on sources of funding.

Although every study focused on an intervention that, by its nature, required access to technology and the internet, relatively few highlighted this as a potential source of selection bias that might lessen an intervention's external validity (11/33, 33%). This suggests that researchers do not perceive technology access as an obstacle to the effective implementation of such interventions in LMICs.

The length of observation after the intervention ended for most of the studies was relatively short, with no studies following up with their participants for more than a year, and half of the studies (17/33, 52%) did not report any follow-up data.

Finally, none of the studies included in this systematic review reported on the methods used to diminish the possibility of interference or data theft on behalf of the participants from

internet service providers, software developers, social media services (where applicable), or other interactive users, despite the sharing of data with a third-party service being a requisite of participation eligibility.

Discussion

Principal Findings

In this systematic review of studies on interventions that use social media to encourage health behavior change in LMICs, we evaluated 33 studies across a range of interventions, settings, and techniques. We identified important gaps in the types and sectors of interventions, length of follow-up, evaluation techniques, use of theoretical and conceptual models, and discussions of the privacy implications of social media use. In addition, we found that although social media interventions have been conducted in a number of LMICs worldwide, few have been conducted in the poorest countries and few have been done in sub-Saharan Africa.

The range of interventions described in the studies included in this review was limited. We found that the body of literature on behavior change in LMICs is not able to address the question of whether social media is generally useful in these settings or even appropriate for certain types of behavior change work specifically. These determinations will likely need to be extremely context-specific, given the variation in social media access and willingness to use it for this type of work.

A conspicuous limitation of most of the studies included in this review was the lack of data on long-term outcomes and impact. Behavior change can take time, and the potential for regression to earlier states is well known. Future research should include a longitudinal follow-up to assess the long-term effects of social media behavioral interventions. In addition, there was a lack of evidence on the effectiveness of theories of change in social media interventions, and future research should focus on testing processes of change.

Given the relative novelty of behavior change interventions conducted via social media, the lack of formative research evaluating the feasibility, appropriateness, and acceptability of specific types of projects is troubling. The earliest study to use formative research was published in 2015, many years after social media became widespread. Formative research helps ensure that a specific intervention is likely to be needed, understood, and accepted by the population of interest. Without this critical step, it is possible to miss important modifications that would have helped shape extant interventions into successful social media–based projects [56]. There are examples of such formative studies in high-income countries, but we found no such formative research in this review. A more rigorous application of the principles of program evaluation will help develop targeted, effective social media–based interventions.

One of the strengths of social media interventions is the availability of objective dosage and exposure data from analytics. However, our review found that some studies reported that their social media efforts were effective without clearly reporting quantitative data (eg, clicks, shares, and views) on social media use. Future research should examine the characteristics of engagement exposure to evaluate dose–response effects, that is, to determine whether more exposure or exposure of a specific type is associated with successful behavior change. It is important to objectively attribute intervention effects to observed behavior changes and build an evidence base in the field.

Our review identified important gaps in the application of theoretical and conceptual models to explain why a particular intervention was needed, how it might have been expected to work, and how it could have been evaluated. The relatively low number of studies embracing this method is concerning, especially given the novelty of social media–based interventions. The use of evidence-based behavioral theory and a program-specific model are hallmarks of well-designed interventions, and these methods need to be explicitly included to support evaluations of future social media interventions in LMICs [57].

Finally, this review raises the question of the ethical implications of using a third-party social media service as a medium for conducting public health experiments. Of notable concern are the various ways in which a study participant's privacy could be violated. When a user interacts with a wide-reaching social media platform (eg, Facebook or Twitter), there are multiple stakeholders who may have a vested interest in harvesting any data or communication provided by the participant: internet service providers, social media platforms, other interactive users, or even a governmental entity with a backdoor encryption policy. Ensuring participant security and protection of privacy are among the most critical components of ethical research; explicit explanation of these risks to personal data loss is necessary to incorporate in every public health social media study.

Limitations

Although this study was conducted following the PRISMA guidelines, there are some important limitations to this systematic review. We did not conduct any quantitative review of the studies or meta-analysis; therefore, we were unable to provide a quantitative assessment of the effectiveness of the interventions of interest. We also limited our search by not availing ourselves of unpublished literature or literature published in a language other than English, where it is possible that we would have found additional intervention descriptions.

Conclusions

In conclusion, as social media becomes a more powerful and omnipresent factor in people's lives, its potential as a platform for public health work has grown rapidly. This systematic review of social media–based behavior change interventions conducted in LMICs highlights the need for diversity and methodological rigor at every step in the planning, implementation, and evaluation stages of programming.

Acknowledgments

This study was completed by the Strategic Analysis, Research, and Training Program at the University of Washington. The Strategic Analysis, Research, and Training Program is a collaborative effort with, and is funded by, the Bill & Melinda Gates Foundation. The funder commissioned the study but did not have control over the design, data collection, analysis, conclusions, or preparation of the manuscript.

Authors' Contributions

JS reviewed titles and abstracts, reviewed full-text manuscripts, verified the data, conducted the data analysis and interpretation, drafted and edited the manuscript, and managed the review process. TEL coordinated search string development, reviewed titles and abstracts, reviewed full-text manuscripts, verified the data, assisted with data analysis and interpretation, and drafted the manuscript. EJ coordinated the initial literature review, reviewed titles and abstracts, reviewed full-text manuscripts, verified the data, assisted with the data analysis and interpretation, and drafted the manuscript. JRL provided project supervision and edited the manuscript. WDE conceived the idea for this work, edited the manuscript, and approved the final work.
All authors have full access to all data in the study and accept responsibility for the decision to submit for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search string for Embase.
[DOCX File , 15 KB - jmir_v24i4e31889_app1.docx]

References

1. Number of monthly active Facebook users worldwide as of 4th quarter 2021 (in millions). Statista. 2020. URL: https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/ [accessed 2020-11-19]
2. Elaheebocus SM, Weal M, Morrison L, Yardley L. Peer-based social media features in behavior change interventions: systematic review. J Med Internet Res 2018;20(2):e20 [FREE Full text] [doi: 10.2196/jmir.8342] [Medline: 29472174]
3. Hsu MS, Rouf A, Allman-Farinelli M. Effectiveness and behavioral mechanisms of social media interventions for positive nutrition behaviors in adolescents: a systematic review. J Adolesc Health 2018;63(5):531-545. [doi: 10.1016/j.jadohealth.2018.06.009] [Medline: 30197198]
4. Chau MM, Burgermaster M, Mamykina L. The use of social media in nutrition interventions for adolescents and young adults-a systematic review. Int J Med Inform 2018;120:77-91 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.10.001] [Medline: 30409348]
5. Brigden A, Anderson E, Linney C, Morris R, Parslow R, Serafimova T, et al. Digital behavior change interventions for younger children with chronic health conditions: systematic review. J Med Internet Res 2020;22(7):e16924 [FREE Full text] [doi: 10.2196/16924] [Medline: 32735227]
6. Chan L, O'Hara B, Phongsavan P, Bauman A, Freeman B. Review of evaluation metrics used in digital and traditional tobacco control campaigns. J Med Internet Res 2020;22(8):e17432 [FREE Full text] [doi: 10.2196/17432] [Medline: 32348272]
7. Acquisti A, Brandimarte L, Loewenstein G. Privacy and human behavior in the age of information. Science 2015;347(6221):509-514. [doi: 10.1126/science.aaa1465] [Medline: 25635091]
8. Fleming P, Bayliss AP, Edwards SG, Seger CR. The role of personal data value, culture and self-construal in online privacy behaviour. PLoS One 2021;16(7):e0253568 [FREE Full text] [doi: 10.1371/journal.pone.0253568] [Medline: 34270577]
9. Aïmeur E, Díaz Ferreyra N, Hage H. Manipulation and malicious personalization: exploring the self-disclosure biases exploited by deceptive attackers on social media. Front Artif Intell 2019;2:26 [FREE Full text] [doi: 10.3389/frai.2019.00026] [Medline: 33733115]
10. Phillips AM. Genomic privacy and direct-to-consumer genetics: big consumer genetic data -- what's in that contract? In: IEEE Symposium on Security and Privacy Workshops. 2015 Presented at: SPW '15; May 21-22, 2015; San Jose, CA p. 60-64. [doi: 10.1109/spw.2015.19]
11. Adams SA, Van Veghel D, Dekker L. Developing a research agenda on ethical issues related to using social media in healthcare. Camb Q Healthc Ethics 2015;24(3):293-302. [doi: 10.1017/S0963180114000619] [Medline: 26059955]
12. Chapman E, Haby MM, Toma TS, de Bortoli MC, Illanes E, Oliveros MJ, et al. Knowledge translation strategies for dissemination with a focus on healthcare recipients: an overview of systematic reviews. Implement Sci 2020;15(1):14 [FREE Full text] [doi: 10.1186/s13012-020-0974-3] [Medline: 32131861]
13. Mansour D, Nashwan A, Abu Rasheed H, Hararah M, Nassar H, Abu Abbas R, et al. Use of social media in breast cancer awareness: GCC countries' experience. J Glob Oncol 2018;4(Supplement 2):30S. [doi: 10.1200/jgo.18.66200]

14. Swanton R, Allom V, Mullan B. A meta-analysis of the effect of new-media interventions on sexual-health behaviours. Sex Transm Infect 2015;91(1):14-20. [doi: 10.1136/sextrans-2014-051743] [Medline: 25433051]
15. Ziegeldorf A, Wagner P, Wulff H. Effects of media-assisted therapeutic approaches on physical activity of obese adults: a systematic review. BMC Endocr Disord 2020;20(1):28 [FREE Full text] [doi: 10.1186/s12902-020-0505-x] [Medline: 32085762]
16. Johns DJ, Langley TE, Lewis S. Use of social media for the delivery of health promotion on smoking, nutrition, and physical activity: a systematic review. Lancet 2017;390:S49. [doi: 10.1016/s0140-6736(17)32984-7]
17. Naslund JA, Kim SJ, Aschbrenner KA, McCulloch LJ, Brunette MF, Dallery J, et al. Systematic review of social media interventions for smoking cessation. Addict Behav 2017;73:81-93 [FREE Full text] [doi: 10.1016/j.addbeh.2017.05.002] [Medline: 28499259]
18. Simeon R, Dewidar O, Trawin J, Duench S, Manson H, Pardo Pardo J, et al. Behavior change techniques included in reports of social media interventions for promoting health behaviors in adults: content analysis within a systematic review. J Med Internet Res 2020;22(6):e16002 [FREE Full text] [doi: 10.2196/16002] [Medline: 32525482]
19. Poushter J, Bishop C, Chwe H. Social media use continues to rise in developing countries but plateaus across developed ones: digital divides remain, both within and across countries. Pew Research Center. 2018. URL: https://tinyurl.com/2p8dph22 [accessed 2020-11-19]
20. Maloney KM, Bratcher A, Wilkerson R, Sullivan PS. Electronic and other new media technology interventions for HIV care and prevention: a systematic review. J Int AIDS Soc 2020;23(1):e25439 [FREE Full text] [doi: 10.1002/jia2.25439] [Medline: 31909896]
21. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
22. Types of evaluation. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/std/Program/pupestd/Types%20of%20Evaluation.pdf [accessed 2021-03-25]
23. Garett R, Menacho L, Young SD. Ethical issues in using social media to deliver an HIV prevention intervention: results from the HOPE Peru study. Prev Sci 2017;18(2):225-232 [FREE Full text] [doi: 10.1007/s11121-016-0739-z] [Medline: 27933425]
24. Young SD, Cumberland WG, Nianogo R, Menacho LA, Galea JT, Coates T. The HOPE social media intervention for global HIV prevention in Peru: a cluster randomised controlled trial. Lancet HIV 2015;2(1):e27-e32 [FREE Full text] [doi: 10.1016/S2352-3018(14)00006-X] [Medline: 26236767]
25. Harding K, Aryeetey R, Carroll G, Lasisi O, Pérez-Escamilla R, Young M. Breastfeed4Ghana: design and evaluation of an innovative social media campaign. Matern Child Nutr 2020;16(2):e12909 [FREE Full text] [doi: 10.1111/mcn.12909] [Medline: 31867865]
26. Cao B, Saha PT, Leuba SI, Lu H, Tang W, Wu D, et al. Recalling, sharing and participating in a social media intervention promoting HIV testing: a longitudinal analysis of HIV testing among MSM in China. AIDS Behav 2019;23(5):1240-1249 [FREE Full text] [doi: 10.1007/s10461-019-02392-0] [Medline: 30632008]
27. Sap S, Kondo E, Sobngwi E, Mbono R, Tatah S, Dehayem M, et al. Effect of patient education through a social network in young patients with type 1 diabetes in a Sub-Saharan context. Pediatr Diabetes 2019;20(3):361-365. [doi: 10.1111/pedi.12835] [Medline: 30779272]
28. Cole J, Ezziane S, Watkins C. Rapid creation of an online discussion space (r/nipah) during a serious disease outbreak: observational study. JMIR Public Health Surveill 2019;5(4):e13753 [FREE Full text] [doi: 10.2196/13753] [Medline: 31719021]
29. Hutchinson P, Mirzoyants A, Leyton A. Empowering youth for social change through the Shujaaz multimedia platform in Kenya. Int J Adolesc Youth 2019;24(1):102-116 [FREE Full text] [doi: 10.1080/02673843.2018.1475287]
30. Goldenhersch E, Thrul J, Ungaretti J, Rosencovich N, Waitman C, Ceberio MR. Virtual reality smartphone-based intervention for smoking cessation: pilot randomized controlled trial on initial clinical efficacy and adherence. J Med Internet Res 2020;22(7):e17571 [FREE Full text] [doi: 10.2196/17571] [Medline: 32723722]
31. Cool CT, Claravall MC, Hall JL, Taketani K, Zepeda JP, Gehner M, et al. Social media as a risk communication tool following typhoon Haiyan. Western Pac Surveill Response J 2015;6 Suppl 1:86-90 [FREE Full text] [doi: 10.5365/WPSAR.2015.6.2.HYN_013] [Medline: 26767143]
32. Hamill S, Turk T, Murukutla N, Ghamrawy M, Mullin S. I 'like' MPOWER: using Facebook, online ads and new media to mobilise tobacco control communities in low-income and middle-income countries. Tob Control 2015;24(3):306-312. [doi: 10.1136/tobaccocontrol-2012-050946] [Medline: 24335477]
33. Chen J, Huang J, Ooi S, Lin L, Chen C, Liu Y, et al. Effect of flexible patterns of health education on enhancing the compliance of pregnant women from Tibet, China. Medicine (Baltimore) 2020;99(1):e18447 [FREE Full text] [doi: 10.1097/MD.0000000000018447] [Medline: 31895773]
34. Mo D, Xiang M, Luo M, Dong Y, Fang Y, Zhang S, et al. Using gamification and social incentives to increase physical activity and related social cognition among undergraduate students in Shanghai, China. Int J Environ Res Public Health 2019;16(5):858 [FREE Full text] [doi: 10.3390/ijerph16050858] [Medline: 30857261]

35. Wu D, Huang W, Zhao P, Li C, Cao B, Wang Y, et al. A crowdsourced physician finder prototype platform for men who have sex with men in China: qualitative study of acceptability and feasibility. JMIR Public Health Surveill 2019;5(4):e13027 [FREE Full text] [doi: 10.2196/13027] [Medline: 31596245]
36. Ahmad N, Shariff ZM, Mukhtar F, Lye MS. Family-based intervention using face-to-face sessions and social media to improve Malay primary school children's adiposity: a randomized controlled field trial of the Malaysian REDUCE programme. Nutr J 2018;17(1):74 [FREE Full text] [doi: 10.1186/s12937-018-0379-1] [Medline: 30071855]
37. Cavalcanti DS, Cabral CS, de Toledo Vianna RP, Osório MM. Online participatory intervention to promote and support exclusive breastfeeding: randomized clinical trial. Matern Child Nutr 2019;15(3):e12806 [FREE Full text] [doi: 10.1111/mcn.12806] [Medline: 30825414]
38. Chen J, Ho E, Jiang Y, Whittaker R, Yang T, Bullen C. Mobile social network-based smoking cessation intervention for Chinese male smokers: pilot randomized controlled trial. JMIR Mhealth Uhealth 2020;8(10):e17522 [FREE Full text] [doi: 10.2196/17522] [Medline: 33095184]
39. Todorovic J, Terzic-Supic Z, Djikanovic B, Nesic D, Piperac P, Stamenkovic Z. Can social media intervention improve physical activity of medical students? Public Health 2019;174:69-73. [doi: 10.1016/j.puhe.2019.05.030] [Medline: 31323599]
40. Chai W, Zou G, Shi J, Chen W, Gong X, Wei X, et al. Evaluation of the effectiveness of a WHO-5A model based comprehensive tobacco control program among migrant workers in Guangdong, China: a pilot study. BMC Public Health 2018;18(1):296 [FREE Full text] [doi: 10.1186/s12889-018-5182-6] [Medline: 29486753]
41. Lwin MO, Vijaykumar S, Foo S, Fernando ON, Lim G, Panchapakesan C, et al. Social media-based civic engagement solutions for dengue prevention in Sri Lanka: results of receptivity assessment. Health Educ Res 2016;31(1):1-11 [FREE Full text] [doi: 10.1093/her/cyv065] [Medline: 26668207]
42. Pereira da Veiga CR, Semprebon E, da Silva JL, Lins Ferreira V, Pereira da Veiga C. Facebook HPV vaccine campaign: insights from Brazil. Hum Vaccin Immunother 2020;16(8):1824-1834 [FREE Full text] [doi: 10.1080/21645515.2019.1698244] [Medline: 31916905]
43. Gamboa J, Lamb MM, de la Cruz P, Bull S, Olson D. Using social media to increase preventative behaviors against arboviral diseases: a pilot study among teens in the Dominican Republic. Mhealth 2019;5:30 [FREE Full text] [doi: 10.21037/mhealth.2019.07.03] [Medline: 31559275]
44. Januraga PP, Izwardi D, Crosita Y, Indrayathi PA, Kurniasari E, Sutrisna A, et al. Qualitative evaluation of a social media campaign to improve healthy food habits among urban adolescent females in Indonesia. Public Health Nutr 2021;24(S2):s98-107. [doi: 10.1017/S1368980020002992] [Medline: 32907648]
45. Thammasarn K, Banchonhattakit P. Effects of food fit for fun program with social media used on health literacy and obesity prevention behaviors among senior-primary school students, in Nakhon Ratchasima province Thailand. Indian J Public Health Res Dev 2020;11(7):1291-1297. [doi: 10.37506/ijphrd.v11i7.10274]
46. He C, Wu S, Zhao Y, Li Z, Zhang Y, Le J, et al. Social media-promoted weight loss among an occupational population: cohort study using a WeChat mobile phone app-based campaign. J Med Internet Res 2017;19(10):e357 [FREE Full text] [doi: 10.2196/jmir.7861] [Medline: 29061555]
47. de Souza CT, de Santana CS, Ferreira P, Nunes JA, de Lourdes Benamor Teixeira M, da Silveira Gouvêa MI. Caring in the age of COVID-19: lessons from science and society. Cad Saude Publica 2020;36(6):e00115020 [FREE Full text] [doi: 10.1590/0102-311X00115020] [Medline: 32609168]
48. Chiu CJ, Menacho L, Fisher C, Young SD. Ethics issues in social media-based HIV prevention in low- and middle-income countries. Camb Q Healthc Ethics 2015;24(3):303-310 [FREE Full text] [doi: 10.1017/S0963180114000620] [Medline: 26059956]
49. Chiu CJ, Menacho L, Young SD. The association between age and ethics-related issues in using social media for HIV prevention in Peru. Ethics Behav 2016;26(2):99-109 [FREE Full text] [doi: 10.1080/10508422.2014.991868] [Medline: 27034609]
50. Purdy CH. Using the internet and social media to promote condom use in Turkey. Reprod Health Matters 2011;19(37):157-165. [doi: 10.1016/S0968-8080(11)37549-0] [Medline: 21555096]
51. Parsapure R, Rahimiforushani A, Majlessi F, Montazeri A, Sadeghi R, Garmarudi G. Impact of health-promoting educational intervention on lifestyle (nutrition behaviors, physical activity and mental health) related to vaginal health among reproductive-aged women with vaginitis. Iran Red Crescent Med J 2016;18(10):e37698 [FREE Full text] [doi: 10.5812/ircmj.37698] [Medline: 28184325]
52. Wu Q, Huang Y, Liao Z, van Velthoven MH, Wang W, Zhang Y. Effectiveness of WeChat for improving exclusive breastfeeding in Huzhu County China: randomized controlled trial. J Med Internet Res 2020;22(12):e23273 [FREE Full text] [doi: 10.2196/23273] [Medline: 33270026]
53. Hutchinson P, Leyton A, Meekers D, Stoecker C, Wood F, Murray J, et al. Evaluation of a multimedia youth anti-smoking and girls' empowerment campaign: SKY Girls Ghana. BMC Public Health 2020;20(1):1734 [FREE Full text] [doi: 10.1186/s12889-020-09837-5] [Medline: 33203403]
54. Diamond-Smith N, Holton AE, Francis S, Bernard D. Addressing anemia among women in India-an informed intervention using Facebook Ad Manager. Mhealth 2020;6:39 [FREE Full text] [doi: 10.21037/mhealth-19-237a] [Medline: 33437835]

55. Chang L, Chattopadhyay K, Li J, Xu M, Li L. Interplay of support, comparison, and surveillance in social media weight management interventions: qualitative study. JMIR Mhealth Uhealth 2021;9(3):e19239 [FREE Full text] [doi: 10.2196/19239] [Medline: 33646130]
56. Evans WD, Harrington C, Patchen L, Andrews V, Gaminian A, Ellis LP, et al. Design of a novel digital intervention to promote healthy weight management among postpartum African American women. Contemp Clin Trials Commun 2019;16:100460 [FREE Full text] [doi: 10.1016/j.conctc.2019.100460] [Medline: 31650078]
57. Bluethmann SM, Bartholomew LK, Murphy CC, Vernon SW. Use of theory in behavior change interventions. Health Educ Behav 2017;44(2):245-253 [FREE Full text] [doi: 10.1177/1090198116647712] [Medline: 27226430]

Abbreviations

LMIC: low- and middle-income country
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by A Mavragani; submitted 08.07.21; peer-reviewed by S Chapman, E Brainin, J Lu, S Tomczyk, A Louren; comments to author 26.08.21; revised version received 21.10.21; accepted 05.02.22; published 14.04.22.

Please cite as:
Seiler J, Libby TE, Jackson E, Lingappa JR, Evans WD
Social Media–Based Interventions for Health Behavior Change in Low- and Middle-Income Countries: Systematic Review
J Med Internet Res 2022;24(4):e31889
URL: https://www.jmir.org/2022/4/e31889
doi:10.2196/31889
PMID:35436220

Review

Recommendations for Defining and Reporting Adherence Measured by Biometric Monitoring Technologies: Systematic Review

Iredia M Olaye[1], MSc, PhD; Mia P Belovsky[2], BSc; Lauren Bataille[3], MSc; Royce Cheng[4], MSc, MBA; Ali Ciger[5], MBA; Karen L Fortuna[6], PhD, LICSW; Elena S Izmailova[7], PhD; Debbe McCall, MBA; Christopher J Miller[8], MStat; Willie Muehlhausen[9], DVM; Carrie A Northcott[10], PhD; Isaac R Rodriguez-Chavez[11], MSc, MHS, PhD; Abhishek Pratap[12,13,14,15], PhD; Benjamin Vandendriessche[16,17], PhD; Yaara Zisman-Ilani[18], MA, PhD; Jessie P Bakker[19], MSc, PhD

[1]Department of Medicine Division of Clinical Epidemiology and Evaluative Sciences Research, Weill Cornell Medical College Cornell University, New York, NY, United States

[2]Sidney Kimmel Medical College at Thomas Jefferson University, Philadelphia, PA, United States

[3]Novartis Pharmaceuticals Corporation, East Hanover, NJ, United States

[4]Health Platforms, Verily Life Sciences, Cambridge, MA, United States

[5]Pfizer, Berlin, Germany

[6]Giesel School of Medicine at Dartmouth College, Hanover, NH, United States

[7]Koneksa Health, New York, NY, United States

[8]AstraZeneca Pharmaceuticals LP, Gaithersburg, MD, United States

[9]SAFIRA Clinical Research, Cloughjordan, Ireland

[10]Pfizer Inc, Cambridge, MA, United States

[11]ICON plc, Blue Bell, PA, United States

[12]CAMH Krembil Center for Neuroinformatics, Toronto, ON, Canada

[13]Vector Institute, Toronto, ON, Canada

[14]Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, United States

[15]Institute of Psychiatry, Psychology, and Neuroscience, Kings College London, London, United Kingdom

[16]Byteflies, Antwerp, Belgium

[17]Department of Electrical, Computer, and Systems Engineering, Case Western Reserve University, Cleveland, OH, United States

[18]Department of Social and Behavioral Sciences; College of Public Health, Temple University, Philadelphia, PA, United States

[19]Signifier Medical Technologies, Needham, MA, United States

Corresponding Author:
Jessie P Bakker, MSc, PhD
Signifier Medical Technologies
175 Highland Ave
Needham, MA
United States
Phone: 1 6173866968
Email: jessie.b@signifiermedical.com

Abstract

Background: Suboptimal adherence to data collection procedures or a study intervention is often the cause of a failed clinical trial. Data from connected sensors, including wearables, referred to here as biometric monitoring technologies (BioMeTs), are capable of capturing adherence to both digital therapeutics and digital data collection procedures, thereby providing the opportunity to identify the determinants of adherence and thereafter, methods to maximize adherence.

Objective: We aim to describe the methods and definitions by which adherence has been captured and reported using BioMeTs in recent years. Identifying key gaps allowed us to make recommendations regarding minimum reporting requirements and consistency of definitions for BioMeT-based adherence data.

Methods: We conducted a systematic review of studies published between 2014 and 2019, which deployed a BioMeT outside the clinical or laboratory setting for which a quantitative, nonsurrogate, sensor-based measurement of adherence was reported. After systematically screening the manuscripts for eligibility, we extracted details regarding study design, participants, the BioMeT or BioMeTs used, and the definition and units of adherence. The primary definitions of adherence were categorized as a continuous variable based on duration (highest resolution), a continuous variable based on the number of measurements completed, or a categorical variable (lowest resolution).

Results: Our PubMed search terms identified 940 manuscripts; 100 (10.6%) met our eligibility criteria and contained descriptions of 110 BioMeTs. During literature screening, we found that 30% (53/177) of the studies that used a BioMeT outside of the clinical or laboratory setting failed to report a sensor-based, nonsurrogate, quantitative measurement of adherence. We identified 37 unique definitions of adherence reported for the 110 BioMeTs and observed that uniformity of adherence definitions was associated with the resolution of the data reported. When adherence was reported as a continuous time-based variable, the same definition of adherence was adopted for 92% (46/50) of the tools. However, when adherence data were simplified to a categorical variable, we observed 25 unique definitions of adherence reported for 37 tools.

Conclusions: We recommend that quantitative, nonsurrogate, sensor-based adherence data be reported for all BioMeTs when feasible; a clear description of the sensor or sensors used to capture adherence data, the algorithm or algorithms that convert sample-level measurements to a metric of adherence, and the analytic validation data demonstrating that BioMeT-generated adherence is an accurate and reliable measurement of actual use be provided when available; and primary adherence data be reported as a continuous variable followed by categorical definitions if needed, and that the categories adopted are supported by clinical validation data and/or consistent with previous reports.

(J Med Internet Res 2022;24(4):e33537) doi:10.2196/33537

KEYWORDS

digital medicine; digital measures; adherence; compliance; mobile phone

Introduction

Background

Suboptimal adherence to clinical study procedures and/or a study intervention is often the root cause of a failed clinical trial, as it contributes to missing data; dilutes the effect of the intervention, thereby reducing statistical power; and may contribute to selection bias [1-3]. Compounding this issue is the fact that adherence is challenging to measure and account for during the analysis [4]. For example, self-reported medication adherence is straightforward to capture but is generally not a reliable measure of actual use; pill counts and pharmacy refill rates are imperfect surrogate measurements, and physical tests such as blood or urine biomarkers are expensive and difficult to administer throughout a trial [5-7]. Connected sensor technologies, such as mobile health and wearables, represent a potential solution for measuring adherence to a therapeutic intervention accurately, given that they are capable of collecting continuous, sensor-based data in real-world settings. Such technology has been increasingly used to capture trial end points [8,9]; therefore, in addition to monitoring adherence to an intervention, there exists an opportunity to monitor adherence to data collection procedures.

When conducting a clinical trial, the goal is to maximize adherence to both study procedures and interventions. However, efforts focused on increasing adherence cannot be developed until the determinants of adherence are understood. In turn, the reasons for suboptimal adherence cannot be identified unless adherence is adequately measured and reported in studies that use biometric monitoring technologies (BioMeTs), defined previously as connected digital tools that process data captured by mobile sensors using algorithms to generate measures of behavioral or physiological function [10]. A growing body of literature has offered standards and guidance to improve digital medicine study design and reporting quality [8-15]; however, best practices regarding measurement and reporting of BioMeT adherence—the extent to which the tool itself or an associated intervention is used as designed—are not as clearly conceptualized [4,16].

Objectives

A research working group from the Digital Medicine Society was formed to conduct a systematic literature review of published studies reporting adherence captured by BioMeTs to (1) identify studies that have used these tools to capture adherence to data collection procedures and/or study interventions, (2) describe the various methods used to measure adherence, and (3) compare the definitions of adherence reported in the literature. We view this description of the current state of the art as a critical first step toward identifying the determinants of adherence, in order to develop adjunct interventions to maximize adherence, ultimately contributing to improving the efficiency of clinical trials using novel technology.

Methods

Literature Search

The PubMed search terms were designed in five layers as follows: (1) used a BioMeT (layer A), (2) reported adherence or compliance (layer B), (3) were clinical studies (layer C), (4) reported original data (layer D), and (5) were published between January 1, 2014, and November 19, 2019 (layer E). Layers B, C, D, and E were based on indexing data available in PubMed, such as Medical Subject Headings terms and publication types. Layer A was designed to identify studies using a BioMeT and comprised 3 Medical Subject Headings terms as well as 34

keywords including *tracker*, *implantable*, *watch*, *mobile*, and *sensor* (see Multimedia Appendix 1 for complete search terms). When developing Layer A, our goal was to be sensitive rather than specific, as we anticipated variability in how BioMeTs are described in the literature.

We have adopted the term *adherence* rather than *compliance* or *concordance* throughout this manuscript, although we recognize that these terms cover a range of inconsistent definitions including patient-driven decisions and behaviors, passively conforming to medical advice, and the extent to which a research participant follows a study protocol [17]. Specifically, we included BioMeTs that measured the use of the tool itself, such as a wrist-worn device containing a skin capacitance sensor to monitor the duration of use, in addition to BioMeTs that measured the use of a diagnostic or therapeutic tool, such as a temperature sensor to measure the use of a dental appliance.

Systematic Review

We developed a Population, Intervention, Comparison, Outcomes, and Study design (PICOS) framework [18] to formulate the eligibility criteria for prospective studies of human participants (Textbox 1). Each study deployed at least one BioMeT outside the clinical setting or a testing facility, for which a quantitative, nonsurrogate, and sensor-based measurement of adherence was reported.

Textbox 1. Eligibility criteria adopted for literature screening in Population, Intervention, Comparison, Outcomes, and Study design order.

Eligibility criteria

Population

- Identify human studies
- Identify studies capturing in vivo data

Intervention

- Identify studies that used at least one biometric monitoring technology (BioMeT):
  - The tool must be used for purposes of measurement, diagnosis, and/or treatment of a behavioral or physiological function related to a disease state or physiological condition.
  - The tool must be mobile, meaning that it is capable of collecting data in real-world settings without oversight from trained personnel or staff.
  - The tool must be connected, meaning that there is a method to move data from the tool to the clinical or laboratory for analysis.
  - The tool must capture data via sensors of a physical property.
- Identify studies that captured BioMeT data outside of the clinical or laboratory setting.

Comparison

- Not applicable

Outcomes

- Identify studies that reported adherence:
  - The tool must measure adherence directly, rather than surrogate data associated with adherence.
  - The adherence data must be quantitative.
  - The adherence data must be sensor-based rather than based on self-report, observation, and/or based on manual adjustment or scoring.

Study design

- Identify studies reporting primary analyses of prospective data collection

Within the aforementioned definition of a BioMeT [10], the term *connected* was interpreted to include any wired or wireless transfer of data; thus, products that used a physical connection for data transfer were included, but devices that only displayed data on a user interface were excluded. Similarly, we interpreted the term *mobile* broadly and included wearables, proximal sensors, ingestibles, implantables, and tools that require a brief interaction such as a smartphone. When considering *sensors*, we included only those that measured physical properties such as temperature, pressure, sound, or acceleration. Therefore, we excluded tools that contained a chronometer that relied on being turned on or off or analyses based on self-report or other subjective assessments.

We stipulated that the measurement of adherence must be quantitative, nonsurrogate, and sensor-based. We defined *nonsurrogate* as an unequivocal reflection of product use. For example, technologies such as smart pill bottles, in which a sensor records the time at which the lid is removed, were considered a surrogate measure of adherence because they do not measure the ingestion of the pills. In contrast, smart pills that combine a pharmaceutical agent with a digital radio-frequency emitter activated by chloride ions in the

digestive system were considered as nonsurrogate adherence measurements, and therefore in scope. Finally, BioMeTs were excluded if the adherence data were based on self-report or observation or if any component of the adherence data required manual adjustment or scoring.

In total, 5 independent investigators (JPB, AC, DM, WM, and IMO) applied the PICOS criteria to a subset of 42 manuscripts for training purposes. The screening results were compared, discrepancies were discussed as a group, and the PICOS criteria were refined and clarified to optimize standardization during the remaining literature screening process. The remaining 898 manuscripts were then divided across 5 trained investigators for screening (LB, AC, WM, IMO, and BV), whereas a sixth investigator (JPB) independently screened a subset of 20% (180/898) of the manuscripts for quality assessment, as described below.

Data Extraction and Analysis

Data extraction fields included the study aim, study design (observational or interventional), therapeutic area, country of data collection, participant demographics (age, sex or gender, and race or ethnicity), information related to the BioMeT (concept of interest as described previously [19], technology type, sensor or sensors, device make and model, software name and version), and adherence data (end point definition and units). End point definitions were identified as the primary metric by which the sample-level data were analyzed to describe BioMeT adherence; for example, the duration of use, the percentage of tasks completed, or the percentage of study participants achieving a specified use goal. Therapeutic areas were categorized according to the Clinical Data Interchange Standards Consortium list [20], with additional categories for healthy and overweight or obesity. The sample size extracted from each manuscript was either the number of participants contributing to the adherence data or, if not reported, the total sample size for the study.

We categorized the primary definition of adherence for each BioMeT according to decreasing levels of data resolution as follows: (1) duration of use, either as a unit of time or percentage (continuous variable); (2) the number of measurements completed or number of days containing a measurement (continuous variable); or (3) the percentage of study participants who achieved a use goal (binary variable). Each BioMeT was categorized as passive (tools designed for continuous use) or active (tools that require user engagement at defined time points). Active BioMeTs were further categorized as session-based tools, such as connected exercise equipment, or task-based tools, such as a smart scale. This distinction was made because duration-based adherence data cannot be extracted from tools that measure one-off tasks; therefore, the highest-resolution adherence data available from these tools are the number of tasks or measurements completed.

All data were presented with descriptive statistics.

Results

Literature Screening Results

The PubMed search identified 940 manuscripts, of which 100 (10.6%) were deemed eligible for inclusion in the systematic review after meeting the PICOS criteria (Figure 1; see Multimedia Appendix 2 for all 100 papers listed). Data were extracted from these 100 manuscripts as described earlier.

After removing the date constraints from the PubMed search terms, we repeated the search by year to assess the number of publications captured from 1975 to 2020 (Figure 2).

Figure 1. Literature screening results per PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. Note that all papers were assessed for eligibility based on information contained in the abstract or full text. Papers were not screened based on the title only, as it was anticipated that many studies would include biometric monitoring technology (BioMeT) data as an exploratory end point and therefore, they would be not captured in the title.

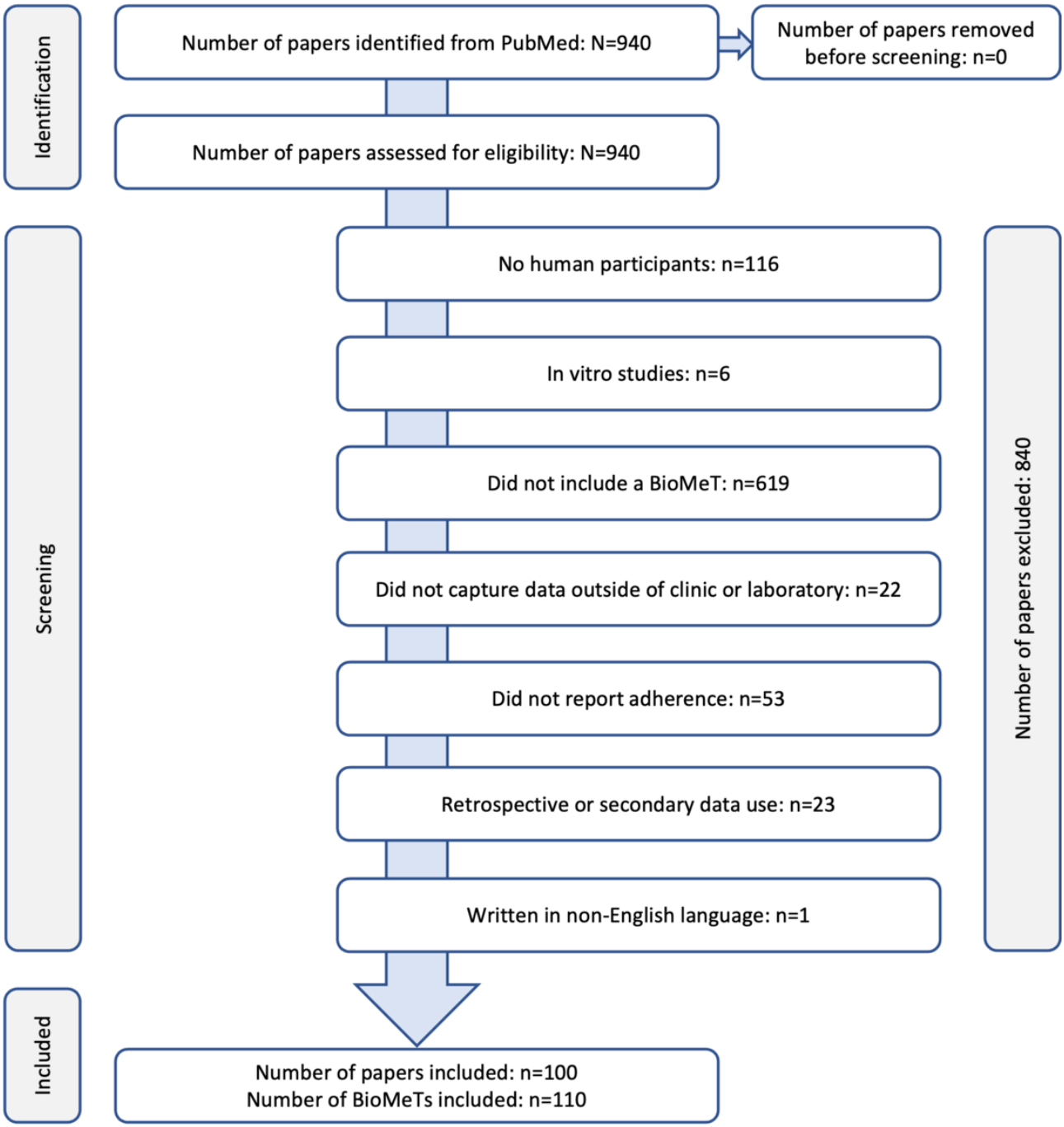
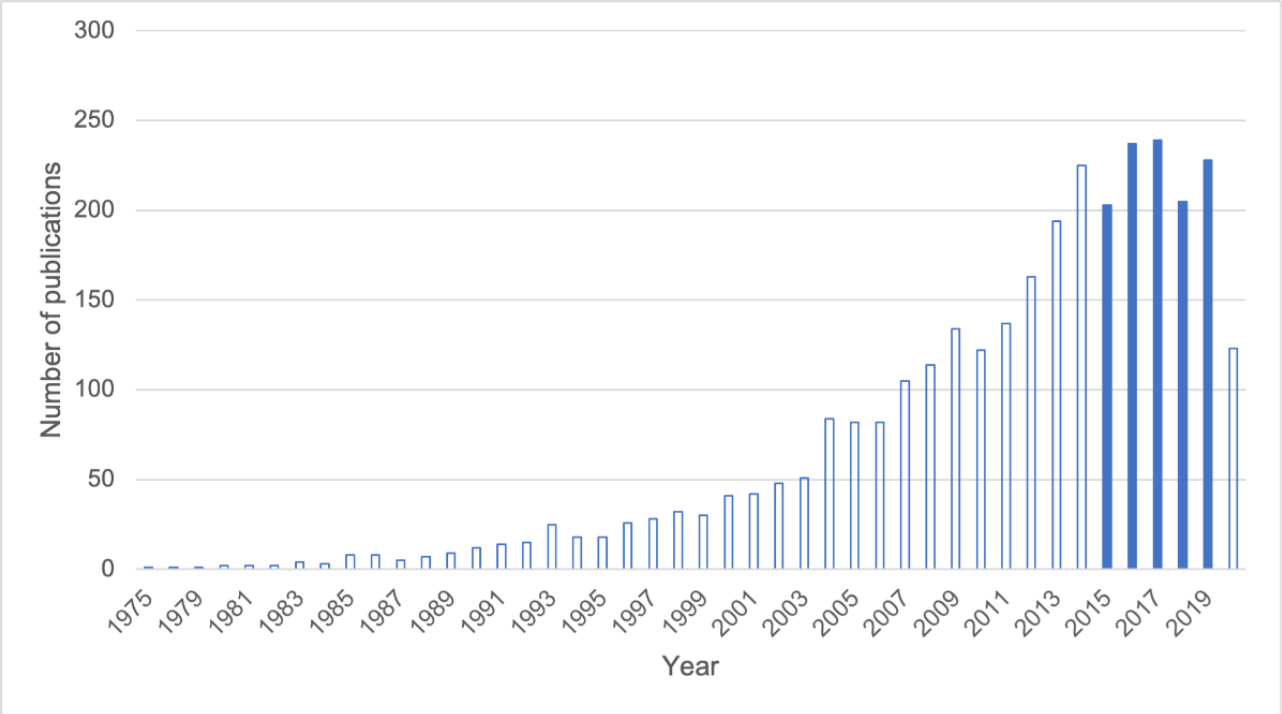


Figure 2. Number of publications captured by our literature search terms over time. The solid bars indicate the publications screened for inclusion in our systematic review.



Literature Screening Quality Assessment

The large number of manuscripts identified during the PubMed search (N=940) precluded the ability of more than one independent investigator to review each against our PICOS criteria. As described earlier, of the 898 manuscripts remaining after assessment of the subset of 42 manuscripts identified for training, we randomly identified 180 (20%) manuscripts to be rescreened by an independent investigator (JPB) for quality assessment. During this process, of the 180 manuscripts, there were 13 (7%) instances in which there was disagreement regarding the classification of the manuscripts. Most of these disagreements resulted from ambiguity of reporting with respect to whether the tool in question met the definition of a BioMeT (8/13, 62% of the manuscripts), whether the device measured adherence via a sensor (1/13, 8% of the manuscripts), or the setting of data collection (1/13, 8% of the manuscripts). Finally, 23% (3/13) of the disagreements were errors in which the manuscript was inadvertently marked as not including human participants or based on in vitro analyses. An additional 8 (1.11%) manuscripts that were not part of the audit subset were marked by the reviewer as ambiguous; all of these were cross-checked by another investigator to determine eligibility.

Descriptive Data

Table 1 summarizes the study design, sample size, and participant demographics of the 100 eligible studies. The sample size ranged from 10 to 128,037 participants, with an overall median of 60 participants (IQR 35-137). Most studies (92/100, 92%) used a single BioMeT; however, a second and third BioMeT was used in 6 (6%) and 2 (2%) manuscripts, respectively. Thus, 100 studies contributed data on 110 BioMeTs.

The manufacturer and/or model were reported for 90.0% (99/110) of the BioMeTs; however, only 30.9% (34/110) reported the software name and/or version. BioMeTs were categorized according to their concept of interest, with exercise or sleep (47/110, 42.7%) and sleep-disordered breathing (25/110, 22.7%) being the most common. These 2 categories also contained the highest number of BioMeT tool types; for example, exercise or sleep was captured by wearables, chest straps, smart clothing or footwear, and smartphones, whereas the sleep-disordered breathing category included positive airway pressure devices, chest straps for the treatment of positional obstructive sleep apnea, oral appliances, and implantable nerve stimulator devices. Notably, proximal sensors were considered in the scope, but none were captured in our systematic review.

Table 1. Study details, demographic data, and biometric monitoring technologies (BioMeTs) by therapeutic area.

	Therapeutic area of focus								
	All (N=100)	Healthy (n=11)	Cardiovascular (n=17)	Endocrine (n=13)	Neural (n=10)	Overweight or obesity (n=6)	Respiratory (n=29)	Pain treatments (n=5)	Other[a] (n=9)
Study design, n (%)									
Observational studies	26 (26)	3 (27)	1 (6)	4 (31)	3 (30)	1 (17)	9 (31)	1 (20)	4 (44)
Interventional studies	74 (74)	8 (73)	16 (94)	9 (69)	7 (70)	5 (83)	20 (69)	4 (80)	5 (56)
Sample size (participants), median; range	60; 10-128,037	179; 42-1381	84; 40-1732	46; 10-234	22; 10-780	86; 11-174	70; 10-128,037	35; 10-68	56; 20-281
Participant characteristics, n (%)									
Sex or gender									
Females or women only[b]	9 (9)	3 (27)	1 (6)	0 (0)	0 (0)	3 (50)	0 (0)	0 (0)	2 (22)
Both sexes or genders	84 (84)	7 (64)	16 (94)	12 (92)	8 (80)	3 (50)	26 (90)	5 (100)	7 (78)
Not reported	7 (7)	1 (9)	0 (0)	1 (8)	2 (20)	0 (0)	3 (10)	0 (0)	0 (0)
Age (years), n (%)									
≥60	24 (24)	4 (36)	4 (24)	5 (38)	2 (20)	0 (0)	5 (17)	1 (20)	3 (33)
>21 to <60	57 (57)	5 (45)	9 (53)	4 (31)	5 (50)	3 (50)	22 (76)[c]	3 (60)	6 (67)
≤21	19 (19)	2 (18)	4 (24)	4 (31)[c]	3 (30)	3 (50)	2 (7)	1 (20)	0 (0)
Race or ethnicity, n (%)									
Reported	39 (39)	7 (64)	11 (65)	6(46)	2 (20)	4 (67)	3 (10)	2 (40)	4 (44)
Not reported	61 (61)	4 (36)	6 (35)	7 (54)	8 (80)	2 (33)	26 (90)	3 (60)	5 (56)
BioMeT tool type, n, (%)									
Wearable	46 (42)	10 (91)	9 (41)	4 (27)	6 (60)	5 (83)	2 (7)	5 (83)	5 (50)
Positive airway pressure device	18 (16)	__[d]	—	—	—	—	18 (60)	—	—
Smart clothing	8 (7)	—	3 (14)	3 (20)	1 (10)	—	—	—	1 (10)
Blood pressure monitor	6 (5)	—	5 (23)	1 (7)	—	—	—	—	—
Chest strap	5 (5)	—	2 (9)	—	—	—	3 (10)	—	—
Smartphone	5 (5)	—	1 (5)	—	2 (20)	1 (17)	—	—	1 (10)
Oral appliance	5 (5)	—	—	—	—	—	2; 7%	—	3 (30)
Glucometer; continuous	3 (3)	—	—	3 (20)	—	—	—	—	—
Glucometer; noncontinuous	3 (3)	—	—	3 (20)	—	—	—	—	—
Ingestible	2 (2)	—	1 (5)	—	—	—	—	1 (17)	—
Implantable	2 (2)	—	—	—	—	—	2 (7)	—	—
Smart scale	2 (2)	—	1 (5)	1 (7)	—	—	—	—	—
Adhesive patch	1 (1)	1 (9)	—	—	—	—	—	—	—
Exercise equipment	1 (1)	—	—	—	—	—	1 (3)	—	—
Muscle trainer	1 (1)	—	—	—	—	—	1 (3)	—	—
Hearing aid	1 (1)	—	—	—	1 (10)	—	—	—	—
Home oxygen	1 (1)	—	—	—	—	—	1 (3)	—	—
Total	110	11	22	15	10	6	30	6	10

[a]Other category included oncology, gastrointestinal, bone structure, anatomy, or orthodontics, pregnancy, and vocal cord dysfunction.

[b]No studies included only males or men.

[c]Each of these categories contained 1 study that reported age only qualitatively or by providing a range; all other studies reported an average age.

[d]No studies falling into that category.

Adherence Data

Overall, we identified 37 unique definitions for the 110 BioMeTs. The most commonly reported definition (duration of use) was reported for 41.8% (46/110) of the tools; however, the next most common definitions (number or percentage of tasks completed and number or percentage of days with data) were reported for only 8.2% (9/110) and 6.4% (7/110) of the BioMeTs, respectively.

As shown in Table 2, each BioMeT was categorized as passive (69/110, 62.7%), session-based (24/110, 21.8%), or task-based (17/110, 15.5%). The duration of use was reported for 46% (32/69) of the passive BioMeTs and 75% (18/24) of the session-based BioMeTs. Of the task-based BioMeTs for which the duration of use could not be meaningfully reported, the highest resolution of adherence data (the number of measurements or days) was reported for 41% (7/17) of the BioMeTs. The lowest resolution of adherence data (achievement of a goal as a binary or categorical variable) was reported for 33.6% (37/110) of all BioMeTs.

Table 2. Adherence data resolution and definition captured by passive and active biometric monitoring technologies (BioMeTs).

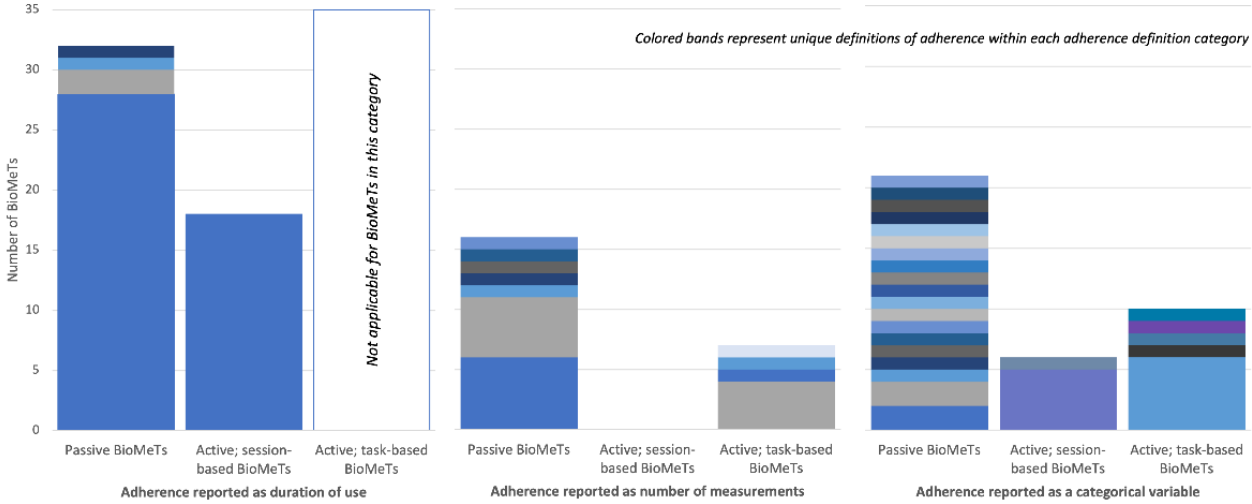
Parameters	Highest resolution adherence data				Lowest resolution adherence data	
	Duration of use (based on a continuous variable)		Number of measurements or days used (based on a continuous variable)		Achievement of a goal (based on a binary variable)	
	Number of BioMeTs, n (%)	Number of unique adherence definitions	Number of BioMeTs, n (%)	Number of unique adherence definitions	Number of BioMeTs, n (%)	Number of unique adherence definitions
Monitoring type						
Passive	32 (64)	4	16 (70)	7	21 (57)	19
Active; session-based	18 (36)	1	0 (0)	0	6 (16)	2
Active; task-based	N/A[a]	N/A	7 (30)	4	10 (27)	5
All BioMeTs	50 (100)	4[b]	23 (100)	8[b]	37 (100)	25[b]
All BioMeTs apart from sleep-disordered breathing	30 (60)	4	23 (100)	8	32 (86)	24

[a]N/A: not applicable.

[b]These data are not simply the sum of the rows above, as there were instances where the same adherence definition was adopted for different tool types.

As shown in Figure 3, the number of unique definitions of adherence increased as the resolution of the reported data decreased. For example, adherence was reported as the duration of use (highest resolution) for 50 BioMeTs, 46 (92%) of which reported an actual unit of time, along with 3 other unique definitions of adherence, such as the duration of use in which a heart rate goal was achieved. In contrast, among the 37 BioMeTs for which adherence was reported as a categorical variable (lowest resolution), 25 unique definitions of adherence were identified. Sleep-disordered breathing was the BioMeT category with the most consistent definitions of adherence (2 definitions reported for 25 BioMeTs); however, the pattern of decreasing uniformity score alongside decreasing data resolution persisted after removal of these BioMeTs (Table 2).

Figure 3. Uniformity of adherence definitions according to whether the biometric monitoring technology (BioMeT) was a passive, session-based, or task-based tool. Passive BioMeTs are those designed for continuous use. Active BioMeTs are those that require user engagement at defined time points, further categorized as session-based (for which duration of use is meaningful) versus task-based (for which the duration of use is not meaningful). The colored bands represent unique definitions of adherence within each bar. The colors are comparable across the bars within each category of adherence definition (duration of use, number of measurements, and categorical variables).



Discussion

Principal Findings

The purpose of this review was to describe the various approaches taken to evaluate adherence to study procedures and/or interventions using BioMeTs in recent clinical studies and discuss best practices that can improve the reliability and comparability of adherence measurements to support further BioMeT evaluation and decision-making in both research and clinical care settings. Notably, we found that 29.9% (53/177) of the studies that used a BioMeT outside the clinical or laboratory setting failed to report a sensor-based, nonsurrogate, quantitative measurement of adherence, thus impeding a complete understanding of the study data. Among the 100 studies that reported sensor-based adherence data, we found substantial variability in terms of the definitions of adherence adopted, and that the degree of variability was associated with the resolution of the data reported. For example, when adherence was reported as a continuous time variable, the same definition of adherence was adopted for 92% (46/50) of the tools. However, when the adherence data were simplified to a categorical variable, we observed 25 unique definitions of adherence reported for 37 tools, and the most common definition was adopted for only 19% (7/37) of the BioMeTs. Examples of adherence definitions that were reported only once each within our data set include the percentage of participants with use <85% of the total time (passive BioMeT), the percentage of participants with use of ≥4 hours on ≥70 days (active, session-based BioMeT), and the percentage of participants completing readings on 100% of days (active, task-based BioMeT). All 3 of these adherence definitions were relevant and useful for the study in question; however, by adopting a specific threshold and reporting adherence as the percentage of the sample that achieved the goal, the adherence data were not readily interpretable against other studies. If adherence data were provided as higher-resolution variables, such as duration of use or number of readings, readers would be better positioned to make comparisons. Considering that adherence to any given procedure or intervention is a critical driver of desired behavior change and improved health outcomes in research and real-world settings [21], greater consistency in defining adherence may help more clearly associate BioMeT adherence to study outcomes and may offer a critical lens in the design and implementation of customized BioMeTs that are fit-for-purpose within their context of use.

In addition to consistent reporting of adherence, it is critical to understand exactly what digital medicine tools are used in a given study, consistent with the 2021 EVIDENCE (Evaluating Connected Sensor Technologies) Publication Checklist [11]. A complete description of the tool used ensures reproducibility, allows for meaningful comparisons across studies, and opens up the possibility of merging data across cohorts. Although the manufacturer or model (or both) were reported for 90.0% (99/110) of the tools captured in our review, this still leaves 10.0% (11/110) for which the tools were described in generic terms. Moreover, we found that the software name or version (or both) used for data processing was reported for only 30.9% (34/110) of the BioMeTs, indicating that the data cannot be reproduced even if the hardware details are known. We also noted key gaps in the descriptive data; most notably, only 41% (41/100) reported the ethnicity or race of participants, which is a persistent problem in clinical research [22,23]. Even among the 46 studies performed in the United States, for which there are clear guidelines for collecting and reporting race and ethnicity [24], 30% (14/46) of the studies did not provide these data. Age, sex or gender, and race or ethnicity are essential for understanding the representativeness of study samples and generalizability of findings and reflect only a subset of a broader set of characteristics such as socioeconomic information that must be captured and reported to understand how BioMeT adherence relates to issues of access, uptake, equity, and equality [25,26].

Differences across studies using BioMeTs are inevitable; however, ideally, there should be standardization and harmonization of collecting and reporting adherence to allow for the evaluation, interpretation, and statistical comparison of outcomes. Thus, although this study was not designed to develop standards, the aforementioned gaps and shortcomings have led us to recommend minimum reporting requirements and that consistent definitions or units for adherence should be adopted. Specifically, we recommend that (1) quantitative, nonsurrogate, sensor-based adherence data be reported for all BioMeTs when feasible; (2) a clear description of the sensor or sensors used to capture adherence data, the algorithm or algorithms that convert sample-level measurements to a metric of adherence, and the analytic validation data demonstrating that BioMeT-generated adherence is an accurate and reliable measurement of actual use be provided when available; and (3) primary adherence data be reported as a continuous variable followed by categorical definitions if needed, and that the categories adopted are supported by clinical validation data or consistent with previous reports (or both). These recommendations are in addition to the minimum requirements recommended elsewhere, such as providing a description of verification, validation, and usability data explaining the fit-for-purpose characteristics of the BioMeT technology used within a specific context as well as detailed demographic and descriptive data for the study sample [10,11,15]. More detailed descriptions of our recommendations for reporting BioMeT adherence are provided in Table 3, which includes a case study that we identified as an exemplar that followed all included recommendations [27].

On a positive note, it is clear that BioMeTs have been increasingly deployed in clinical research studies. Repeating our PubMed search over successive years revealed that the number of publications captured in our search terms increased steadily between 1975 and 2005 and became increasingly prevalent until 2015. The reduced number of papers captured in 2020 may reflect a delay in PubMed indexing and possibly a reduced submission and/or acceptance rate during the initial stages of the COVID-19 pandemic. It is encouraging to observe adherence data reported from a wide range of monitoring, diagnostic, and therapeutic tools from studies conducted in 22 different countries. Furthermore, although accelerometry-based tools for estimating sleep and activity have been in use for several decades [28], our literature search captured adherence data for more recently developed tools, such as automated speech assessments [29] and upper-limb training systems for motor disorders [30].

Table 3. Recommendations for capturing and reporting adherence measured by biometric monitoring technologies (BioMeTs).

Identified gaps and recommendations	Case study [27]
Gap 1: Quantitative, nonsurrogate, sensor-based adherence data were not reported in 29.9% of screened manuscripts that captured BioMeT data outside the clinical or laboratory setting.	
Recommendation 1: Investigators are encouraged to develop and/or use BioMeT sensors to capture sensor-based adherence data in addition to their primary purpose.	This study aimed to evaluate adherence to a physical activity among students recruited from 20 schools. Quantitative adherence data were derived from wrist-worn accelerometers, considered a direct reflection of wear-time.
Recommendation 2: Where feasible, we encourage investigators to collect and report adherence data that are a direct reflection of actual use, rather than a surrogate.	N/A[a]
Gap 2: BioMeT manufacturer or model and software information was missing for 10% and 68% of included tools, respectively.	
Recommendation 3: In addition to reporting the BioMeT manufacturer or model and software used for generating adherence data (where applicable), we recommend that investigators provide a clear description of the sensor or sensors capturing adherence data.	BioMeT model: GENEActiv wrist-worn device (ActivInsights Ltd). Sensor description: 3-axis accelerometer. Software: GENEActiv PC software (version 2.9), with subsequent signal processing performed in R-package (GGIR; version 1.2-2).
Recommendation 4: We recommend that investigators describe the algorithm or algorithms that convert sample-level measurements into a measurement of adherence. If a description is not available from the manufacturer, this should be stated.	The paper included the data sampling frequency (100 Hz); a description of the signal processing steps including calibration; the epoch length (5 seconds) over which the sample-level data were averaged; and the units (milligravitational units; m g). A description of the nonwear detection algorithm was summarized as, “Non-wear is estimated on the basis of the SD and value range of each axis, calculated for 60-min windows with 15-min sliding window. The window is classified as non-wear if, for at least two of the three axes, the SD is less than 13 mg or the value range is less than 50 mg.”
Recommendation 5: We recommend that investigators describe the analytic validation data supporting the adherence algorithm; that is, the data indicating that adherence per the BioMeT is an accurate estimate of actual use. If analytic validation data is not available, this should be stated.	A reference to previous verification and analytic validation work was included.
Gap 3: Heterogeneity of adherence definitions increased alongside decreasing resolution of adherence data reported.	
Recommendation 6: We recommend that investigators using BioMeTs that are either passive (designed to capture data passively over long periods) or session-based (designed for user engagement at certain time points, for which the duration of use is meaningful) report primary adherence as a continuous variable of time; that is, total minutes or hours or days, or average hours per day, days per week, and so on. Example of a passive BioMeT: smart clothing. Example of a session-based BioMeT: connected exercise equipment.	The BioMeT was categorized as passive, as the wrist-worn accelerometer was designed to capture data continuously over 3 separate periods of 7 days. Adherence was reported as the total hours of wear-time, and hours per day of wear-time.
Recommendation 7: We recommend that investigators using BioMeTs that are task based (designed for user engagement at certain time points, for which the duration of use is not meaningful) report primary adherence as a continuous variable; that is, the number of tasks or days completed. Example of a task-based BioMeT: connected scale.	N/A, as the BioMeT was categorized as passive rather than task based.
Recommendation 8: We recommend that categorical adherence data are reported only in addition to continuous adherence data; for example, the percentage of participants with use >x hours per day or percentage of participants completing >y tasks.	Categorical adherence data included the number of participants with ≥16 hours of wear-time per day.
Recommendation 9: We recommend that categorical definitions of adherence be based on clinical validation data indicating the level of adherence associated with a clinically meaningful change in the outcome of interest, when available. If clinical validation data are not available, this should be stated.	The investigators include a reference to previous work that adopted the threshold of ≥16 hours of wear-time per day and describe another study that compared thresholds of 8 hours, 16 hours, and 24 hours of wear-time.

[a]N/A: not applicable.

Strengths and Limitations

To our knowledge, this is the first systematic review to focus specifically on BioMeT adherence. Further strengths of our study include the large number of manuscripts screened for potential inclusion, and the quality control processes that we implemented, which resulted in few disagreements among the reviewers. The sensitive, rather than specific, search terms we adopted increased our confidence that we were able to capture the relevant set of literature, given our initial concern that many BioMeTs were used for exploratory analyses and therefore not referred to in study titles, abstracts, or keywords. The quality

control process was particularly important, given that 76.4% (718/940) of the manuscripts were screened by a single investigator. Alongside these strengths, this review has several limitations that should be noted. Owing to the inconsistencies in the study outcome measures and variability in the definition of adherence adopted across studies, we did not undertake a methodological assessment and could not determine statistical inference. We also did not extract every element of the study design, such as the duration of the study itself or the length of time the BioMeTs were used per protocol. We included only peer reviewed publications indexed in PubMed; therefore, our findings may not be representative of all studies capturing BioMeT data in related fields, such as engineering. Finally, due to the vast number of manuscripts captured in our PubMed search terms, we limited the time frame to a 5-year period, and we limited our review to studies reporting nonsurrogate measurements of adherence, thereby excluding technology such as smart pill bottles which may offer valuable adherence data where a nonsurrogate measurement is not feasible.

Conclusions

This review provides a description of the numerous methods that have been used in recent years to measure BioMeT adherence, allowing us to identify gaps and make specific reporting recommendations. Several important questions remain, which we hope will be addressed in future studies. For example, it will be interesting to compare our findings to a similar review covering the subsequent 5-year period (2020-2025), as the abrupt acceleration of digital monitoring and interventions including telemedicine during the current COVID-19 era [31,32] will likely, in hindsight, be considered a paradigm shift in both research and health care delivery. We hope that with increased consistency and reporting of data elements, it will become possible to meta-analyze adherence data to identify the possible determinants of BioMeT use patterns. Only when adherence data are adequately reported will the field of digital medicine be able to advance our understanding of the reasons underlying acceptance and adherence, which will ultimately allow investigators to optimize the design of tools, studies, implementation methods, and user engagement strategies to realize the full potential of BioMeTs as digital monitoring, diagnostic, and therapeutic tools. To support these actions, we recommend that (1) quantitative, nonsurrogate, sensor-based adherence data be reported for all BioMeTs when feasible; (2) a clear description of the sensor or sensors used to capture adherence data, the algorithm or algorithms that convert sample-level measurements to a metric of adherence, and the analytic validation data demonstrating that BioMeT-generated adherence is an accurate and reliable measurement of actual use be provided when available; and (3) primary adherence data be reported as a continuous variable followed by categorical definitions if needed, and that the categories adopted are supported by clinical validation data and/or consistent with previous reports.

Acknowledgments

This publication is a result of collaborative research performed under the auspices of the Digital Medicine Society (DiMe). DiMe is a 510(c)(3) nonprofit professional society for the digital medicine community and is not considered a sponsor of this work. All authors are members of DiMe, who volunteered to participate in this systematic review. All DiMe research activities are overseen by a research committee, the members of which were invited to comment on the manuscript before submission.

Conflicts of Interest

The authors have reported the following conflicts of interest through employment or stock ownership: JPB (Signifier Medical Technologies and Philips), LB (Novartis), RC (Verily Life Sciences), AC (Pfizer and Ali Ciger Ventures UG [haftungsbeschränkt]), KLF (K Health, Trusst Health Inc, InquistHealth, and Social Wellness), ESI (Koneksa Health), CJM (Astra Zeneca and Abbvie), CAN (Pfizer), IRRC (ICON plc), and BV (Byteflies).

Multimedia Appendix 1
PubMed search terms.
[DOCX File , 100 KB - jmir_v24i4e33537_app1.docx]

Multimedia Appendix 2
All manuscripts identified for data extraction.
[DOCX File , 124 KB - jmir_v24i4e33537_app2.docx]

References

1. Breckenridge A, Aronson JK, Blaschke TF, Hartman D, Peck CC, Vrijens B. Poor medication adherence in clinical trials: consequences and solutions. Nat Rev Drug Discov 2017;16(3):149-150. [doi: 10.1038/nrd.2017.1] [Medline: 28154411]
2. Shiovitz TM, Bain EE, McCann DJ, Skolnick P, Laughren T, Hanina A, et al. Mitigating the effects of nonadherence in clinical trials. J Clin Pharmacol 2016;56(9):1151-1164 [FREE Full text] [doi: 10.1002/jcph.689] [Medline: 26634893]
3. Tripepi G, Jager KJ, Dekker FW, Zoccali C. Selection bias and information bias in clinical research. Nephron Clin Pract 2010;115(2):c94-c99 [FREE Full text] [doi: 10.1159/000312871] [Medline: 20407272]

4. Roe D, Jones N, Hasson-Ohayon I, Zisman-Ilani Y. Conceptualization and study of antipsychotic medication use: from adherence to patterns of use. Psychiatr Serv 2021;72(12):1464-1466. [doi: 10.1176/appi.ps.202100006] [Medline: 34126781]
5. Valdés Y Llorca C, Cortés-Castell E, Ribera-Casado JM, de Lucas-Ramos P, de Palacio-Guerrero LM, Mugarza-Borqué F, et al. Validation of self-reported adherence in chronic patients visiting pharmacies and factors associated with the overestimation and underestimation of good adherence. Eur J Clin Pharmacol 2020;76(11):1607-1614. [doi: 10.1007/s00228-020-02950-9] [Medline: 32613537]
6. Anghel LA, Farcas AM, Oprean RN. An overview of the common methods used to measure treatment adherence. Med Pharm Rep 2019;92(2):117-122 [FREE Full text] [doi: 10.15386/mpr-1201] [Medline: 31086837]
7. 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;5(4):470-482 [FREE Full text] [doi: 10.1007/s13142-015-0315-2] [Medline: 26622919]
8. Perry B, Herrington W, Goldsack JC, Grandinetti CA, Vasisht KP, Landray MJ, et al. Use of mobile devices to measure outcomes in clinical research, 2010-2016: a systematic literature review. Digit Biomark 2018;2(1):11-30 [FREE Full text] [doi: 10.1159/000486347] [Medline: 29938250]
9. Walton MK, Cappelleri JC, Byrom B, Goldsack JC, Eremenco S, Harris D, et al. Considerations for development of an evidence dossier to support the use of mobile sensor technology for clinical outcome assessments in clinical trials. Contemp Clin Trials 2020;91:105962 [FREE Full text] [doi: 10.1016/j.cct.2020.105962] [Medline: 32087341]
10. Goldsack JC, Coravos A, Bakker JP, Bent B, Dowling AV, Fitzer-Attas C, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). NPJ Digit Med 2020;3:55 [FREE Full text] [doi: 10.1038/s41746-020-0260-4] [Medline: 32337371]
11. Manta C, Mahadevan N, Bakker J, Ozen Irmak S, Izmailova E, Park S, et al. EVIDENCE publication checklist for studies evaluating connected sensor technologies: explanation and elaboration. Digit Biomark 2021;5(2):127-147 [FREE Full text] [doi: 10.1159/000515835] [Medline: 34179682]
12. Godfrey A, Goldsack JC, Tenaerts P, Coravos A, Aranda C, Hussain A, et al. BioMeT and algorithm challenges: a proposed digital standardized evaluation framework. IEEE J Transl Eng Health Med 2020;8:0700108 [FREE Full text] [doi: 10.1109/JTEHM.2020.2996761] [Medline: 32542118]
13. Coravos A, Doerr M, Goldsack J, Manta C, Shervey M, Woods B, et al. Modernizing and designing evaluation frameworks for connected sensor technologies in medicine. NPJ Digit Med 2020;3:37 [FREE Full text] [doi: 10.1038/s41746-020-0237-3] [Medline: 32195372]
14. Goldsack JC, Dowling AV, Samuelson D, Patrick-Lake B, Clay I. Evaluation, acceptance, and qualification of digital measures: from proof of concept to endpoint. Digit Biomark 2021;5(1):53-64 [FREE Full text] [doi: 10.1159/000514730] [Medline: 33977218]
15. Coran P, Goldsack JC, Grandinetti CA, Bakker JP, Bolognese M, Dorsey ER, et al. Advancing the use of mobile technologies in clinical trials: recommendations from the clinical trials transformation initiative. Digit Biomark 2019;3(3):145-154 [FREE Full text] [doi: 10.1159/000503957] [Medline: 32095773]
16. Klonoff DC. Behavioral theory: the missing ingredient for digital health tools to change behavior and increase adherence. J Diabetes Sci Technol 2019;13(2):276-281 [FREE Full text] [doi: 10.1177/1932296818820303] [Medline: 30678472]
17. Chakrabarti S. What's in a name? Compliance, adherence and concordance in chronic psychiatric disorders. World J Psychiatry 2014;4(2):30-36 [FREE Full text] [doi: 10.5498/wjp.v4.i2.30] [Medline: 25019054]
18. Tacconelli E. Systematic reviews: CRD's guidance for undertaking reviews in health care. Lancet Infect Dis 2010;10(4):226. [doi: 10.1016/s1473-3099(10)70065-7]
19. Bakker JP, Goldsack JC, Clarke M, Coravos A, Geoghegan C, Godfrey A, et al. A systematic review of feasibility studies promoting the use of mobile technologies in clinical research. NPJ Digit Med 2019;2:47 [FREE Full text] [doi: 10.1038/s41746-019-0125-x] [Medline: 31304393]
20. Therapeutic areas. Clinical Data Interchange Standards Consortium. URL: https://www.cdisc.org/standards/therapeutic-areas [accessed 2021-08-11]
21. Brown MT, Bussell JK. Medication adherence: WHO cares? Mayo Clin Proc 2011;86(4):304-314 [FREE Full text] [doi: 10.4065/mcp.2010.0575] [Medline: 21389250]
22. Rochon PA, Mashari A, Cohen A, Misra A, Laxer D, Streiner DL, et al. The inclusion of minority groups in clinical trials: problems of under representation and under reporting of data. Account Res 2004;11(3-4):215-223. [doi: 10.1080/08989620490891412] [Medline: 15812967]
23. Medina LD, Torres S, Gioia A, Ochoa Lopez A, Wang J, Cirino PT. Reporting of demographic variables in neuropsychological research: an update of O'Bryant et al.'s trends in the current literature. J Int Neuropsychol Soc 2021;27(5):497-507. [doi: 10.1017/S1355617720001083] [Medline: 33176898]
24. Evaluation and reporting of age-, race-, and ethnicity-specific data in medical device clinical studies: guidance for industry and food and drug administration staff. U.S. Food & Drug Administration. 2017. URL: https://www.fda.gov/media/98686/download [accessed 2022-03-15]

25. Kontos E, Blake KD, Chou WY, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
26. Azzopardi-Muscat N, Sørensen K. Towards an equitable digital public health era: promoting equity through a health literacy perspective. Eur J Public Health 2019;29(Supplement_3):13-17 [FREE Full text] [doi: 10.1093/eurpub/ckz166] [Medline: 31738443]
27. Rowlands AV, Harrington DM, Bodicoat DH, Davies MJ, Sherar LB, Gorely T, et al. Compliance of adolescent girls to repeated deployments of wrist-worn accelerometers. Med Sci Sports Exerc 2018;50(7):1508-1517. [doi: 10.1249/MSS.0000000000001588] [Medline: 29474208]
28. Martin JL, Hakim AD. Wrist actigraphy. Chest 2011;139(6):1514-1527 [FREE Full text] [doi: 10.1378/chest.10-1872] [Medline: 21652563]
29. Whitling S, Lyberg-Åhlander V, Rydell R. Absolute or relative voice rest after phonosurgery: a blind randomized prospective clinical trial. Logoped Phoniatr Vocol 2018;43(4):143-154. [doi: 10.1080/14015439.2018.1504985] [Medline: 30183437]
30. Gerber CN, Kunz B, van Hedel HJ. Preparing a neuropediatric upper limb exergame rehabilitation system for home-use: a feasibility study. J Neuroeng Rehabil 2016;13:33 [FREE Full text] [doi: 10.1186/s12984-016-0141-x] [Medline: 27008504]
31. Seshadri DR, Davies EV, Harlow ER, Hsu JJ, Knighton SC, Walker TA, et al. Wearable sensors for COVID-19: a call to action to harness our digital infrastructure for remote patient monitoring and virtual assessments. Front Digit Health 2020;2:8 [FREE Full text] [doi: 10.3389/fdgth.2020.00008] [Medline: 34713021]
32. Goldsack JC, Izmailova ES, Menetski JP, Hoffmann SC, Groenen PM, Wagner JA. Remote digital monitoring in clinical trials in the time of COVID-19. Nat Rev Drug Discov 2020;19(6):378-379. [doi: 10.1038/d41573-020-00094-0] [Medline: 32409759]

Abbreviations

BioMeT: biometric monitoring technology
DiMe: Digital Medicine Society
PICOS: Population, Intervention, Comparison, Outcomes, and Study design

Edited by A Mavragani; submitted 11.09.21; peer-reviewed by B Peterson, C Drummond, M Nissen; comments to author 20.10.21; revised version received 03.11.21; accepted 14.01.22; published 14.04.22.

Please cite as:
Olaye IM, Belovsky MP, Bataille L, Cheng R, Ciger A, Fortuna KL, Izmailova ES, McCall D, Miller CJ, Muehlhausen W, Northcott CA, Rodriguez-Chavez IR, Pratap A, Vandendriessche B, Zisman-Ilani Y, Bakker JP
Recommendations for Defining and Reporting Adherence Measured by Biometric Monitoring Technologies: Systematic Review
J Med Internet Res 2022;24(4):e33537
URL: https://www.jmir.org/2022/4/e33537
doi:10.2196/33537
PMID:35436221

Review

Contributions of Artificial Intelligence Reported in Obstetrics and Gynecology Journals: Systematic Review

Ferdinand Dhombres[1,2], MD, PhD; Jules Bonnard[3], MSc; Kévin Bailly[3], PhD; Paul Maurice[1], MD, MSc; Aris T Papageorghiou[4], MD, PhD; Jean-Marie Jouannic[1,2], MD, PhD

[1]Fetal Medicine Department, Armand Trousseau University Hospital, Sorbonne University, Paris, France

[2]Laboratory in Medical Informatics and Knowledge Engineering in e-Health, Institut National de la Santé et de la Recherche Médicale, Sorbonne University, Paris, France

[3]Institute for Intelligent Systems and Robotics, Sorbonne University, Paris, France

[4]Oxford Maternal & Perinatal Health Institute, Green Templeton College, Oxford, United Kingdom

Corresponding Author:
Ferdinand Dhombres, MD, PhD
Fetal Medicine Department
Armand Trousseau University Hospital
Sorbonne University
26 Avenue du Dr Arnold Netter
Paris, 75012
France
Phone: 33 4473 5117
Email: ferdinand.dhombres@inserm.fr

Abstract

Background: The applications of artificial intelligence (AI) processes have grown significantly in all medical disciplines during the last decades. Two main types of AI have been applied in medicine: symbolic AI (eg, knowledge base and ontologies) and nonsymbolic AI (eg, machine learning and artificial neural networks). Consequently, AI has also been applied across most obstetrics and gynecology (OB/GYN) domains, including general obstetrics, gynecology surgery, fetal ultrasound, and assisted reproductive medicine, among others.

Objective: The aim of this study was to provide a systematic review to establish the actual contributions of AI reported in OB/GYN discipline journals.

Methods: The PubMed database was searched for citations indexed with "artificial intelligence" and at least one of the following medical subject heading (MeSH) terms between January 1, 2000, and April 30, 2020: "obstetrics"; "gynecology"; "reproductive techniques, assisted"; or "pregnancy." All publications in OB/GYN core disciplines journals were considered. The selection of journals was based on disciplines defined in Web of Science. The publications were excluded if no AI process was used in the study. Review, editorial, and commentary articles were also excluded. The study analysis comprised (1) classification of publications into OB/GYN domains, (2) description of AI methods, (3) description of AI algorithms, (4) description of data sets, (5) description of AI contributions, and (6) description of the validation of the AI process.

Results: The PubMed search retrieved 579 citations and 66 publications met the selection criteria. All OB/GYN subdomains were covered: obstetrics (41%, 27/66), gynecology (3%, 2/66), assisted reproductive medicine (33%, 22/66), early pregnancy (2%, 1/66), and fetal medicine (21%, 14/66). Both machine learning methods (39/66) and knowledge base methods (25/66) were represented. Machine learning used imaging, numerical, and clinical data sets. Knowledge base methods used mostly omics data sets. The actual contributions of AI were method/algorithm development (53%, 35/66), hypothesis generation (42%, 28/66), or software development (3%, 2/66). Validation was performed on one data set (86%, 57/66) and no external validation was reported. We observed a general rising trend in publications related to AI in OB/GYN over the last two decades. Most of these publications (82%, 54/66) remain out of the scope of the usual OB/GYN journals.

Conclusions: In OB/GYN discipline journals, mostly preliminary work (eg, proof-of-concept algorithm or method) in AI applied to this discipline is reported and clinical validation remains an unmet prerequisite. Improvement driven by new AI research guidelines is expected. However, these guidelines are covering only a part of AI approaches (nonsymbolic) reported in this review; hence, updates need to be considered.

(J Med Internet Res 2022;24(4):e35465) doi:10.2196/35465

KEYWORDS

artificial intelligence; systematic review; knowledge bases; machine learning; obstetrics; gynaecology; perinatology; medical informatics

Introduction

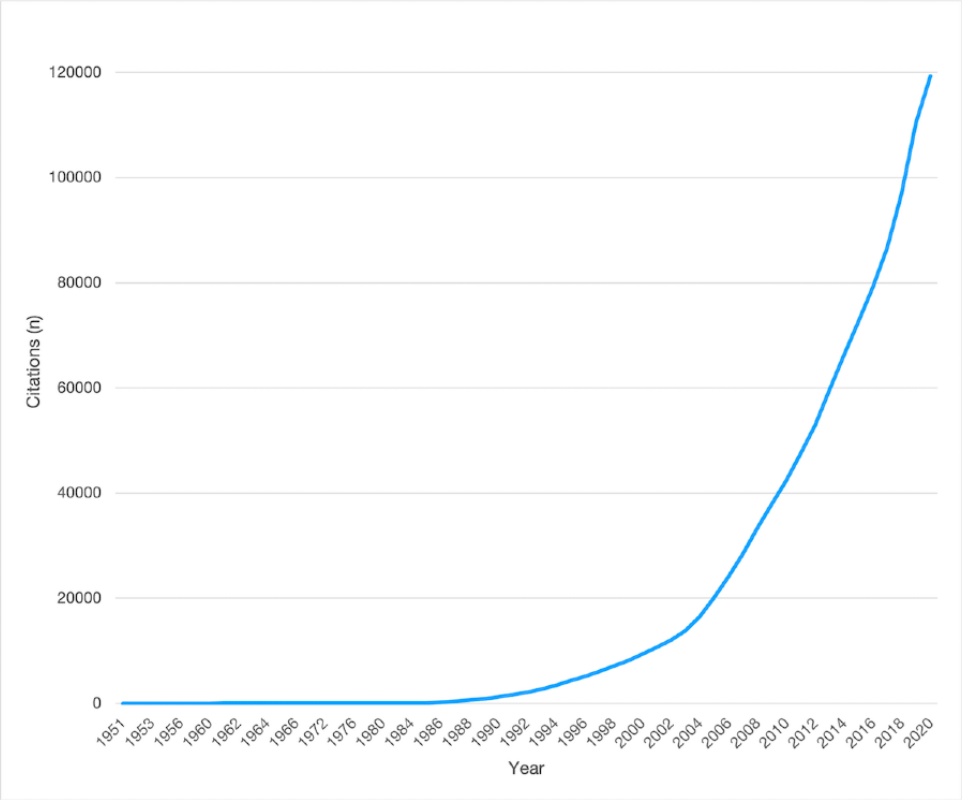
The foundations of artificial intelligence (AI) as a discipline were established in the 1950s, under the hypothesis formulated by John McCarthy as "Every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it" [1]. Developing AI was a 3-fold challenge: collect an unprecedented amount of data for training and validation of algorithms, build computers with sufficient computational power, and create algorithms to simulate human intelligence functions (eg, reasoning, learning, adaptation, vision, interaction).

At the dawn of the 21st century, all 3 challenges have been taken up in many fields, leveraging different types of AI approaches. Two general directions in AI research approaches have been pursued: symbolic approaches and nonsymbolic approaches. The symbolic AI approach, also known as "Good Old-Fashioned AI" (GOFAI) [2], encompasses formal logic, knowledge representation, and rule-based and semantic reasoning. These approaches are generally explainable and human-readable, need human curation and design, and do not rely on a large amount of data to develop. The first GOFAI-related publications in the field of medicine emerged 60 years ago [3], and these approaches provided the first significant results with expert systems [4,5] and are now widely used in knowledge-based systems [6,7], mostly through the application of ontologies and semantic web technologies [8-10]. Nonsymbolic AI, defined as intelligence without specific knowledge representations, encompasses various approaches to simulate other human intelligence processes such as learning, perception, and pattern recognition. Machine learning (ML) has become the main approach in this area [11], mostly through algorithms such as artificial neural networks (ANNs), and relies on a large amount of high-quality data to learn, train, and validate, along with significant computational resources. This AI is generally "nonexplainable," with the process occurring inside ANNs (architecture, layers, and connections) remaining in the form of a "black box" to the users.

Publications in AI in medicine have grown rapidly during the last two decades: 119,325 citations are referenced in PubMed, 93% of which have been published since 2000 (Figure 1). The obstetrics and gynecology (OB/GYN) domain represents a wide range of medical activities (obstetrics, fetal medicine, open and endoscopic surgery, ultrasound and other imaging modalities, reproductive biology, and assisted reproductive technologies). These activities are leveraging various types of data (eg, textual data; 2D, 3D, and 4D imaging data; genomic and proteomic data; fetal monitoring data). However, it is only recently that AI concepts (ML principles) were described in an OB/GYN ultrasound imaging journal [12]. Interestingly, the general emergence of AI in the OB/GYN domain, and more specifically in OB/GYN journals, has not been investigated.

Our aim was to establish the actual contributions of AI reported in OB/GYN journals with a systematic review to investigate, within all OB/GYN subdomains, the AI methods, sources of data, and the contribution and validation of the AI processes. Most of the recent publications about AI usually focus on ML, ANNs, and deep learning. In this review, we considered all AI contributions, including both symbolic and nonsymbolic AI.

Figure 1. Trend of the 119,325 citations in PubMed indexed with the MeSH (Medical Subject Heading) term "artificial intelligence" between 1951 and 2020.



Methods

Design

This systematic review was performed in accordance with the recommended PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [13]. The PRISMA Checklist for this study is presented in Multimedia Appendix 1.

Ethics Approval

As no patients were involved in the study, this study was exempted from institutional review board approval.

Literature Search Strategy

The PubMed database was searched for citations published between January 1, 2000, and April 30, 2020. We used the National Library of Medicine Medical Subject Headings (MeSH) terms to search for citations indexed with "artificial intelligence" and at least one other MeSH term from the OB/GYN domain: "obstetrics"; "gynecology"; "reproductive techniques, assisted"; or "pregnancy." This search was restricted to English-language publications with an abstract, using the following query: "artificial intelligence" [MeSH Terms] AND ("obstetrics" [MeSH Terms] OR "gynecology" [MeSH Terms] OR "reproductive techniques, assisted" [MeSH Terms] OR "pregnancy" [MeSH Terms]) AND 2000/1/1:2020/4/30 [Date Publication].

The results of the query were considered final on November 30, 2020, to cover all publications with potential delayed indexation in PubMed.

All retrieved citations were classified according to disciplines defined in Web of Science (WoS) for the Journal Citation Reports (JCR) and grouped in the following 9 discipline categories: OB/GYN core disciplines journals, other medical clinical disciplines journals, medical nonclinical disciplines journals, medical genetics/biology disciplines journals, medical imaging journals, medical informatics journals, computer science disciplines journals, engineering disciplines journals, and other science disciplines journals. The detailed disciplines and discipline categories are presented in Table S1 of Multimedia Appendix 2 for all journals.

We included all publications from journals or conference proceedings of the core OB/GYN WoS disciplines, namely *Obstetrics & Gynecology*, *Surgery*, *Oncology*, *Developmental Biology*, *Reproductive Biology*, *Andrology*, or *Urology & Nephrology*. The publications were excluded if no AI process was used in the study. Review, editorial, and commentary articles were also excluded.

The publication selection was independently performed by two researchers (FD, JB) by full-text review to assess the actual use of any AI process in the study. Discrepancies on AI process assessments between reviewers were resolved during meetings with KB and JMJ.

Data Collection and Analysis

The data collection and the qualitative analysis of the citations comprised six different tasks: (1) classification of publications into 5 OB/GYN domains (obstetrics, gynecology, assisted reproductive medicine, early pregnancy, and fetal medicine), (2) description of the AI method used in the study (eg, ML, knowledge base), (3) description of the AI algorithm used in

the study (eg, ANN, support vector machine, bioinformatics knowledge bases), (4) description of the type of data used in the AI process (eg, image data set, omics data set), (5) contribution of the AI process (eg, new algorithm, hypothesis generation, fully functional software), and (6) description of the validation of the AI process (eg, validation on one data set, validation on more than one data set, clinical validation). The statistical synthesis of this systematic review was performed by computing the proportion of publications by groups defined in the qualitative analysis.

The evolution over time of the scientific production related to AI in OB/GYN was assessed by a trend analysis of publications per year during the entire review period for OB/GYN core journals and other science journals. The respective contributions of all scientific disciplines in the retrieved citations were assessed by the analysis of their distribution across all WoS disciplines and the proportion of citations in each of the 9 science discipline categories.

Results

Study Selection

The PubMed search retrieved 579 citations. The 161 publications from OB/GYN core disciplines journals were reviewed for eligibility assessment. A total of 66 publications met the selection criteria [14-79]. The flow chart of the publications reviewed is presented in Figure 2.

All OB/GYN domains were represented in these selected publications (N=66): obstetrics (n=27, 41%), gynecology (n=2, 3%), assisted reproductive medicine (n=22, 33%), early pregnancy (n=1, 2%), and fetal medicine (n=14, 21%). The detailed distribution of the publications in these domains is presented in Figure 3.

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram for the selection process of the studies included in this review.

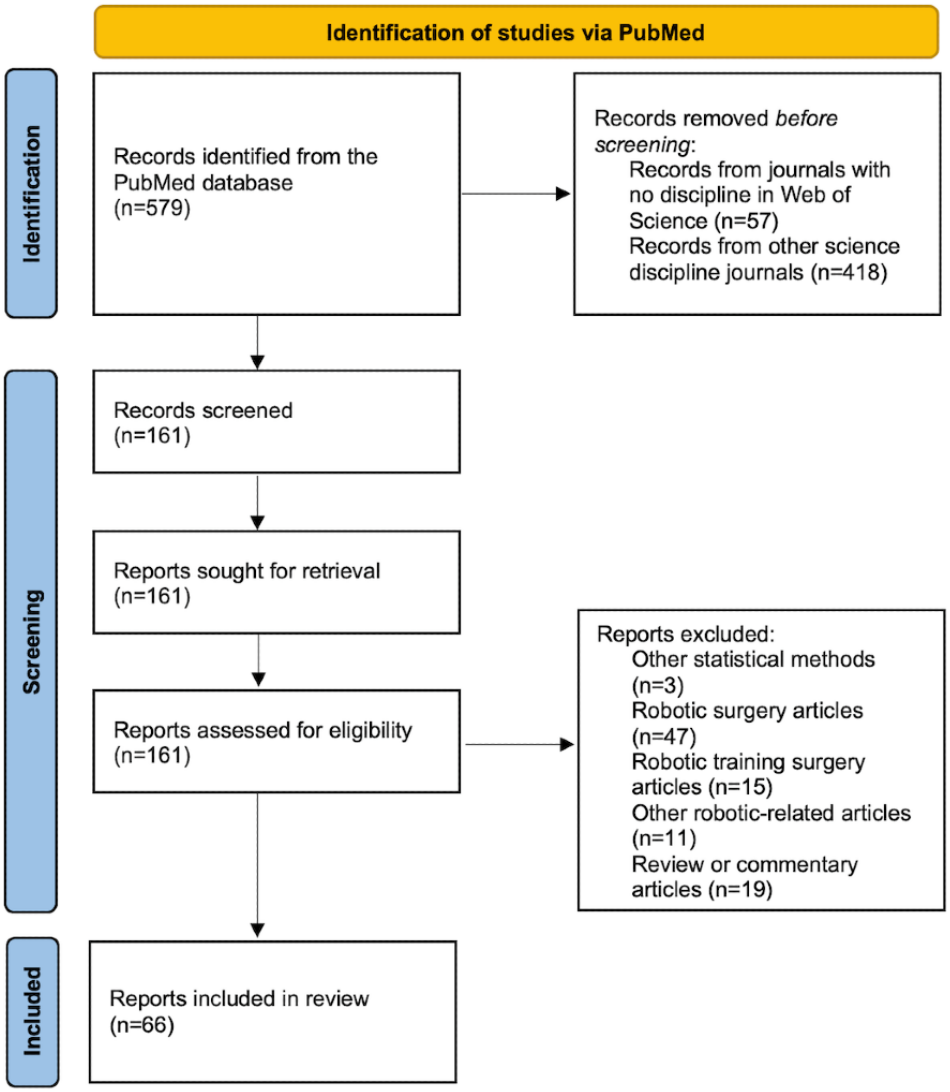
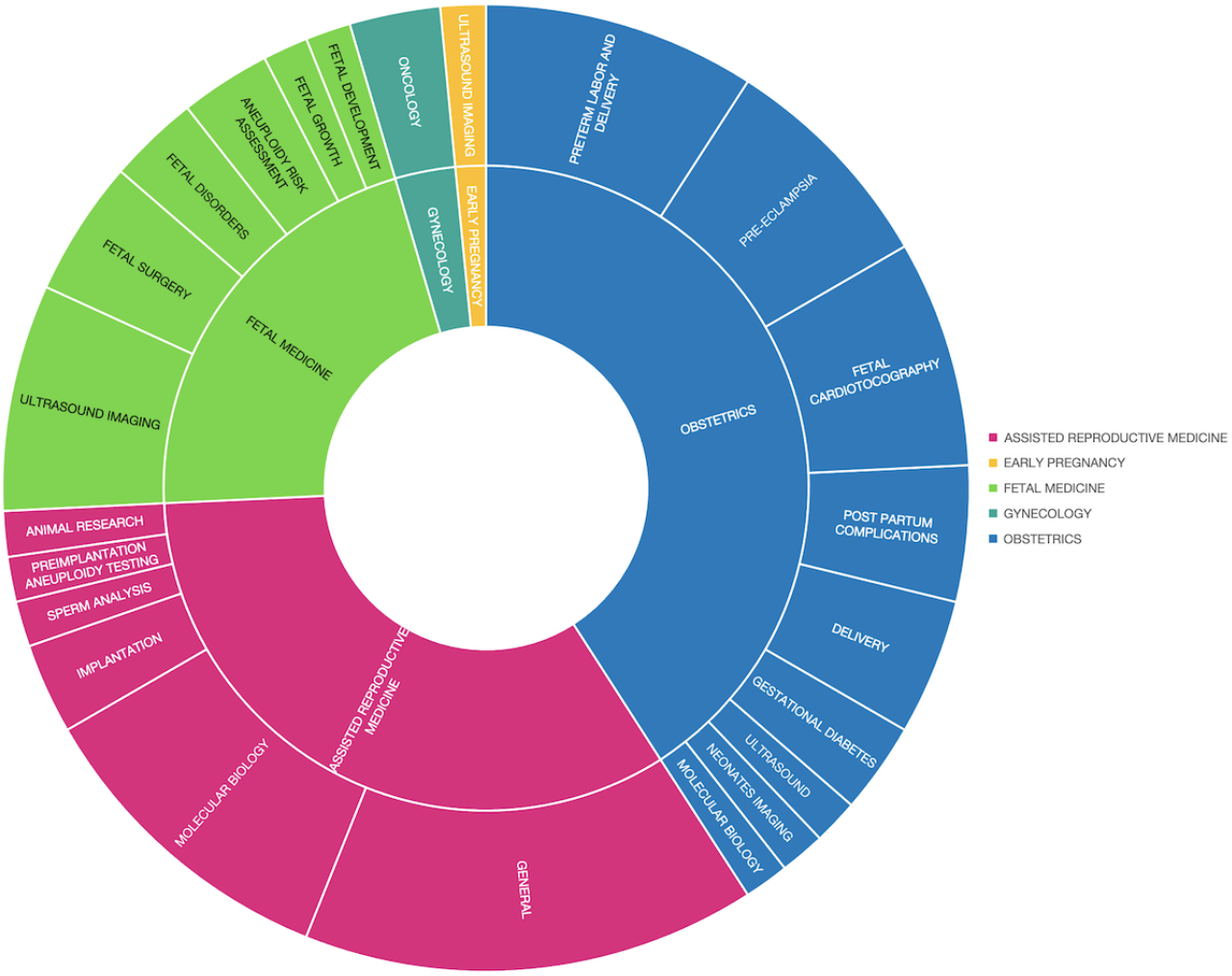


Figure 3. Distribution of the 66 artificial intelligence publications in obstetrics and gynecology journals, across subdomains.



Description of AI Methods

The two main AI methods in the 66 selected papers were represented by ML methods [14-47,66-70] (59%, n=39) and by knowledge base methods [50-65,71-79] (38%, n=25). In obstetrics and in fetal medicine, ML was more common than other AI methods in 70% (19/27) and 71% (10/14) of the publications, respectively. ML methods and knowledge bases were used in 45% (10/22) and 55% (12/22) of the publications in in the assisted reproductive medicine domain, respectively.

The ML methods comprised mainly ANNs (25/39, 64%) [14-25,27-30,33-35,41,43-45,47,68]. Diverse other ML approaches are used, ranging from classical ML tools such as support vector machine [26] and genetic algorithms [31] to more recent methods such as random forest [38,46] or gradient boosted trees [37,42,67]. Very recently, more evolved and combined neural networks are used in deep learning to process complex data for image segmentation (eg, [41,44]) or classification (eg, [39,43]). The knowledge base methods comprised bioinformatic processes involving mainly Gene Ontology (88%, 22/25) but also other omics knowledge bases, text-mining processes leveraging ontologies, and semantic reasoning processes based on domain ontologies.

The data sets used with all AI methods in the selected studies are detailed in Table 1. ML methods dealt primarily with ultrasound imaging (2D, 3D, video), numerical, and clinical data sets, whereas knowledge base methods dealt mostly with omics data sets.

The contribution of using AI methods were for algorithm development (53%), hypothesis generation (42%), or software development (3%).

When using knowledge base methods, the main AI contribution was to generate hypotheses in physiology or physiopathology (reproduction and implantation, preeclampsia, fetal growth, or breast cancer). When using ML methods, the AI contribution was to build prediction algorithms (implantation success, neonatal outcome, preterm delivery, fetal weight, aneuploidy risk, or postpartum complications). The detailed contributions for all AI methods are presented in Table 2.

Most ML methods were applied to one data set (87%, 34/39) and the use of two or more data sets was less common (13%, 5/39). No external clinical validation of ML methods was identified in the selected articles. Knowledge base methods were applied on one data set in all cases and validated in one clinical study.

Table 1. Type of data and artificial intelligence methods used in the 66 selected articles.

Type of data	Articles, n
Knowledge base method data sets	
cDNA[a]/RNA-sequencing	16
Mixed (clinical and transcriptomic data)	3
Proteomic/spectrometry	2
Other: text (publications), imaging (2D ultrasound), mixed (clinical and proteomic data), genomic data repository	4
Machine learning method data sets	
Clinical (numeric/categorical variables)	16
Numeric (fetal biometry)	4
Numeric (fetal heart monitoring/FSpO2[b] data)	4
Image (microscopy)	3
Video (fetoscopy)	3
Image (2D ultrasound)	2
Other: administrative (numerical/categorical variables), registry (numerical/categorical variables), numeric (electromyography), numeric (maternal EKG[c]), mixed (clinical and genomic data), DNA methylation, proteomic/spectrometry	7
Fuzzy logic data sets: numeric (fetal heart monitoring/FSpO2 data)	1
Other data sets, artificial intelligence method not specified (image dataset: 3D ultrasound)	1

[a]cDNA: complementary DNA.

[b]FSpO2: fetal oxygen saturation.

[c]EKG: electrocardiogram.

Table 2. Contributions of artificial intelligence methods used in the 66 selected articles.

Contribution of artificial intelligence methods	Articles, n
Knowledge base method contributions	
Hypothesis generation: ART[a] techniques/implantation physiology	7
Hypothesis generation: preeclampsia physiopathology	3
Hypothesis generation: reproduction physiology	3
Hypothesis generation: breast cancer physiopathology	2
Hypothesis generation: fetal growth/development physiology	2
Method: variant characterization	1
Method: guided ultrasound image analysis	1
Other hypothesis generation: pregnancy physiology, diabetes physiopathology, preterm labor physiopathology, recurrent pregnancy loss physiopathology, stem cell profiling, candidate gene/variant	6
Machine learning method contributions	
Algorithm: implantation/ART method success prediction	6
Algorithm: neonatal outcome prediction	4
Algorithm: preterm delivery prediction	3
Algorithm: delivery route prediction	3
Algorithm: fetal weight/growth abnormalities prediction	3
Algorithm: aneuploidy prediction/aneuploidy risk assessment	2
Algorithm: postpartum complications prediction	2
Other algorithms: gestational age prediction, preeclampsia prediction, blastocyst grading, classification of lung disorders, muscle image segmentation	5
Method: fetoscopic images annotation	2
Other methods: placental blood vessels detection, preterm outcome risk assessment, fertility phenotyping	3
Hypothesis generation: diabetes physiopathology, fetal alcohol disorder spectrum physiopathology, gastroschisis physiopathology, coagulation physiopathology, uterus physiology	5
Prototype software: ART success prediction	1
Fuzzy logic method contributions: functional software (3D fetal heart analysis)	1
Other contributions, artificial intelligence method not specified: algorithm (neonatal outcome prediction)	1

[a]ART: assisted reproductive technology.

General Trend in AI Publications

We observed a significant rising trend in the scientific production over the last two decades, mainly outside the OB/GYN core journals (Figure 4). A total of 67 science disciplines covered this scientific production (579 PubMed indexed citations), 18% of which were in OB/GYN core disciplines journals. The distribution of citations in the other discipline categories is shown in Table 3. The science discipline was not defined in WoS/JCR for 6% of the citations.

Figure 4. Trends in PubMed artificial intelligence citations between 2000 and 2020 in obstetrics and gynecology (OB/GYN) journals and in other scientific disciplines journals.

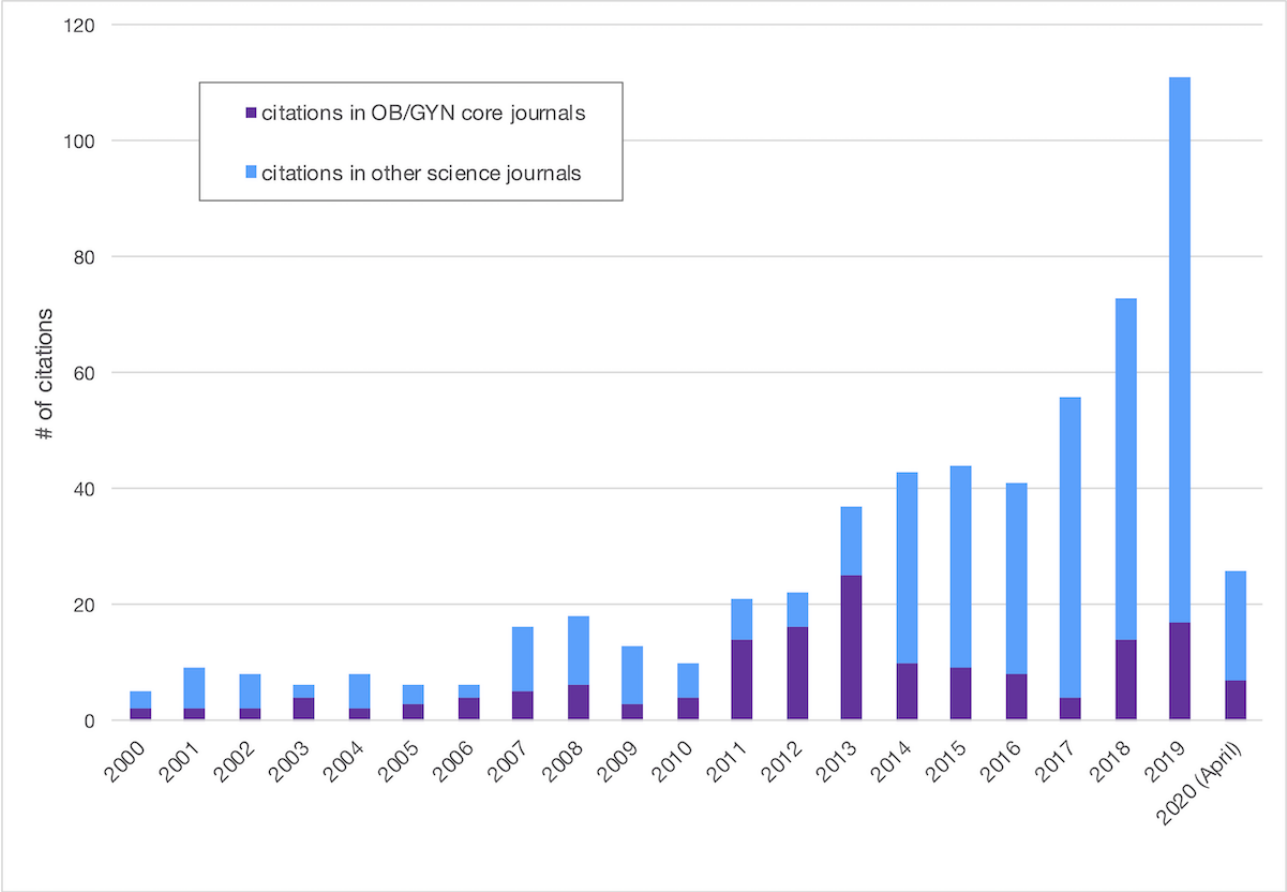


Table 3. Distribution of the 579 PubMed artificial intelligence citations between 2000 and 2020 among the 67 science disciplines.

Science disciplines	Articles (N=874)[a], n (%)
OB/GYN[b] core disciplines	161 (18.4)
Medical imaging discipline	50 (5.7)
Other medical clinical discipline	47 (5.4)
Medical nonclinical discipline	115 (13.2)
Medical informatics discipline	60 (6.9)
Medical genetics/biology disciplines	58 (6.6)
Engineering disciplines	79 (9.0)
Computer science disciplines	66 (7.6)
Other science disciplines	181 (2.1)
Absence of discipline in Web of Science	57 (6.5)

[a]Since some citations are multidisciplinary, the total is higher.

[b]OB/GYN: obstetrics and gynecology.

Discussion

Main Findings

In this review, we have demonstrated that AI contributions are emerging in OB/GYN journals and that a wide range of AI approaches (symbolic and nonsymbolic) are applied across all OB/GYN subdomains. ML is the most common nonsymbolic AI approach (59%) and articles are based mainly on ANNs (64%). Knowledge bases are the most common symbolic AI approach (38%) and are based on ontologies in most articles (88%).

However, most of the AI publications related to AI in OB/GYN (82%) remain out of the scope of the usual OB/GYN journals. Additionally, formally validated AI contributions reported to date suffer from an overall poor level of validation (one data set in most cases and no external validation in all cases) and

actual AI contributions remain at the level of "proof of concept" or "proof of feasibility."

Publications in OB/GYN Discipline Journals

The reported AI contribution to OB/GYN in the core discipline journals was 18% in comparison with 82% in journals of other disciplines. This can be explained by the early stage of research in AI or by the absence of clinical validation, meaning that the results are more relevant for the AI and computer science community. When novel algorithms are developed, computer science journals are naturally preferred [80-84]; for example, one of the first convolutional neural networks able to perform automated plane recognition during a fetal ultrasound scan was reported in a computer science journal [85]. In addition a clinically validated ML-based fetal biometric system was reported in a general medical imaging journal, not in a core OB/GYN discipline journal [86]. Another contribution based on logic and semantic reasoning for early pregnancy diagnosis was reported in a medical informatics journal [87]. These examples illustrate that core OB/GYN discipline journals await clinical value demonstration of AI-based research rather than reporting on novel systems. This pattern might also suggest that the time has come for the OB/GYN community to take over some valuable early-stage AI contributions within its core discipline journals.

Additionally, we have observed more advanced AI techniques and architectures applied to OB/GYN in computer science journals than in OB/GYN discipline journals. Moreover, the well-established and most robust ANN architectures (eg, U-net, ResNet) are no longer published in computer science journals and are largely published in OB/GYN discipline journals to present another application context [39,41,44]. As a result, a strong representation of experts in AI methods in editorial boards could improve editorial choices, which would help to fill in the delay of translation of advanced AI to the OB/GYN readership.

Interestingly, reported AI methods are applied in unconnected data silos in the field of OB/GYN (images, omics data, clinical data, other data modalities) and mixed AI methods in the field of OB/GYN are in early stages. Thus, approaches involving both ML and knowledge bases is a new direction that we expect to emerge. For example, the Smart Ultrasound in Obstetrics and Gynecology (SUOG) initiative (EIT-Health Innovation program) [88] combines knowledge bases for differential diagnosis and ML for image analysis to develop an AI-based ultrasound diagnosis assistant.

Quality of AI Research Reporting in the OB/GYN Field

The low level of validation of AI processes in medicine has been previously reported [89]. We also observed significant heterogeneity in the description of AI processes in this review, with an overall limited level of description in most publications and with a poor level of clinical validation. This can be explained because, until recently, there were no AI-specific guidelines for medical publications. Indeed, most AI notions are new to the medical readership, medical editorial boards, and medical literature indexing. Some medical publications have proposed glossaries and definitions of basic AI notions, and the first reference guidelines for reporting medical studies involving AI were published in 2020 [90-95]. Although these initiatives should improve the reporting of AI-related publications, these guidelines only cover ML approaches. For example, the extension of SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) guidelines for clinical trial protocols using interventions involving AI (SPIRIT-AI) [92] lists the items of interest for AI publications but does not cover knowledge representations, ontologies, semantic reasoning, nor knowledge bases. In addition, we found that 38% of the articles in this review leverage these AI approaches. Consequently, a further extension of these recommendations could advantageously provide guidelines for symbolic AI approaches.

Albeit not covered in the guidelines for AI-related research, some "routine" methods in statistics (eg, logistic regression, multivariate logistic regression) and in data visualization (eg, K-means clustering) are also considered as ML approaches [96]. In this review, we excluded studies based on these methods [97-99]. However, from a perspective of consistency, some statistical methods involving ML techniques could also be covered by AI-related research guidelines.

There are recurrent debates on ethical and legal considerations in AI methods in the news and social media; therefore, we were surprised that most publications do not elaborate on these aspects. The majority use nonexplainable approaches such as ANNs; while using such nonexplainable methods is acceptable, some limitations need disclosure, and their reproducibility needs proper assessment. The most straightforward assessment method of reproducibility relies on external validation, which remains critical prior to application of all methods, but even more so if nonexplainable. Human responsibility in using AI-based processes also depends on the level of autonomy of the process [100] and on recommendations to use such processes [101].

Limitations of MeSH Indexation in PubMed

This is the first systematic review on AI contributions reported in OB/GYN core journals. This study was performed by a pluridisciplinary group of experts from both the OB/GYN and computer science communities [102]. We have limited our paper selection to citations in PubMed and used the science disciplines as defined by WoS/JCR, thus controlling potential bias in the definition of journal domains. Although our method is reproducible and complies with systematic review guidelines, it is by essence subject to bias in publication indexation. For example, articles with ML methods mentioned only in one paragraph (eg, [103]) are not covered in this study. In addition, for papers with a scope in decision support (eg, [104,105]), the indexation will not fall under the MeSH term "artificial intelligence" in PubMed but rather under the MeSH term "diagnosis, computer-assisted" that is a distinct notion. However, unlike systematic reviews of clinical therapies, this limitation is less of a problem as we were still able to ascertain general trends in this relatively novel field of study.

All reviewed papers on robotic surgery were indexed in PubMed with the MeSH term "robotics" and under the MeSH term "artificial intelligence." Currently, in MeSH, "robotics" is a subcategory of "artificial intelligence." As a result, all robotic surgery papers are considered to be AI papers, which is not

always the case. A revision of MeSH terms and/or indexation policies could be a solution for disambiguation. Additionally, the use of appropriate AI-oriented keywords provided by authors at the time of submission could improve the characterization of AI-based research.

Conclusions

Until mid-2020, mostly preliminary work in AI applied in OB/GYN has been reported and published outside the usual OB/GYN journals. When published in OB/GYN journals, multiple data set validation and clinical validation of AI processes remain unmet prerequisites. Clarification in AI methods could be achieved by improving the MeSH indexing of publications in PubMed. Additionally, the reporting of AI applications should be improved by the new 2020 guidelines and recommendations in medical research involving AI. These are promising for future clinically relevant and methodologically valid clinical trials publications. However, these guidelines are covering only a part of AI approaches involved in the articles reviewed in this study, and updates need to be considered, especially to cover symbolic AI approaches.

Acknowledgments

This work was partially supported by the EIT-HEALTH Innovation (SUOG project, BP2020#20062 and BP2021#211015; principal investigator FD) and by the Sorbonne University's Institute of Technology for Health (IUIS 2019 Doctoral Program Grant to JB).

Authors' Contributions

FD designed the review and drafted the original manuscript. FD, JB, KB, and JMJ reviewed the articles. AP, JMJ, and PM interpreted the data and reviewed/edited the manuscript. All authors have reviewed the submitted version.

Conflicts of Interest

FD and JMJ are inventors of an ontology-based imaging protocol, patented by their university (Sorbonne University) and implemented in the SUOG project. ATP is a co-founder and senior scientific advisor for Intelligent Ultrasound. The other authors have no conflicts of interest to declare.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOC File , 65 KB - jmir_v24i4e35465_app1.doc]

Multimedia Appendix 2
List of journals covered by the review, by science discipline (source Web of Science) and grouped by discipline categories.
[DOCX File , 41 KB - jmir_v24i4e35465_app2.docx]

References

1. McCarthy J, Minsky M, Rochester N, Shannon C. A proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. AI Magazine 2006 Aug 31;27(4):12. [doi: 10.1609/aimag.v27i4.1904]
2. Haugeland J. Artificial intelligence: the very idea. Cambridge, MA: MIT Press; 1985.
3. Ledley RS, Lusted LB. Reasoning foundations of medical diagnosis; symbolic logic, probability, and value theory aid our understanding of how physicians reason. Science 1959 Jul 03;130(3366):9-21. [doi: 10.1126/science.130.3366.9] [Medline: 13668531]
4. Shortliffe EH, Axline SG, Buchanan BG, Merigan TC, Cohen SN. An artificial intelligence program to advise physicians regarding antimicrobial therapy. Comput Biomed Res 1973 Dec;6(6):544-560. [doi: 10.1016/0010-4809(73)90029-3] [Medline: 4589706]
5. Shortliffe EH, Davis R, Axline SG, Buchanan BG, Green C, Cohen SN. Computer-based consultations in clinical therapeutics: explanation and rule acquisition capabilities of the MYCIN system. Comput Biomed Res 1975 Aug;8(4):303-320. [doi: 10.1016/0010-4809(75)90009-9] [Medline: 1157471]
6. Dhombres F, Charlet J, Section Editors for the IMIA Yearbook Section on Knowledge RepresentationManagement. Formal medical knowledge representation supports deep learning algorithms, bioinformatics pipelines, genomics data analysis, and big data processes. Yearb Med Inform 2019 Aug 16;28(1):152-155 [FREE Full text] [doi: 10.1055/s-0039-1677933] [Medline: 31419827]
7. Dhombres F, Charlet J, Section Editors for the IMIA Yearbook Section on Knowledge RepresentationManagement. As ontologies reach maturity, artificial intelligence starts being fully efficient: findings from the section on knowledge representation and management for the Yearbook 2018. Yearb Med Inform 2018 Aug 29;27(1):140-145 [FREE Full text] [doi: 10.1055/s-0038-1667078] [Medline: 30157517]

8. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res 2019 Jan 08;47(D1):D330-D338 [FREE Full text] [doi: 10.1093/nar/gky1055] [Medline: 30395331]
9. Reference Genome Group of the Gene Ontology Consortium T. The Gene Ontology's Reference Genome Project: a unified framework for functional annotation across species. PLoS Comput Biol 2009 Jul 3;5(7):e1000431 [FREE Full text] [doi: 10.1371/journal.pcbi.1000431] [Medline: 19578431]
10. Köhler S, Carmody L, Vasilevsky N, Jacobsen JOB, Danis D, Gourdine J, et al. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. Nucleic Acids Res 2019 Jan 08;47(D1):D1018-D1027 [FREE Full text] [doi: 10.1093/nar/gky1105] [Medline: 30476213]
11. 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]
12. Drukker L, Noble JA, Papageorghiou AT. Introduction to artificial intelligence in ultrasound imaging in obstetrics and gynecology. Ultrasound Obstet Gynecol 2020 Oct;56(4):498-505. [doi: 10.1002/uog.22122] [Medline: 32530098]
13. 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]
14. Lockwood CJ, Kuczynski E. Risk stratification and pathological mechanisms in preterm delivery. Paediatr Perinat Epidemiol 2001 Jul;15(Suppl 2):78-89. [doi: 10.1046/j.1365-3016.2001.00010.x] [Medline: 11520402]
15. Macones GA, Hausman N, Edelstein R, Stamilio DM, Marder SJ. Predicting outcomes of trials of labor in women attempting vaginal birth after cesarean delivery: a comparison of multivariate methods with neural networks. Am J Obstet Gynecol 2001 Feb;184(3):409-413. [doi: 10.1067/mob.2001.109386] [Medline: 11228495]
16. Salamalekis E, Thomopoulos P, Giannaris D, Salloum I, Vasios G, Prentza A, et al. Computerised intrapartum diagnosis of fetal hypoxia based on fetal heart rate monitoring and fetal pulse oximetry recordings utilising wavelet analysis and neural networks. BJOG 2002 Oct;109(10):1137-1142. [doi: 10.1111/j.1471-0528.2002.01388.x] [Medline: 12387467]
17. Maeda K, Noguchi Y, Matsumoto F. Evaluation of prolonged fetal monitoring with normal and pathologic outcome probabilities determined by artificial neural network. Fetal Diagn Ther 2003 Aug 15;18(5):370-375. [doi: 10.1159/000071982] [Medline: 12913350]
18. Maeda K, Nagasawa T. Automatic computerized diagnosis of fetal sinusoidal heart rate. Fetal Diagn Ther 2005 Aug 19;20(5):328-334. [doi: 10.1159/000086807] [Medline: 16113548]
19. Wald M, Sparks AE, Sandlow J, Van-Voorhis B, Syrop CH, Niederberger CS. Computational models for prediction of IVF/ICSI outcomes with surgically retrieved spermatozoa. Reprod Biomed Online 2005 Sep;11(3):325-331. [doi: 10.1016/s1472-6483(10)60840-1] [Medline: 16176672]
20. Valensise H, Facchinetti F, Vasapollo B, Giannini F, Monte ID, Arduini D. The computerized fetal heart rate analysis in post-term pregnancy identifies patients at risk for fetal distress in labour. Eur J Obstet Gynecol Reprod Biol 2006 Apr 01;125(2):185-192. [doi: 10.1016/j.ejogrb.2005.06.034] [Medline: 16459010]
21. Shi S, Maner WL, Mackay LB, Garfield RE. Identification of term and preterm labor in rats using artificial neural networks on uterine electromyography signals. Am J Obstet Gynecol 2008 Feb;198(2):235 [FREE Full text] [doi: 10.1016/j.ajog.2007.08.039] [Medline: 18226633]
22. Elaveyini U, Devi SP, Rao KS. Neural networks prediction of preterm delivery with first trimester bleeding. Arch Gynecol Obstet 2011 May 7;283(5):971-979. [doi: 10.1007/s00404-010-1469-2] [Medline: 20449599]
23. Ma Y, Chen B, Wang H, Hu K, Huang Y. Prediction of sperm retrieval in men with non-obstructive azoospermia using artificial neural networks: leptin is a good assistant diagnostic marker. Hum Reprod 2011 Feb 07;26(2):294-298. [doi: 10.1093/humrep/deq337] [Medline: 21138907]
24. Tejera E, Jose Areias M, Rodrigues A, Ramõa A, Manuel Nieto-Villar J, Rebelo I. Artificial neural network for normal, hypertensive, and preeclamptic pregnancy classification using maternal heart rate variability indexes. J Matern Fetal Neonatal Med 2011 Sep 21;24(9):1147-1151. [doi: 10.3109/14767058.2010.545916] [Medline: 21250912]
25. Cheng Y, Yan G, Chiu YH, Chang F, Chang C, Chung K. Efficient fetal size classification combined with artificial neural network for estimation of fetal weight. Taiwan J Obstet Gynecol 2012 Dec;51(4):545-553 [FREE Full text] [doi: 10.1016/j.tjog.2012.09.009] [Medline: 23276557]
26. Santos Filho E, Noble JA, Poli M, Griffiths T, Emerson G, Wells D. A method for semi-automatic grading of human blastocyst microscope images. Hum Reprod 2012 Sep;27(9):2641-2648. [doi: 10.1093/humrep/des219] [Medline: 22736327]
27. Cheng Y, Chiu YH, Wang H, Chang F, Chung K, Chang C, et al. Using Akaike information criterion and minimum mean square error mode in compensating for ultrasonographic errors for estimation of fetal weight by new operators. Taiwan J Obstet Gynecol 2013 Mar;52(1):46-52 [FREE Full text] [doi: 10.1016/j.tjog.2013.01.008] [Medline: 23548217]
28. Manna C, Nanni L, Lumini A, Pappalardo S. Artificial intelligence techniques for embryo and oocyte classification. Reprod Biomed Online 2013 Jan;26(1):42-49. [doi: 10.1016/j.rbmo.2012.09.015] [Medline: 23177416]
29. Ramasamy R, Padilla WO, Osterberg EC, Srivastava A, Reifsnyder JE, Niederberger C, et al. A comparison of models for predicting sperm retrieval before microdissection testicular sperm extraction in men with nonobstructive azoospermia. J Urol 2013 Feb;189(2):638-642. [doi: 10.1016/j.juro.2012.09.038] [Medline: 23260551]

30. Guzmán-Bárcenas J, Hernández JA, Arias-Martínez J, Baptista-González H, Ceballos-Reyes G, Irles C. Estimation of umbilical cord blood leptin and insulin based on anthropometric data by means of artificial neural network approach: identifying key maternal and neonatal factors. BMC Pregnancy Childbirth 2016 Jul 21;16(1):179 [FREE Full text] [doi: 10.1186/s12884-016-0967-z] [Medline: 27440187]
31. Papageorghiou AT, Kemp B, Stones W, Ohuma EO, Kennedy SH, Purwar M, International FetalNewborn Growth Consortium for the 21st Century (INTERGROWTH-21st). Ultrasound-based gestational-age estimation in late pregnancy. Ultrasound Obstet Gynecol 2016 Dec;48(6):719-726. [doi: 10.1002/uog.15894] [Medline: 26924421]
32. Veeramani SK, Muthusamy E. Detection of abnormalities in ultrasound lung image using multi-level RVM classification. J Matern Fetal Neonatal Med 2016 Jul 02;29(11):1844-1852. [doi: 10.3109/14767058.2015.1064888] [Medline: 26135771]
33. Bahado-Singh RO, Sonek J, McKenna D, Cool D, Aydas B, Turkoglu O, et al. Artificial intelligence and amniotic fluid multiomics: prediction of perinatal outcome in asymptomatic women with short cervix. Ultrasound Obstet Gynecol 2019 Jul 09;54(1):110-118. [doi: 10.1002/uog.20168] [Medline: 30381856]
34. Beksac MS, Tanacan A, Bacak HO, Leblebicioglu K. Computerized prediction system for the route of delivery (vaginal birth versus cesarean section). J Perinat Med 2018 Oct 25;46(8):881-884. [doi: 10.1515/jpm-2018-0022] [Medline: 29570455]
35. Neocleous AC, Syngelaki A, Nicolaides KH, Schizas CN. Two-stage approach for risk estimation of fetal trisomy 21 and other aneuploidies using computational intelligence systems. Ultrasound Obstet Gynecol 2018 Apr 04;51(4):503-508. [doi: 10.1002/uog.17558] [Medline: 28640401]
36. Kuhle S, Maguire B, Zhang H, Hamilton D, Allen AC, Joseph KS, et al. Comparison of logistic regression with machine learning methods for the prediction of fetal growth abnormalities: a retrospective cohort study. BMC Pregnancy Childbirth 2018 Aug 15;18(1):333. [doi: 10.1186/s12884-018-1971-2] [Medline: 30111303]
37. Lussier AA, Morin AM, MacIsaac JL, Salmon J, Weinberg J, Reynolds JN, et al. DNA methylation as a predictor of fetal alcohol spectrum disorder. Clin Epigenetics 2018 Jan 12;10(1):5-14 [FREE Full text] [doi: 10.1186/s13148-018-0439-6] [Medline: 29344313]
38. Blank C, Wildeboer RR, DeCroo I, Tilleman K, Weyers B, de Sutter P, et al. Prediction of implantation after blastocyst transfer in in vitro fertilization: a machine-learning perspective. Fertil Steril 2019 Feb;111(2):318-326. [doi: 10.1016/j.fertnstert.2018.10.030] [Medline: 30611557]
39. Vasconcelos F, Brandão P, Vercauteren T, Ourselin S, Deprest J, Peebles D, et al. Towards computer-assisted TTTS: Laser ablation detection for workflow segmentation from fetoscopic video. Int J Comput Assist Radiol Surg 2018 Oct 27;13(10):1661-1670 [FREE Full text] [doi: 10.1007/s11548-018-1813-8] [Medline: 29951938]
40. Liang B, Gao Y, Xu J, Song Y, Xuan L, Shi T, et al. Raman profiling of embryo culture medium to identify aneuploid and euploid embryos. Fertil Steril 2019 Apr;111(4):753-762. [doi: 10.1016/j.fertnstert.2018.11.036] [Medline: 30683589]
41. Sadda P, Imamoglu M, Dombrowski M, Papademetris X, Bahtiyar MO, Onofrey J. Deep-learned placental vessel segmentation for intraoperative video enhancement in fetoscopic surgery. Int J Comput Assist Radiol Surg 2019 Feb 27;14(2):227-235 [FREE Full text] [doi: 10.1007/s11548-018-1886-4] [Medline: 30484115]
42. Betts KS, Kisely S, Alati R. Predicting common maternal postpartum complications: leveraging health administrative data and machine learning. BJOG 2019 May;126(6):702-709. [doi: 10.1111/1471-0528.15607] [Medline: 30628159]
43. Tran D, Cooke S, Illingworth PJ, Gardner DK. Deep learning as a predictive tool for fetal heart pregnancy following time-lapse incubation and blastocyst transfer. Hum Reprod 2019 Jun 04;34(6):1011-1018 [FREE Full text] [doi: 10.1093/humrep/dez064] [Medline: 31111884]
44. van den Noort F, van der Vaart CH, Grob ATM, van de Waarsenburg MK, Slump CH, van Stralen M. Deep learning enables automatic quantitative assessment of puborectalis muscle and urogenital hiatus in plane of minimal hiatal dimensions. Ultrasound Obstet Gynecol 2019 Aug 26;54(2):270-275. [doi: 10.1002/uog.20181] [Medline: 30461079]
45. Vogiatzi P, Pouliakis A, Siristatidis C. An artificial neural network for the prediction of assisted reproduction outcome. J Assist Reprod Genet 2019 Jul 19;36(7):1441-1448 [FREE Full text] [doi: 10.1007/s10815-019-01498-7] [Medline: 31218565]
46. Weber KA, Yang W, Carmichael SL, Padula AM, Shaw GM. A machine learning approach to investigate potential risk factors for gastroschisis in California. Birth Defects Res 2019 Mar 01;111(4):212-221 [FREE Full text] [doi: 10.1002/bdr2.1441] [Medline: 30588769]
47. Bano S, Vasconcelos F, Vander Poorten E, Vercauteren T, Ourselin S, Deprest J, et al. FetNet: a recurrent convolutional network for occlusion identification in fetoscopic videos. Int J Comput Assist Radiol Surg 2020 May 29;15(5):791-801 [FREE Full text] [doi: 10.1007/s11548-020-02169-0] [Medline: 32350787]
48. Czabański R, Jezewski J, Wróbel J, Sikora J, Jezewski M. Application of fuzzy inference systems for classification of fetal heart rate tracings in relation to neonatal outcome. Ginekol Pol 2013 Jan;84(1):38-43. [doi: 10.17772/gp/1538] [Medline: 23488308]
49. Yeo L, Romero R. Fetal Intelligent Navigation Echocardiography (FINE): a novel method for rapid, simple, and automatic examination of the fetal heart. Ultrasound Obstet Gynecol 2013 Sep 02;42(3):268-284. [doi: 10.1002/uog.12563] [Medline: 24000158]
50. Adona PR, Leal CL, Biase FH, De Bem TH, Mesquita LG, Meirelles FV, et al. In vitro maturation alters gene expression in bovine oocytes. Zygote 2016 Feb 17;24(4):624-633. [doi: 10.1017/s0967199415000672]

51. Fang P, Zeng P, Wang Z, Liu M, Xu W, Dai J, et al. Estimated diversity of messenger RNAs in each murine spermatozoa and their potential function during early zygotic development. Biol Reprod 2014 May;90(5):94. [doi: 10.1095/biolreprod.114.117788] [Medline: 24671878]
52. Franczak A, Wojciechowicz B, Kotwica G. Transcriptomic analysis of the porcine endometrium during early pregnancy and the estrous cycle. Reprod Biol 2013 Sep;13(3):229-237. [doi: 10.1016/j.repbio.2013.07.001] [Medline: 24011194]
53. Forde N, Maillo V, O'Gaora P, Simintiras CA, Sturmey RG, Ealy AD, et al. Sexually dimorphic gene expression in bovine conceptuses at the initiation of implantation. Biol Reprod 2016 Oct 03;95(4):92-92 [FREE Full text] [doi: 10.1095/biolreprod.116.139857] [Medline: 27488033]
54. Santonocito M, Vento M, Guglielmino MR, Battaglia R, Wahlgren J, Ragusa M, et al. Molecular characterization of exosomes and their microRNA cargo in human follicular fluid: bioinformatic analysis reveals that exosomal microRNAs control pathways involved in follicular maturation. Fertil Steril 2014 Dec;102(6):1751-1761. [doi: 10.1016/j.fertnstert.2014.08.005] [Medline: 25241362]
55. Peng G, Suo S, Chen J, Chen W, Liu C, Yu F, et al. Spatial transcriptome for the molecular annotation of lineage fates and cell identity in mid-gastrula mouse embryo. Dev Cell 2016 Mar 21;36(6):681-697 [FREE Full text] [doi: 10.1016/j.devcel.2016.02.020] [Medline: 27003939]
56. Hao Y, Zhang Z, Han D, Cao Y, Zhou P, Wei Z, et al. Gene expression profiling of human blastocysts from in vivo and 'rescue IVM' with or without melatonin treatment. Mol Med Rep 2017 Aug;16(2):1278-1288 [FREE Full text] [doi: 10.3892/mmr.2017.6742] [Medline: 28627630]
57. Légaré C, Akintayo A, Blondin P, Calvo E, Sullivan R. Impact of male fertility status on the transcriptome of the bovine epididymis. Mol Hum Reprod 2017 Jun 01;23(6):355-369. [doi: 10.1093/molehr/gax019] [Medline: 28379507]
58. Maurice P, Dhombres F, Blondiaux E, Friszer S, Guilbaud L, Lelong N, et al. Towards ontology-based decision support systems for complex ultrasound diagnosis in obstetrics and gynecology. J Gynecol Obstet Hum Reprod 2017 May;46(5):423-429. [doi: 10.1016/j.jogoh.2017.03.004] [Medline: 28934086]
59. Aghaeepour N, Lehallier B, Baca Q, Ganio EA, Wong RJ, Ghaemi MS, et al. A proteomic clock of human pregnancy. Am J Obstet Gynecol 2018 Mar;218(3):347. [doi: 10.1016/j.ajog.2017.12.208] [Medline: 29277631]
60. Hu X, Ao J, Li X, Zhang H, Wu J, Cheng W. Competing endogenous RNA expression profiling in pre-eclampsia identifies hsa_circ_0036877 as a potential novel blood biomarker for early pre-eclampsia. Clin Epigenetics 2018 Apr 10;10(1):48 [FREE Full text] [doi: 10.1186/s13148-018-0482-3] [Medline: 29643944]
61. Cao C, Wen Y, Wang X, Fang N, Yuan S, Huang X. Testicular piRNA profile comparison between successful and unsuccessful micro-TESE retrieval in NOA patients. J Assist Reprod Genet 2018 May 19;35(5):801-808 [FREE Full text] [doi: 10.1007/s10815-018-1134-4] [Medline: 29460056]
62. Lykoudi A, Kolialexi A, Lambrou GI, Braoudaki M, Siristatidis C, Papaioanou GK, et al. Dysregulated placental microRNAs in early and late onset preeclampsia. Placenta 2018 Jan;61:24-32. [doi: 10.1016/j.placenta.2017.11.005] [Medline: 29277268]
63. Lu C, Yan Z, Song X, Xu Y, Zheng X, Li R, et al. Effect of exogenous gonadotropin on the transcriptome of human granulosa cells and follicular fluid hormone profiles. Reprod Biol Endocrinol 2019 Jun 24;17(1):49 [FREE Full text] [doi: 10.1186/s12958-019-0489-4] [Medline: 31234873]
64. Ntostis P, Kokkali G, Iles D, Huntriss J, Tzetis M, Picton H, et al. Can trophectoderm RNA analysis predict human blastocyst competency? Syst Biol Reprod Med 2019 Aug 27;65(4):312-325 [FREE Full text] [doi: 10.1080/19396368.2019.1625085] [Medline: 31244343]
65. Santucci-Pereira J, Zeleniuch-Jacquotte A, Afanasyeva Y, Zhong H, Slifker M, Peri S, et al. Genomic signature of parity in the breast of premenopausal women. Breast Cancer Res 2019 Mar 28;21(1):46 [FREE Full text] [doi: 10.1186/s13058-019-1128-x] [Medline: 30922380]
66. Hamilton EF, Dyachenko A, Ciampi A, Maurel K, Warrick PA, Garite TJ. Estimating risk of severe neonatal morbidity in preterm births under 32 weeks of gestation. J Matern Fetal Neonatal Med 2020 Jan 18;33(1):73-80. [doi: 10.1080/14767058.2018.1487395] [Medline: 29886760]
67. Kamphuis C, Duenk P, Veerkamp RF, Visser B, Singh G, Nigsch A, et al. Machine learning to further improve the decision which boar ejaculates to process into artificial insemination doses. Theriogenology 2020 Mar 01;144:112-121 [FREE Full text] [doi: 10.1016/j.theriogenology.2019.12.017] [Medline: 31927416]
68. Lipschuetz M, Guedalia J, Rottenstreich A, Novoselsky Persky M, Cohen SM, Kabiri D, et al. Prediction of vaginal birth after cesarean deliveries using machine learning. Am J Obstet Gynecol 2020 Jun;222(6):613. [doi: 10.1016/j.ajog.2019.12.267] [Medline: 32007491]
69. Venkatesh KK, Strauss RA, Grotegut CA, Heine RP, Chescheir NC, Stringer JSA, et al. Machine learning and statistical models to predict postpartum hemorrhage. Obstet Gynecol 2020 Apr;135(4):935-944. [doi: 10.1097/AOG.0000000000003759] [Medline: 32168227]
70. Miyagi Y, Tada K, Yasuhi I, Maekawa Y, Okura N, Kawakami K, et al. New method for determining fibrinogen and FDP threshold criteria by artificial intelligence in cases of massive hemorrhage during delivery. J Obstet Gynaecol Res 2020 Feb 24;46(2):256-265. [doi: 10.1111/jog.14166] [Medline: 31762151]

71. Shi C, Han HJ, Fan LJ, Guan J, Zheng XB, Chen X, et al. Diverse endometrial mRNA signatures during the window of implantation in patients with repeated implantation failure. Hum Fertil 2018 Sep 19;21(3):183-194. [doi: 10.1080/14647273.2017.1324180] [Medline: 28523980]
72. Song Y, Liu J, Huang S, Zhang L. Analysis of differentially expressed genes in placental tissues of preeclampsia patients using microarray combined with the Connectivity Map database. Placenta 2013 Dec;34(12):1190-1195. [doi: 10.1016/j.placenta.2013.09.013] [Medline: 24125805]
73. Su Y, Zhang Y. Identification of biological processes and genes for gestational diabetes mellitus. Arch Gynecol Obstet 2015 Sep 4;292(3):635-640. [doi: 10.1007/s00404-015-3649-6] [Medline: 25736406]
74. Triche EW, Uzun A, DeWan AT, Kurihara I, Liu J, Occhiogrosso R, et al. Bioinformatic approach to the genetics of preeclampsia. Obstet Gynecol 2014 Jun;123(6):1155-1161 [FREE Full text] [doi: 10.1097/AOG.0000000000000293] [Medline: 24807322]
75. Weng S, Chang S, Cheng Y, Wang H, Wang T, Chang MD, et al. Comparative transcriptome analysis reveals a fetal origin for mesenchymal stem cells and novel fetal surface antigens for noninvasive prenatal diagnosis. Taiwan J Obstet Gynecol 2011 Dec;50(4):447-457 [FREE Full text] [doi: 10.1016/j.tjog.2011.10.009] [Medline: 22212316]
76. Xue D, Zhang Y, Wang Y, Wang J, An F, Sun X, et al. Quantitative proteomic analysis of sperm in unexplained recurrent pregnancy loss. Reprod Biol Endocrinol 2019 Jul 09;17(1):52 [FREE Full text] [doi: 10.1186/s12958-019-0496-5] [Medline: 31288842]
77. Ye K, Dai J, Liu L, Peng M. Network-based gene function inference method to predict optimal gene functions associated with fetal growth restriction. Mol Med Rep 2018 Sep 29;18(3):3003-3010. [doi: 10.3892/mmr.2018.9232] [Medline: 30015878]
78. Yuan W, Heesom K, Phillips R, Chen L, Trinder J, López Bernal A. Low abundance plasma proteins in labour. Reproduction 2012 Oct;144(4):505-518. [doi: 10.1530/REP-12-0114] [Medline: 22859520]
79. Zhou Q, Sun E, Ling L, Liu X, Zhang M, Yin H, et al. Bioinformatic analysis of computational identified differentially expressed genes in tumor stoma of pregnancy-associated breast cancer. Mol Med Rep 2017 Sep;16(3):3345-3350. [doi: 10.3892/mmr.2017.6947] [Medline: 28713995]
80. Wang G, Zuluaga MA, Li W, Pratt R, Patel PA, Aertsen M, et al. DeepIGeoS: A deep interactive geodesic framework for medical image segmentation. IEEE Trans Pattern Anal Mach Intell 2019 Jul 1;41(7):1559-1572. [doi: 10.1109/tpami.2018.2840695]
81. Meng Q, Housden J, Matthew J, Rueckert D, Schnabel JA, Kainz B, et al. Weakly supervised estimation of shadow confidence maps in fetal ultrasound imaging. IEEE Trans Med Imaging 2019 Dec;38(12):2755-2767. [doi: 10.1109/tmi.2019.2913311]
82. Mohseni Salehi SS, Khan S, Erdogmus D, Gholipour A. Real-time deep pose estimation with geodesic loss for image-to-template rigid registration. IEEE Trans Med Imaging 2019 Feb;38(2):470-481. [doi: 10.1109/tmi.2018.2866442]
83. Rueda S, Fathima S, Knight CL, Yaqub M, Papageorghiou AT, Rahmatullah B, et al. Evaluation and comparison of current fetal ultrasound image segmentation methods for biometric measurements: a grand challenge. IEEE Trans Med Imaging 2014 Apr;33(4):797-813. [doi: 10.1109/tmi.2013.2276943]
84. Rahmatullah B, Papageorghiou A, Noble J. Integration of local and global features for anatomical object detection in ultrasound. Med Image Comput Comput Assist Interv 2012;15(Pt 3):402-409. [doi: 10.1007/978-3-642-33454-2_50] [Medline: 23286156]
85. Baumgartner CF, Kamnitsas K, Matthew J, Fletcher TP, Smith S, Koch LM, et al. SonoNet: real-time detection and localisation of fetal standard scan planes in freehand ultrasound. IEEE Trans Med Imaging 2017 Nov;36(11):2204-2215. [doi: 10.1109/tmi.2017.2712367]
86. Ambroise Grandjean G, Hossu G, Bertholdt C, Noble P, Morel O, Grangé G. Artificial intelligence assistance for fetal head biometry: assessment of automated measurement software. Diagn Interv Imaging 2018 Nov;99(11):709-716 [FREE Full text] [doi: 10.1016/j.diii.2018.08.001] [Medline: 30177447]
87. Dhombres F, Maurice P, Guilbaud L, Franchinard L, Dias B, Charlet J, et al. A novel intelligent scan assistant system for early pregnancy diagnosis by ultrasound: clinical decision support system evaluation study. J Med Internet Res 2019 Jul 03;21(7):e14286 [FREE Full text] [doi: 10.2196/14286] [Medline: 31271152]
88. Smart Ultrasound in Obstetrics and Gynecology (SUOG) project. URL: https://www.suog.org/ [accessed 2022-04-13]
89. 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]
90. Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. Lancet 2019 Apr 20;393(10181):1577-1579. [doi: 10.1016/S0140-6736(19)30037-6] [Medline: 31007185]
91. Andaur Navarro CL, Damen JAAG, Takada T, Nijman SWJ, Dhiman P, Ma J, et al. Protocol for a systematic review on the methodological and reporting quality of prediction model studies using machine learning techniques. BMJ Open 2020 Nov 11;10(11):e038832 [FREE Full text] [doi: 10.1136/bmjopen-2020-038832] [Medline: 33177137]
92. Cruz Rivera S, Liu X, Chan A, Denniston AK, Calvert MJ, SPIRIT-AICONSORT-AI Working Group, SPIRIT-AICONSORT-AI Steering Group, SPIRIT-AICONSORT-AI Consensus Group. Guidelines for clinical trial

protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Nat Med 2020 Sep;26(9):1351-1363 [FREE Full text] [doi: 10.1038/s41591-020-1037-7] [Medline: 32908284]
93. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, SPIRIT-AICONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. BMJ 2020 Sep 09;370:m3164 [FREE Full text] [doi: 10.1136/bmj.m3164] [Medline: 32909959]
94. Norgeot B, Quer G, Beaulieu-Jones BK, Torkamani A, Dias R, Gianfrancesco M, et al. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. Nat Med 2020 Sep 09;26(9):1320-1324 [FREE Full text] [doi: 10.1038/s41591-020-1041-y] [Medline: 32908275]
95. Sounderajah V, Ashrafian H, Aggarwal R, De Fauw J, Denniston AK, Greaves F, et al. Developing specific reporting guidelines for diagnostic accuracy studies assessing AI interventions: The STARD-AI Steering Group. Nat Med 2020 Jun 08;26(6):807-808. [doi: 10.1038/s41591-020-0941-1] [Medline: 32514173]
96. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. New York, NY: Springer; 2009.
97. Ballester M, Oppenheimer A, d'Argent EM, Touboul C, Antoine J, Coutant C, et al. Nomogram to predict pregnancy rate after ICSI-IVF cycle in patients with endometriosis. Hum Reprod 2012 Feb 23;27(2):451-456. [doi: 10.1093/humrep/der392] [Medline: 22114107]
98. Bottomley C, Van Belle V, Kirk E, Van Huffel S, Timmerman D, Bourne T. Accurate prediction of pregnancy viability by means of a simple scoring system. Hum Reprod 2013 Jan;28(1):68-76. [doi: 10.1093/humrep/des352] [Medline: 23111205]
99. Souza RT, Cecatti JG, Passini R, Pacagnella RC, Oliveira PF, Silva CM, Brazilian Multicentre Study on Preterm Birth study group. Cluster analysis identifying clinical phenotypes of preterm birth and related maternal and neonatal outcomes from the Brazilian Multicentre Study on Preterm Birth. Int J Gynaecol Obstet 2019 Jul 21;146(1):110-117. [doi: 10.1002/ijgo.12839] [Medline: 31055833]
100. Jaremko JL, Azar M, Bromwich R, Lum A, Alicia Cheong LH, Gibert M, Canadian Association of Radiologists (CAR) Artificial Intelligence Working Group. Canadian Association of Radiologists White Paper on Ethical and Legal Issues Related to Artificial Intelligence in Radiology. Can Assoc Radiol J 2019 May 01;70(2):107-118 [FREE Full text] [doi: 10.1016/j.carj.2019.03.001] [Medline: 30962048]
101. Price WN, Gerke S, Cohen IG. Potential liability for physicians using artificial intelligence. JAMA 2019 Nov 12;322(18):1765-1766. [doi: 10.1001/jama.2019.15064] [Medline: 31584609]
102. Littmann M, Selig K, Cohen-Lavi L, Frank Y, Hönigschmid P, Kataka E, et al. Validity of machine learning in biology and medicine increased through collaborations across fields of expertise. Nat Mach Intell 2020 Jan 13;2(1):18-24. [doi: 10.1038/s42256-019-0139-8]
103. Jugović D, Tumbri J, Medić M, Jukić MK, Kurjak A, Arbeille P, et al. New Doppler index for prediction of perinatal brain damage in growth-restricted and hypoxic fetuses. Ultrasound Obstet Gynecol 2007 Sep;30(3):303-311. [doi: 10.1002/uog.4094] [Medline: 17721870]
104. Huang Y, Chen D, Jiang Y, Kuo S, Wu H, Moon WK. Computer-aided diagnosis using morphological features for classifying breast lesions on ultrasound. Ultrasound Obstet Gynecol 2008 Sep;32(4):565-572. [doi: 10.1002/uog.5205] [Medline: 18383556]
105. Huang Y, Kuo S, Chang C, Liu Y, Moon WK, Chen D. Image retrieval with principal component analysis for breast cancer diagnosis on various ultrasonic systems. Ultrasound Obstet Gynecol 2005 Oct 08;26(5):558-566. [doi: 10.1002/uog.1951] [Medline: 16086435]

Abbreviations

AI: artificial intelligence
ANN: artificial neural network
GOFAI: Good Old-Fashioned Artificial Intelligence
JCR: Journal Citation Reports
MeSH: Medical Subject Heading
ML: machine learning
OB/GYN: obstetrics and gynecology
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SPIRIT-AI: Standard Protocol Items: Recommendations for Interventional Trials involving Artificial Intelligence
SUOG: Smart Ultrasound in Obstetrics and Gynecology
WoS: Web of Science

Edited by A Mavragani; submitted 05.12.21; peer-reviewed by R Rastmanesh, H Ayatollahi; comments to author 18.12.21; revised version received 11.02.22; accepted 15.03.22; published 20.04.22.

Please cite as:
Dhombres F, Bonnard J, Bailly K, Maurice P, Papageorghiou AT, Jouannic JM
Contributions of Artificial Intelligence Reported in Obstetrics and Gynecology Journals: Systematic Review
J Med Internet Res 2022;24(4):e35465
URL: https://www.jmir.org/2022/4/e35465
doi:10.2196/35465
PMID:35297766

Review

Publicly Available, Interactive Web-Based Tools to Support Advance Care Planning: Systematic Review

Charlèss Dupont[1], MSc; Tinne Smets[1,2], PhD; Fanny Monnet[1,2], MSc; Lara Pivodic[1,2], PhD; Aline De Vleminck[1,2], PhD; Chantal Van Audenhove[3], PhD; Lieve Van den Block[1,2], PhD

[1]End-of-Life Care Research Group, Vrije Universiteit Brussel & Ghent University, Brussels, Belgium

[2]Department of Family Medicine and Chronic Care, Vrije Universiteit Brussel, Brussel, Belgium

[3]LUCAS Center for Care Research and Consultancy, KU Leuven, Leuven, Belgium

Corresponding Author:
Charlèss Dupont, MSc
End-of-Life Care Research Group
Vrije Universiteit Brussel & Ghent University
Laarbeeklaan 103
Brussels, 1090
Belgium
Phone: 32 484 433 257
Email: charless.dupont@vub.be

Abstract

Background: There is an increasing number of interactive web-based advance care planning (ACP) support tools, which are web-based aids in any format encouraging reflection, communication, and processing of publicly available information, most of which cannot be found in the peer-reviewed literature.

Objective: This study aims to conduct a systematic review of web-based ACP support tools to describe the characteristics, readability, and quality of content and investigate whether and how they are evaluated.

Methods: We systematically searched the web-based gray literature databases OpenGrey, ClinicalTrials.gov, ProQuest, British Library, Grey Literature in the Netherlands, and Health Services Research Projects in Progress, as well as Google and app stores, and consulted experts using the following eligibility criteria: web-based, designed for the general population, accessible to everyone, interactive (encouraging reflection, communication, and processing of information), and in English or Dutch. The quality of content was evaluated using the *Quality Evaluation Scoring Tool* (score 0-28—a higher score indicates better quality). To synthesize the characteristics of the ACP tools, readability and quality of content, and whether and how they were evaluated, we used 4 data extraction tables.

Results: A total of 30 tools met the eligibility criteria, including 15 (50%) websites, 10 (33%) web-based portals, 3 (10%) apps, and 2 (7%) with a combination of formats. Of the 30 tools, 24 (80%) mentioned a clear aim, including 7 (23%) that supported reflection or communication, 8 (27%) that supported people in making decisions, 7 (23%) that provided support to document decisions, and 2 (7%) that aimed to achieve all these aims. Of the 30 tools, 7 (23%) provided information on the development, all of which were developed in collaboration with health care professionals, and 3 (10%) with end users. Quality scores ranged between 11 and 28, with most of the lower-scoring tools not referring to information sources.

Conclusions: A variety of ACP support tools are available on the web, varying in the quality of content. In the future, users should be involved in the development process of ACP support tools, and the content should be substantiated by scientific evidence.

Trial Registration: PROSPERO CRD42020184112; https://tinyurl.com/mruf8b43

(J Med Internet Res 2022;24(4):e33320) doi:10.2196/33320

KEYWORDS

advance care planning; systematic review; web-based tools; health communication; quality of online content

Introduction

Via a European consensus process, advance care planning (ACP) has been defined as a process that enables individuals to define goals and preferences for future medical care, discuss these preferences with family and health care providers, and record these preferences and choices [1,2]. In recent decades, the concept of ACP has changed considerably, shifting from a clinician-led and documentation-focused process that emphasizes the need for advance directives to a broader concept of ongoing communication regarding various aspects of future care and treatment planning [3,4]. In the recent public health literature, the concept has been broadened further by emphasizing the opportunities that ACP conversations offer for normalizing and reshaping how we think, talk, and make decisions about the last chapters of our lives [5,6].

To support people in having such conversations, a wide variety of ACP support tools have been developed in several formats such as print or websites. They exist in different kinds of modalities (guides, card games, and videos) and for different target groups: people with specific diseases and their families, family caregivers, or the general public [7,8]. With the growing use of the internet and international efforts to promote ACP [9], more web-based ACP support tools are becoming publicly available [10]. An advantage of these web-based tools is that they are easily accessible to a large audience, are often interactive, and can thus be tailored to the needs of the user. Interactive elements include, for example, questions or exercises to encourage reflection and communication and process the information provided [10-12].

A recent review of the published peer-reviewed studies assessed the feasibility and effectiveness of interactive web-based ACP support tools for adult patients, relatives, and healthy individuals and found that users considered the tools easy to use and not burdensome. It also demonstrated that they could improve a user's knowledge of ACP, ACP communication with relatives and health care professionals, and ACP documentation [8]. However, this review was not able to include all ACP support tools available to the general public, as many exist only on the web and have not been published in academic journals [7,8]. Reviewing these web-based tools is important as the quality of web-based content can vary widely or be based on personal opinions and experiences rather than on scientific evidence [13-16]. This can be particularly problematic with regard to ACP, as the content may be biased in favor of or against certain medical interventions [15], whereas the primary purpose of ACP should be to promote choices based on individual values and preferences [1,2].

Currently, there is no comprehensive overview of interactive ACP tools available on the internet. Therefore, this systematic review aims to answer the following research questions:

1. What are the characteristics and functionalities of interactive web-based ACP support tools?
2. How is ACP addressed in these tools?
3. What is the readability and quality of their content, and have they been evaluated in a study, and if so, what is their level of evidence?

Methods

Review Design and Protocol Registration

We conducted a systematic review of web-based interactive ACP support tools (hereafter called tools) following the 4 search strategies for web-based gray literature by Godin et al [17]. Search, selection, and data synthesis were performed between September 2020 and January 2021. The protocol for this systematic review was registered with PROSPERO (International Prospective Register of Systematic Reviews; CRD42020184112).

Eligibility Criteria

We searched for tools that met the following inclusion criteria:

1. Designed to support the general population; that is, people with or without serious illnesses and their families
2. Available on the internet
3. Accessible to whoever visits the tool and can be used by anyone
4. Interactive; that is, encourage the user to reflect, communicate, formulate decisions, or document wishes [10-12]
5. In English or Dutch

Tools that exclusively aimed to support health care professionals in the ACP process were excluded.

Search Strategies

We systematically searched for tools using the four search strategies recommended by Godin et al [17]: (1) web-based gray literature databases, (2) search engines, (3) app stores, and (4) expert consultation. The first three search strategies were conducted separately by 2 researchers (CD and FM), who both used the same search combinations (Multimedia Appendices 1-3) on different computers with *clean* (deleted cookies and history) browsers without being logged into a Google account. Because, as with peer-reviewed databases, every search database (gray literature databases, search engines, and app stores) has its own search functionalities and filters, we adapted the search terms to fit into the search fields of each database.

Furthermore, as search engines have their own algorithms for showing relevant results, we used several different search combinations (ie, combining the search terms, their permutations, and trending keywords) in search strategy 2 to minimize the risk of omitting relevant sources (Multimedia Appendix 2). As a search engine can give an overwhelming amount of *hits*, we screened the first 15 pages (first 150 hits) of each search combination.

We screened the available content of the results, such as executive summaries, the webpage *about*, or the explanation of the tool—when available—until we could answer the following question: *is this tool potentially a web-based ACP support tool for the general population*?. If the answer was *yes*, we included the tool for full screening and transferred the results to a Microsoft Excel file (with the name of the tool or website if there was no specific name for the tool).

For the first search strategy (gray literature databases), we used the following databases: *OpenGrey*, *ClinicalTrials.gov*, *ProQuest*, *British Library*, *Grey Literature in the Netherlands*, and *Health Services Research Projects in Progress*. For the second search strategy, we used the *Google search engine*, and for the search in app stores, we used *Google Play Store* and *Apple App Store.*

The fourth search strategy entailed consultation with experts on ACP. We identified experts via frequently listed and cited authors of the relevant literature, known stakeholders, and suggestions from other key informants. We consulted these experts by email and, to achieve saturation, asked them whether they knew any other tools that we had not found using the first three search strategies. The responses were added to the Microsoft Excel file and saved for the final selection.

Selection of Web-Based ACP Support Tools

For the final selection, the 2 researchers (CD and FM) separately screened all available content of each potentially relevant tool against the eligibility criteria. In cases of disagreement about whether to include a tool, a third reviewer (TS) screened it and made the final decision on including it. We used Archive.is to archive the home page or the first page of the tools.

Evaluation of the Readability and Quality of the Content of the Included Tools

Readability was evaluated using web-based readability analysis tools. These readability tools calculate several readability scores based on the characteristics of the text, such as the number of syllables per word and number of words per sentence. As the web-based readability analysis tools were exclusively for one language, we used two tools: one for the English tools [18] and one for the Dutch tools [19]. To determine readability, we used the Common European Framework of Reference for Languages (CEFR) levels [20,21]. The CEFR levels can be used to determine both English and Dutch readability and are calculated by the algorithm of web-based readability analysis tools and look at the number of words per sentence, number of pronouns and prepositions in a sentence, and number of simple words [22]. CEFR comprises six reading levels (A1, A2, B1, B2, C1, and C2), with A1 as the easiest level and C2 as the most difficult. The recommended readability standard for CEFR is B1 [22].

To evaluate the quality of the content of the tools, we used the validated *Quality Evaluation Scoring Tool* (QUEST) [23]. This quality assessment tool can be used to assess web-based health content by evaluating 7 items, each assigned with a weighted score. Six items have a possible score between 0 to 1 or 0 to 2 and weight between 1 and 3: authorship (score 0-2×1), conflicts of interest (score 0-2×3), complementarity (whether they support the patient-physician relationship; score 0-1×1), currency (if the content is frequently updated; score 0-2×1) and the tone of the content (whether the content was *fully supported* using strong vocabulary such as *cure*, *guarantee*, and *easy*; *mainly supported* where the authors mainly support their claims but with more cautious vocabulary; or *balanced/cautious support* with statements of limitations or contrasting findings; score 0-2×3). The seventh item is attribution (whether and what kind of sources are used to create the content) and is measured in two steps: first, identifying the presence of references to scientific studies (score 0-3×3) and, second, when scoring >1, identifying the type of studies referred to in vitro, observational studies, or meta-analyses or clinical trials (score 0-2×1) [23,24]. Each tool was evaluated for each of the 7 items. The scores for each item were summed to create a total quality score ranging between 0 and 28, with higher scores indicating a better quality of the content in the tool.

Readability was evaluated by 2 researchers (CD and FM) who copy-pasted the text of a full webpage into the text fields of the web-based readability score calculator tools. The same 2 researchers assessed the content of the tools to determine the QUEST score for each tool. Any disagreements on readability and QUEST scores were discussed to reach a consensus. If no consensus was found, a third researcher (TS) made the final decision.

Evaluation of the Level of Evidence of the Included Tools

To assess whether the included tools had been evaluated as part of a research study, we (1) screened each one for any information on evaluation; (2) we searched the gray literature databases OpenGrey, ClinicalTrials.gov, ProQuest, British Library, Grey Literature in the Netherlands, and Health Services Research Projects in Progress; and (3) we searched for publications of primary peer-reviewed studies in PubMed, Web of Science, PsycINFO, and CINAHL. Search terms in the database included the names of the tools, websites, or persons or organizations involved in the development of the tool.

We screened the available content (for scientific article abstracts or, when needed, full texts) to check whether the result was a primary study on one of the included tools. If it was, where possible, we determined the level of evidence using the Hierarchy of Evidence from the National Health and Medical Research Council [25,26] to determine how the tools were evaluated.

Two researchers (CD and FM) independently assessed the full texts of the peer-reviewed studies to determine the level of evidence and assess what was evaluated (usability and effectiveness). Any disagreements were discussed to reach a consensus. If no consensus was reached, a third researcher (TS) made the final decision.

Data Synthesis and Outcomes of Interest

To answer the research questions, we developed 4 data extraction tables. For the first 3 extraction tables, the 2 researchers (CD and FM) independently assessed the content of the tools to summarize the addressed characteristics, functionalities, and key elements of ACP. Any disagreements were discussed to reach a consensus. If no consensus was reached, a third researcher (TS) made the final decision.

The first extraction table was used to assess the aim; target group; available languages; format; and where, by whom, and how the tools were developed. We evaluated the used functionalities by developing a second extraction table based on a review of peer-reviewed studies by van der Smissen et al

[8]. We slightly changed the 12 assessed functionalities by removing *can be used without assistance,* as in our review, we only included tools that were designed to be used by the general population, and we added the functionalities *predetermined path and possibility to save input and return* as these allow the user to conduct the ACP process at their own pace and *input can be printed* as this functionality can increase the accessibility of websites [27]. Therefore, in this review, we extracted the following data with regard to the functionalities of ACP tools:

1. The tool is free of charge.
2. Registration is needed.
3. There is a save and return option.
4. It is possible for the user to give input.
5. The tool is tailored to the user.
6. The tools provide feedback based on the input of the user.
7. The input from the user can be printed.
8. The tool suggests a predetermined ACP path.
9. The tool gives an indication of the progress of the user.
10. Videos are used, as well as hyperlinks to other web pages.
11. There is a text-to-speech option.
12. A privacy policy is mentioned.
13. Data log analysis is mentioned.

The manner in which ACP is addressed in the tools was extracted using 14 ACP key elements (the third extraction table) as an analytic framework. The 14 key elements were self-developed and based on the recommendations of Rietjens et al [1], the consensus definition of ACP of Sudore et al [2], and the review by van der Smissen et al [8]. From these definitions, we aimed to extract all relevant elements that can be part of the ACP; that is, the following elements:

1. Providing information on ACP
2. Providing information on legal frameworks
3. Address readiness and timing for ACP
4. Stimulates to explore personal values and goals
5. Stimulates to explore preferences regarding future care
6. Stimulates to explore uncertainties and consequences
7. Stimulates to explore preferences regarding the last days of life
8. Stimulates to explore preferences for a possible proxy decision-maker
9. Encourages to appoint a proxy decision-maker
10. Encourages to discuss ACP with family
11. Encourages to discuss ACP with health care professionals
12. Encourages to document ACP
13. Encourages to generate that document (in the tool)
14. Encourages to share that document.

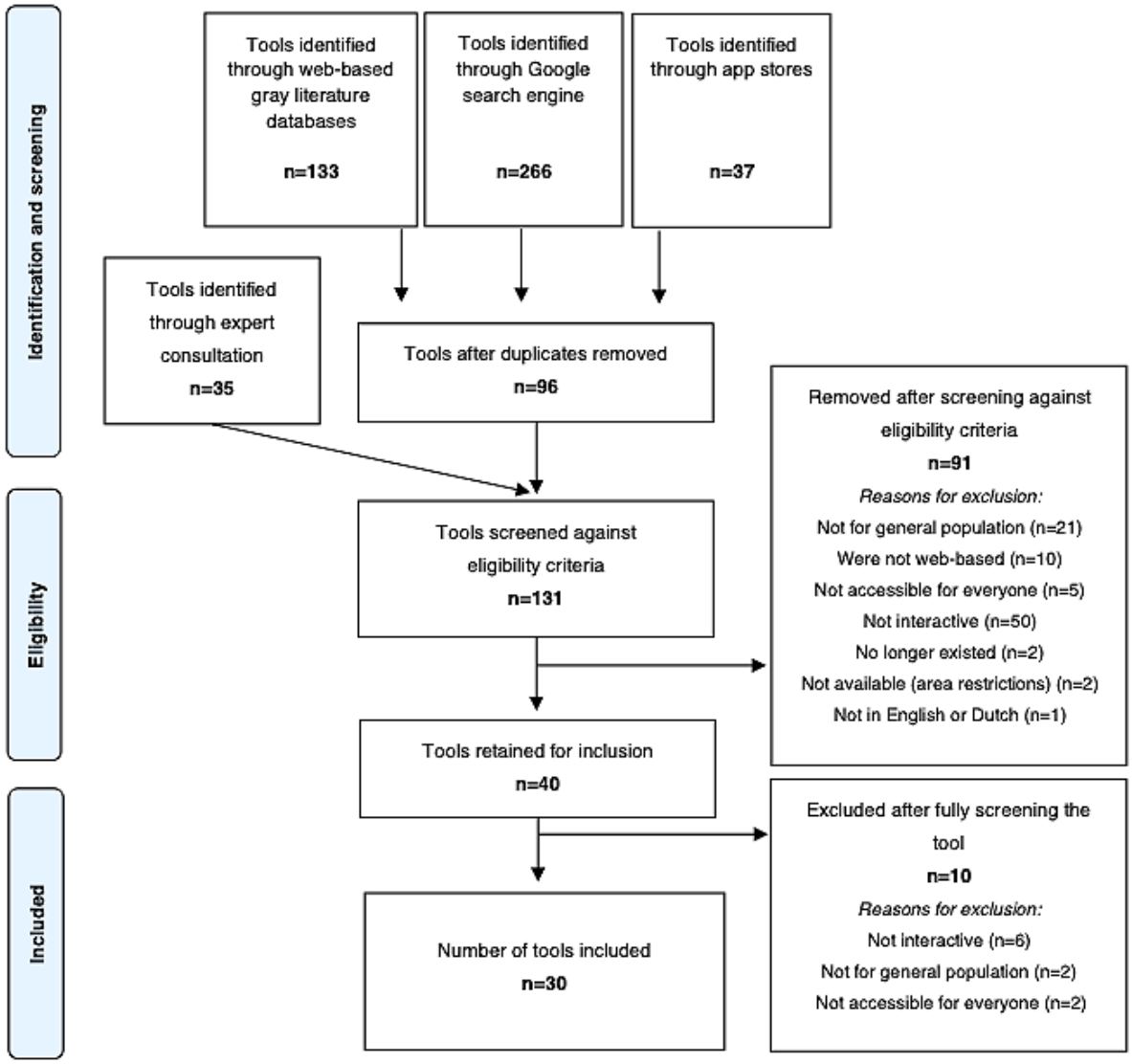
The readability, quality of the content, information about the evaluation of the tools, and their level of evidence were summarized in the fourth extraction table.

Results

Selection and Inclusion of the Tools

We found 436 tools using the first 3 search strategies. After removing duplicates (the tools retrieved and the removed duplicates per search strategy can be found in Multimedia Appendix 4), a list of 96 potential interactive web-based ACP support tools for the general population was sent to 15 experts on ACP. Of the 15 experts, 14 (93%) replied, and together they identified 35 additional tools. Two researchers (CD and FM) subsequently screened the 131 tools against the eligibility criteria. Approximately 69.5% (91/131) were excluded (Figure 1), and there were no conflicts to resolve regarding exclusion. The remaining 30.5% (40/131) of tools were fully screened (ie, all text available in the tool for data extraction), and we ended up excluding 10 more in an agreement between both researchers. Thus, the total number of included tools was 30. An overview of screening, selection, and reasons for exclusion and inclusion is shown in Figure 1.

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



Characteristics of the Tools

Approximately 80% (24/30) of tools targeted the general population, although 20% (6/30) also aimed to target health care professionals (Multimedia Appendix 5 [28-57]). For instance, the app *ACDCare* could be used by patients to make an advance care directive that can be uploaded to their health record and to which the health care professional has access [28]. The tools *Mydirectives*, *Be My Voice*, *My Living Will*, *Speak Up*, and *Considering your own future health care* not only supported the general population in ACP but also provided information on ACP for health care professionals [29-33]. Of the 30 tools, 24 (80%) did mention a clear aim; 7 (23%) supported reflection or communication, 8 (27%) supported people in making decisions, 7 (23%) provided support to document decisions, and 2 (7%) aimed to achieve all these objectives.

Of the 30 tools, 15 (50%) were websites, 10 (33%) were web-based portals with the possibility of logging in and returning to personal information, 3 (10%) were apps, and 2 (7%) had a combination of formats. For example, *PREPARE for Your Care* (*PREPARE*) can be used via the website or by logging in to a web-based portal [34].

Of the 30 tools, 12 (40%) were developed in the United States, 6 (20%) in Canada, 4 (13%) in Australia, 3 (10%) in the Netherlands, 3 (10%) in the United Kingdom, 1 (3%) in New Zealand, and 1 (3%) in Belgium. Approximately 83% (25/30) were owned by nonprofit organizations. Approximately 23% (7/30) provided information on their development; all were developed in collaboration with health care professionals, and 10% (3/30) were also developed with end users.

Functionalities of the Tools

Of the 30 tools, 27 (90%) were free of charge, and 17 (57%) were available without registration (Multimedia Appendix 6 [28-57]). Approximately 97% (29/30) of tools offered users the option of providing input by responding to a question or statement, for example, by asking them to write in an empty

box. With regard to the flow of the tool (the steps or path a user must take to go through it), 97% (29/30) used a predetermined path [28-44,46-57]. For example, if there are *x* number of steps in a tool, the user needs to go through these steps to *finish* the ACP process provided in the tool. One of the tools (the *Go Wish* card game) does not have a set path but uses a mechanism in which users can sort cards on particular wishes and preferences to stimulate reflection and communication [45]. Approximately 17% (5/30) used the input of the user to tailor information, redirect them to a specific webpage with more information, or ask for clarification on the input. For example, in *My Values*, when identifying *not wanting to be a burden* as important, a screen pops up asking them to explain briefly (by typing) what *becoming a burden* means to them [49]. One also uses input to give feedback (Multimedia Appendix 6 [28-57])—*PREPARE* not only provides information based on input but also provides tips when, for example, the user indicates *not ready to choose a proxy decision-maker* [34].

ACP Elements Addressed in the Tools

All tools contained at least seven of the 14 ACP key elements identified in our analytical framework, and 20% (6/30) of tools comprised all (Multimedia Appendix 7 [28-57]). With regard to the information available, 40% (12/30) provided both information on ACP and on the legal frameworks of ACP. Of the 30 tools, readiness and timing of ACP were addressed in 15 (50%), encouraging people to explore personal values and goals in 28 (93%), and preferences regarding future care in 29 (97%) tools; uncertainties and consequences of hypothetical serious illness scenarios were addressed in 25 (83%), preferences regarding the last days of life in 27 (90%), and the possibility of appointing a proxy decision-maker in 21 (70%) tools. Of the 30 tools, 20 (67%) encouraged appointing a proxy decision-maker, and 27 (90%) encouraged discussing ACP with family or a health care professional and documenting ACP outcomes (for example, using an advance directive); in 27 (90%) tools, it was possible to document one's wishes and preferences within the tool itself and share this document with others (by printing, via email, or via a direct link; Multimedia Appendix 7 [28-57]).

Readability and Quality of the Content

The readability of the ACP tools varied (Table 1); however, 83% (25/30) of ACP tools had a B1 or lower CEFR level. The QUEST scores of the tools varied between 11 and 28 (theoretical scale scores between 0 and 28; Table 2). Most of the lower-scoring tools did not refer to any sources to support the information they contained, were not current (ie, not updated in the past 5 years), or did not provide any information on authorship. The tools that scored ≥21 used at least one reference to a scientific study to support the information in the tool.

Table 1. Readability scores of the interactive, web-based ACP[a] support tools.[b]

Tools	CEFR[c] scale
My decisions	A1
My living voice	A1
Speak up	A1
PREPARE[d]	A2
Go Wish card game	A2
MyDirectives	A2
Plan Your Life Span	A2
Lets Think Ahead—My ACP	A2
Advance Care Planning: Should I Stop Treatment That Prolongs My Life?	B1
The Letter project Advance Directive	B1
Advance Care Planning: Should I Receive CPR[e] and Life Support?	B1
MyWishes	B1
My Living Will	B1
Everplans	B1
Considering your own future health care	B1
Tijdig nadenken over het levenseinde	B1
Oog in Oog	B1
Verken uw wensen voor zorg en behandeling	B1
Dying to Talk	B1
Be my voice	B1
ACDCare	B1
Five Wishes	B1
MyValues	B1
Dementia Values and Priorities Tool	B1
Advance Care Planning: Should I Have Artificial Hydration and Nutrition?	B2
Planning for Your Future	B2
Beslishulp—Vroegtijdige zorgplanning	B2
Cake	B2
NVLivingWill	B2
Advance Care Planning: Should I Stop Kidney Dialysis?	C1

[a]ACP: advance care planning.

[b]Ranked from lowest (A1) to highest score (C2) possible.

[c]CEFR: Common European Framework of Reference for Languages.

[d]PREPARE: PREPARE for Your Care.

[e]CPR: cardiopulmonary resuscitation.

Table 2. QUEST[a] scores of the interactive, web-based ACP[b] support tools.[c]

Tools	QUEST score
PREPARE[d]	28
Go Wish card game	23
Considering your own future health care	20
Everplans	20
Oog in Oog	20
Dementia Values and Priorities Tool	19
Dying to Talk	19
My Living Will	19
MyDirectives	19
NVLivingWill	19
Plan your Life Span	19
Cake	17
Five Wishes	17
Lets Think Ahead–My ACP	17
Planning for Your Future	17
Be my voice	16
Beslishulp-Vroegtijdige zorgplanning	16
MyValues	16
MyWishes	16
My decisions	14
Advance Care Planning: Should I Have Artificial Hydration and Nutrition?	13
Advance Care Planning: Should I Receive CPR[e] and Life Support?	13
Advance Care Planning: Should I Stop Kidney Dialysis?	13
Advance Care Planning: Should I Stop Treatment That Prolongs My Life?	13
My living voice	12
ACDCare	11

[a]QUEST: Quality Evaluation Scoring Tool.

[b]ACP: advance care planning.

[c]Ranked from lowest (0) to highest score (28) possible.

[d]PREPARE: PREPARE for Your Care.

[e]CPR: cardiopulmonary resuscitation.

Evaluated Tools and Their Level of Evidence

Of the 30 included tools, 5 (16%) tools had been evaluated in a study, all of which were published in the peer-reviewed literature (Table 3). *Verken uw wensen voor zorg en behandeling* [57] is under evaluation; however, the results have not yet been published. *MyDirectives* and *NVLivingWill* were studied using a posttest design [58-60], and *PREPARE*, *Plan Your Life Span*, and *The Letter project Advance Directive* were studied in at least one properly designed randomized controlled trial [61-71]. The study on *NVLivingWill* evaluated the ease of use of the tool [58]; the studies on *PREPARE* evaluated ease of use, effectiveness, acceptability, and understandability (information in the tool was easy to read) [63-71]; *Plan Your Life Span* evaluated effectiveness [61]; and the study on *The Letter project Advance Directive* evaluated understandability [72].

Table 3. Level of evidence of the interactive, web-based advance care planning support tools.

Tool, authors, and title of the study		The hierarchy of evidence[a]
***MyDirectives* [29]**		Evidence obtained from a posttest
Holland et al [59]	Nurse-led patient-centered advance care planning in primary care	
Fine et al [60]	Early experience with digital advance care planning and directives, a novel consumer-driven program	
***NVLivingWill* [51]**		Evidence obtained from a posttest
Klugman and Usatine [58]	An evaluation of 2 online advance directive programs	
***Plan Your Life Span* [53]**		Evidence obtained from at least one properly designed randomized controlled trial
Ramirez-Zohfeld et al [62]	Longitudinal follow-up of long-term care planning using PlanYourLifespan.org	
Lindquist et al [61]	PlanYourLifeSpan.org—an intervention to help seniors make choices for their fourth quarter of life: results from the randomized clinical trial	
***PREPARE*[b] [34]**		Evidence obtained from at least one properly designed randomized controlled trial
Howard et al [71]	Effect of an interactive website to engage patients in advance care planning in outpatient settings	
Freytag et al [70]	Empowering older adults to discuss advance care planning during clinical visits: the PREPARE randomized trial	
Lum et al [69]	Improving a full range of advance care planning behavior change and action domains: the PREPARE randomized trial	
Zapata et al [68]	Feasibility of a video-based advance care planning website to facilitate group visits among diverse adults from a safety-net health system	
Sudore et al [67]	Engaging diverse English- and Spanish-speaking older adults in advance care planning: the PREPARE randomized clinical trial	
Sudore et al [2,66]	Effect of the PREPARE website vs an easy-to-read advance directive on advance care planning documentation and engagement among veterans	
Cresswell et al [65]	Evaluation of an advance care planning web-based resource: applicability for cancer treatment patients	
Ouchi et al [64]	Preparing older adults with serious illness to formulate their goals for medical care in the emergency department	
Sudore et al [63]	A novel website to prepare diverse older adults for decision-making and advance care planning: a pilot study	
The Letter project Advance Directive [55]		Evidence obtained from at least one properly designed randomized controlled trial
Periyakoil et al [72]	A randomized controlled trial comparing the letter project advance directive to traditional advance directive	

[a]Hierarchy of evidence from the National Health and Medical Research Council was assessed per tool. Level I: evidence obtained from a systematic review of all relevant randomized controlled trials; level II: evidence obtained from at least one properly designed randomized controlled trial; level III-1: evidence obtained from well-designed pseudorandomized controlled trials (alternate allocation or some other method); level III-2: evidence obtained from comparative studies with concurrent controls and allocation not randomized (cohort studies), case control studies, or interrupted time series with a control group; level III-3: evidence obtained from comparative studies with historical control, ≥2 single-arm studies, or interrupted time series without a parallel control group; level IV: evidence obtained from case series, either posttest or pretest, and posttest [25,26].

[b]PREPARE: PREPARE for Your Care.

Discussion

Principal Findings

This review included 30 ACP tools developed in North America, Europe, and Oceania. Most tools mention a clear aim (ie, to support reflection and communication, support people in making decisions, or support documenting decisions); however, only 7% (2/30) aimed to achieve all 3 aims. Of the 30 tools, 7 (23%) were developed in collaboration with health care professionals, but only 3 (10%) also involved end users. All tools, except 1, encouraged users to follow steps in a predetermined order to go through the ACP process. With regard to the ACP elements, almost all tools stimulated the user to explore personal values,

goals, and preferences regarding future care; 40% (12/30) provided both information on ACP and its legal frameworks. Of the 3 tools, 2 (67%) also encouraged the user to appoint a proxy decision-maker. Most of the ACP tools had a good readability score; however, the quality of the content varied between 11 and 28 on the QUEST scale. Most of the included ACP tools had not been evaluated in a study.

We found great variety among the tools available on the web in terms of their aims, functionalities, approaches to addressing ACP, and quality. However, the included tools also shared important commonalities.

First, we found that many tools did not provide information about their development process. If they did, they involved health care professionals such as physicians, experts on end-of-life care, ethicists, or lawyers. End users were only involved in the development process of 10% (3/30) of tools, although this is highly recommended in the literature on developing web-based technologies [73-75]. Research on developing new technologies shows that the involvement of end users inevitably yields improvements in usability and quality and ensures that the tool is tailored to the needs of prospective end users [73-75].

Second, most tools stimulated the user to explore personal values, goals, and preferences regarding future care and the last days of life, which is the primary purpose of ACP [1,2], and most also encouraged the appointment of a proxy decision-maker and discussion with family and health care professionals. However, all but 1 tool in this review used an approach in which all users are encouraged to follow a predetermined path or step to go through the ACP process. ACP is a process of exploration; discussion; and recording of preferences, wishes, and decisions. How and when to best perform these aspects of ACP depends, among other things, on personal barriers, needs, preferences, and readiness [76,77]. Using steps in a predetermined order suggests that ACP is linear instead of an iterative process [1]. Moreover, using predetermined steps may not be appropriate for all users, as some may just want to explore possibilities without making decisions, whereas others, for instance, because of their illness, may prefer to focus on anticipatory decision-making [76,78].

Third, with regards to the quality of the content of the included tools, as rated by the QUEST score, we found that 20% (6/30) scored ≥20, of which 17% (1/6; PREPARE) scored a maximum of 28 points. All other tools had a medium to low quality score as they did not refer to any information sources and were not up to date; that is, they had not been updated in at least 5 years. Especially with regard to ACP, having evidence-based information that can be verified by the user is important as people use this information to plan and make health-related care decisions [79,80]. We also found that most of the included tools had not been evaluated in a study. Only the PREPARE tool had been evaluated for its ease of use, effectiveness, acceptability, and understandability. When people look for support in ACP, they may use tools that have a low or even nonexistent level of evidence regarding their usability and effectiveness [13-16,81].

Strengths, Limitations, and Future Research

This is the first systematic review to provide an overview of interactive web-based ACP support tools available on the internet for the general population. Previous reviews have provided an overview of web-based ACP support tools that could be found in the peer-reviewed literature and emphasized the absence of an overview of those available in the gray literature [7,8]. Furthermore, our review is the first to assess the quality of the content of web-based ACP support tools [7,8]. This study had some limitations. First, as we limited our third search strategy (ie, search engine) to 150 *hits* per search combination, there is a possibility that some existing tools were not included in our review. However, we consider this unlikely, as our search was conducted systematically by 2 researchers using a broad range of search terms, and we searched more *hits* than recommended (recommended 100) by Godin et al [17]. We also consulted experts and asked them whether they knew of any other tools that we had not yet found. Furthermore, although we archived the tools when performing data synthesis using Archive.is, there is a possibility that new content or functionalities have been added to the tools included in our review or that new tools have been released since the search was conducted.

For future developments, users should be involved in aligning their preferences and needs with the content and functionalities of the tools. Involving users early in the development process can improve the usability of tools and increase their uptake [82]. In addition, we would recommend following a thorough design process using existing road maps when developing new ACP tools, rigorously evaluating through usability and effectiveness testing before deployment, and transparently reporting on development and evaluation. Second, the ACP content provided to users should be regularly updated and supported by sources; hence, we recommend that the content in ACP tools should be substantiated with the most recent scientific literature. Moreover, future research should focus on how ACP tools are used by the general population and how they can support ACP in the medical context; that is, between patients and health care professionals. Finally, it would be interesting to know how the general population would assess these tools.

Conclusions

There are numerous interactive web-based ACP support tools that are publicly available, varying in terms of their characteristics, functionalities, readability, quality of content, and level of evidence. Most tools were not codeveloped with end users; were of low or medium quality; and, with a few exceptions, had not been evaluated in research. In the future, users should be involved in the development of ACP support tools, and their content should be substantiated by scientific evidence. In addition, we recommend that developers should follow a rigorous design process and evaluate the usability and effectiveness of tools before their deployment. Future research should focus on how tools are used by the general population and how they can support ACP in the medical context; that is, between patients and health care professionals.

Acknowledgments

The authors would like to thank all the experts for their time and effort in assisting in data collection and Jane Ruthven for English editing.

This study is part of the project *CAPACITY: Flanders Project to Develop Capacity in Palliative Care Across Society*, a collaboration among the Vrije Universiteit Brussels, Ghent University, and the Catholic University Leuven, Belgium. This study was supported by a grant from the Research Foundation, Flanders (file number S002219N).

Authors' Contributions

All authors have substantially contributed to the research and approved the version to be published. The search, screening, selection, and synthesis were performed by CD and FM. The original draft was prepared by CD. TS, LP, ADV, CVA, and LVdB were involved in the review and editing of the manuscript. TS and LVdB were involved in supervision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms for the web-based gray literature databases.
[DOCX File , 13 KB - jmir_v24i4e33320_app1.docx]

Multimedia Appendix 2
Search terms for the Google search engine.
[DOCX File , 13 KB - jmir_v24i4e33320_app2.docx]

Multimedia Appendix 3
Search terms for app stores.
[DOCX File , 13 KB - jmir_v24i4e33320_app3.docx]

Multimedia Appendix 4
Tools retrieved per search strategy.
[DOCX File , 13 KB - jmir_v24i4e33320_app4.docx]

Multimedia Appendix 5
Characteristics of the included interactive and web-based advance care planning support tools.
[DOCX File , 22 KB - jmir_v24i4e33320_app5.docx]

Multimedia Appendix 6
Functionalities of included interactive and web-based advance care planning support tools.
[DOCX File , 23 KB - jmir_v24i4e33320_app6.docx]

Multimedia Appendix 7
Advance care planning elements addressed in interactive web-based advance care planning support tools.
[DOCX File , 22 KB - jmir_v24i4e33320_app7.docx]

References

1. Rietjens JA, Sudore RL, Connolly M, van Delden JJ, Drickamer MA, Droger M, European Association for Palliative Care. Definition and recommendations for advance care planning: an international consensus supported by the European Association for Palliative Care. Lancet Oncol 2017 Sep;18(9):e543-e551. [doi: 10.1016/S1470-2045(17)30582-X] [Medline: 28884703]
2. Sudore RL, Lum HD, You JJ, Hanson LC, Meier DE, Pantilat SZ, et al. Defining advance care planning for adults: a consensus definition from a multidisciplinary Delphi panel (S740). J Pain Symptom Manage 2017 May;53(5):821-32.e1 [FREE Full text] [doi: 10.1016/j.jpainsymman.2016.12.331] [Medline: 28062339]
3. Sinclair JB, Oyebode JR, Owens RG. Consensus views on advance care planning for dementia: a Delphi study. Health Soc Care Community 2016 Mar;24(2):165-174. [doi: 10.1111/hsc.12191] [Medline: 25684278]
4. Van den Block L. Advancing research on advance care planning in dementia. Palliat Med 2019 Mar;33(3):259-261. [doi: 10.1177/0269216319826411] [Medline: 30758274]

5. Prince-Paul M, DiFranco E. Upstreaming and normalizing advance care planning conversations-a public health approach. Behav Sci (Basel) 2017 Apr 12;7(2):18 [FREE Full text] [doi: 10.3390/bs7020018] [Medline: 28417931]
6. Tishelman C, Eneslätt M, Menkin ES, Van Den Block L. Tishelman et al's response to Morrison: advance directives/care planning: clear, simple, and wrong (DOI: 10.1089/jpm.2020.0272). J Palliat Med 2021 Jan;24(1):16-17. [doi: 10.1089/jpm.2020.0540] [Medline: 33393888]
7. Butler M, Ratner E, McCreedy E, Shippee N, Kane RL. Decision aids for advance care planning: an overview of the state of the science. Ann Intern Med 2014 Sep 16;161(6):408-418 [FREE Full text] [doi: 10.7326/M14-0644] [Medline: 25069709]
8. van der Smissen D, Overbeek A, van Dulmen S, van Gemert-Pijnen L, van der Heide A, Rietjens JA, et al. The feasibility and effectiveness of web-based advance care planning programs: scoping review. J Med Internet Res 2020 Mar 17;22(3):e15578 [FREE Full text] [doi: 10.2196/15578] [Medline: 32181750]
9. Narang AK, Wright AA, Nicholas LH. Trends in advance care planning in patients with cancer: results from a national longitudinal survey. JAMA Oncol 2015 Aug;1(5):601-608 [FREE Full text] [doi: 10.1001/jamaoncol.2015.1976] [Medline: 26181909]
10. Green MJ, Levi BH. The era of "e": the use of new technologies in advance care planning. Nurs Outlook 2012;60(6):376-83.e2 [FREE Full text] [doi: 10.1016/j.outlook.2012.08.005] [Medline: 23141197]
11. De Vito Dabbs A, Myers BA, Mc Curry KR, Dunbar-Jacob J, Hawkins RP, Begey A, et al. User-centered design and interactive health technologies for patients. Comput Inform Nurs 2009;27(3):175-183 [FREE Full text] [doi: 10.1097/NCN.0b013e31819f7c7c] [Medline: 19411947]
12. Cline RJ, Haynes KM. Consumer health information seeking on the internet: the state of the art. Health Educ Res 2001 Dec;16(6):671-692. [doi: 10.1093/her/16.6.671] [Medline: 11780707]
13. Berland GK, Elliott MN, Morales LS, Algazy JI, Kravitz RL, Broder MS, et al. Health information on the internet: accessibility, quality, and readability in English and Spanish. JAMA 2001;285(20):2612-2621 [FREE Full text] [doi: 10.1001/jama.285.20.2612] [Medline: 11368735]
14. Seçkin G, Yeatts D, Hughes S, Hudson C, Bell V. Being an informed consumer of health information and assessment of electronic health literacy in a national sample of internet users: validity and reliability of the e-HLS Instrument. J Med Internet Res 2016 Jul 11;18(7):e161 [FREE Full text] [doi: 10.2196/jmir.5496] [Medline: 27400726]
15. Powell J, Inglis N, Ronnie J, Large S. The characteristics and motivations of online health information seekers: cross-sectional survey and qualitative interview study. J Med Internet Res 2011 Feb 23;13(1):e20 [FREE Full text] [doi: 10.2196/jmir.1600] [Medline: 21345783]
16. Fahy E, Hardikar R, Fox A, Mackay S. Quality of patient health information on the Internet: reviewing a complex and evolving landscape. Australas Med J 2014;7(1):24-28 [FREE Full text] [doi: 10.4066/AMJ.2014.1900] [Medline: 24567763]
17. Godin K, Stapleton J, Kirkpatrick SI, Hanning RM, Leatherdale ST. Applying systematic review search methods to the grey literature: a case study examining guidelines for school-based breakfast programs in Canada. Syst Rev 2015 Oct 22;4:138 [FREE Full text] [doi: 10.1186/s13643-015-0125-0] [Medline: 26494010]
18. Readability score | Readability test | Reading level calculator. Readable. URL: https://readable.com/ [accessed 2022-02-17]
19. Leesniveau. Accessibility Foundation. URL: https://www.accessibility.nl/tools/leesniveau [accessed 2022-02-17]
20. Bletsis PP, Egeler S, van der Lei B, Lee BT, Lin SJ. Readability of Dutch online patient-directed health information on breast reconstruction. Eur J Plast Surg 2019 Aug;42(4):343-350. [doi: 10.1007/s00238-019-1498-4]
21. Meppelink CS, van Weert JC, Brosius A, Smit EG. Dutch health websites and their ability to inform people with low health literacy. Patient Educ Couns 2017 Nov;100(11):2012-2019. [doi: 10.1016/j.pec.2017.06.012] [Medline: 28624261]
22. North B, Piccardo N. Common European framework of reference for languages: learning, teaching, assessment?; companion volume. Strasbourg, France: Council of Europe Publishing; Jul 2020:55-59.
23. Robillard JM, Jun JH, Lai JA, Feng TL. The QUEST for quality online health information: validation of a short quantitative tool. BMC Med Inform Decis Mak 2018 Oct 19;18(1):87 [FREE Full text] [doi: 10.1186/s12911-018-0668-9] [Medline: 30340488]
24. Roberts L. Health information and the Internet: the 5 Cs website evaluation tool. Br J Nurs 2010;19(5):322-325. [doi: 10.12968/bjon.2010.19.5.47075] [Medline: 20335904]
25. National Health and Medical Research Council. A guide to the development, implementation and evaluation of clinical practice guidelines. Canberra, Australia: Commonwealth of Australia; 1999:15-17.
26. Anderson CF. Guide to clinical preventive services: an assessment of the effectiveness of 169 interventions. Mayo Clin Proc 1989 Dec 1;64(12):1566-1567. [doi: 10.1016/s0025-6196(12)65716-4]
27. Make web content printer friendly. Health Literacy Online. URL: https://health.gov/healthliteracyonline/display/section-3-10/ [accessed 2022-01-12]
28. ACDCare. GP Partners Australia. URL: http://archive.today/2021.04.16-082031/https://apps.apple.com/uy/app/acdcare/id1500681626 [accessed 2021-03-15]
29. MyDirectives. URL: http://archive.today/2021.04.16-083317/https://mydirectives.com/ [accessed 2021-03-15]
30. Be My Voice. Capital Health Network. URL: http://archive.today/2021.04.16-083021/http://www.bemyvoice.com.au/Home.aspx [accessed 2021-03-15]

31. My Living Will. Vital Decisions. URL: http://archive.today/2021.04.16-083023/https://www.mylivingwill.org.uk/static/wizard?Type=Statement [accessed 2021-03-15]
32. Speak up. Canadian Hospice Palliative Care Association. URL: http://archive.today/2021.04.16-083319/https://www.advancecareplanning.ca/my-plan/workbook/ [accessed 2021-03-15]
33. Considering your own future health care. Health Quality & Safety Commission, New Zealand. 2018. URL: http://archive.today/2021.04.16-082903/https://www.hqsc.govt.nz/our-programmes/advance-care-planning/publications-and-resources/ [accessed 2021-03-15]
34. PREPARE for your care. The Regents of the University of California. 2018. URL: http://archive.today/2021.04.16-083623/https://prepareforyourcare.org/welcome [accessed 2021-03-15]
35. Advance Care Planning: Should I Have Artificial Hydration and Nutrition? Healthwise and Alberta. 2019. URL: http://archive.today/2021.04.16-082516/https://www.healthlinkbc.ca/health-topics/tu4431 [accessed 2021-03-15]
36. Advance Care Planning: should I receive CPR and life support? Healthwise and Alberta. 2019. URL: http://archive.today/2021.04.16-082531/https://www.healthlinkbc.ca/health-topics/tu2951 [accessed 2021-03-15]
37. Advance Care Planning: should I stop kidney dialysis? Healthwise and Alberta. 2019. URL: http://archive.today/2021.04.16-082541/https://myhealth.alberta.ca/Health/Pages/conditions.aspx?hwid=tu6095 [accessed 2021-03-15]
38. Advance Care Planning: should I stop treatment that prolongs my life? Healthwise and Alberta. 2019. URL: http://archive.today/2021.04.16-082605/https://myhealth.alberta.ca/Health/pages/conditions.aspx?hwid=tu1430&lang=en-ca [accessed 2021-03-15]
39. Vroegtijdige zorgplanning (voor patiënt) - Beslishulp. Centrum voor Evidence-Based Medicine (Cebam). URL: http://archive.today/2021.04.16-083644/https://www.gezondheidenwetenschap.be/beslissingshulp/vroegtijdige-zorgplanning [accessed 2021-03-15]
40. Cake. URL: http://archive.today/2021.04.16-084034/https://www.joincake.com/ [accessed 2021-03-15]
41. Dementia Values and Priorities Tool. Compassion & Choices. URL: http://archive.today/2021.04.16-084038/https://values-tool.compassionandchoices.org/ [accessed 2021-03-15]
42. Dying to Talk - Discussion Starters. Palliative Care Australia. URL: https://dyingtotalk.org.au/discussion-starters/ [accessed 2021-03-15]
43. Everplans. 2012. URL: http://archive.today/2021.04.16-084153/https://webapps.everplans.com/%23/auth/sign-up [accessed 2021-03-15]
44. Five Wishes. URL: http://archive.today/2021.04.16-084416/https://fivewishes.org/shop/order/product/five-wishes-online [accessed 2021-03-15]
45. Go Wish card game online version. CODA Alliance. 2018. URL: http://archive.today/2021.04.16-083901/http://www.gowish.org/gowish/gowish.html [accessed 2021-03-15]
46. Lets Think Ahead - My ACP. University of the West Scotland. URL: http://archive.today/2021.04.16-082949/https://www.scet.co/fullscreen-page/comp-iqtuyfne/e2469d5e-3e56-4452-803d-ea83e55b2f32/4/%3Fi%3D4 [accessed 2021-03-22]
47. Compassion in Dying charity. My Decisions. URL: http://archive.today/2021.04.16-084547/https://mydecisions.org.uk/ [accessed 2021-03-15]
48. My Living Voice. Vital Decisions. URL: http://archive.today/2021.04.16-084549/https://vitaldecisions.mylivingvoice.com/ [accessed 2021-03-15]
49. My Values. Barwon Health. URL: http://archive.today/2021.04.16-083754/https://www.myvalues.org.au/ [accessed 2021-03-15]
50. MyWishes. URL: https://www.app.mywishes.co.uk/register [accessed 2021-03-15]
51. NVLivingWill Living Will Online Completion. Nevada Center for Ethics & Health Policy. URL: http://archive.today/2021.04.16-084642/http://www.nvlivingwill.com/ [accessed 2021-03-15]
52. Bureau MORBidee. Oog in Oog. Google Play. URL: http://archive.today/2021.04.16-084820/https://play.google.com/store/apps/details?id=com.eyetoeye.eye_to_eye&hl=en_IE [accessed 2021-03-15]
53. Plan your Life Span. Northwestern University. URL: http://archive.today/2021.04.16-084828/https://www.planyourlifespan.org/ [accessed 2021-03-15]
54. Prepare an advance directive. People's Law School. 2019. URL: https://www.peopleslawschool.ca/prepare-advance-directive/ [accessed 2021-03-15]
55. Use the Letter Project form to complete your Advance Directives in Minutes!. Stanford Medicine Letter Project. 2015. URL: https://med.stanford.edu/letter/advancedirective.html [accessed 2021-03-15]
56. Tijdig nadenken over het levenseinde [Timely thinking about the end of life]. NPV Zorg voor het leven. URL: https://ehealth-npvzorg.serviant.nl/login/?back=https://ehealth-npvzorg.serviant.nl/client/module%3FtrajectId%3D3692%26q%3DlTTLQmY [accessed 2021-03-15]
57. Verken uw wensen voor zorg en behandeling [Explore your care and treatment needs]. Dutch College of General Practitioners, Erasmus MC and thuisarts.nl. URL: https://www.thuisarts.nl/keuzehulp/verken-uw-wensen-voor-zorg-en-behandeling [accessed 2021-03-15]
58. Klugman CM, Usatine RP. An evaluation of 2 online advance directive programs. Am J Hosp Palliat Care 2013 Nov;30(7):657-663. [doi: 10.1177/1049909112463116] [Medline: 23087248]

59. Holland DE, Vanderboom CE, Dose AM, Ingram CJ, Delgado A, Austin CM, et al. Nurse-led patient-centered advance care planning in primary care. J Hosp Palliat Nurs 2017;19(4):368-375. [doi: 10.1097/NJH.0000000000000358]
60. Fine RL, Yang Z, Spivey C, Boardman B, Courtney M. Early experience with digital advance care planning and directives, a novel consumer-driven program. Proc (Bayl Univ Med Cent) 2016 Jul;29(3):263-267 [FREE Full text] [doi: 10.1080/08998280.2016.11929430] [Medline: 27365867]
61. Lindquist LA, Ramirez-Zohfeld V, Sunkara PD, Forcucci C, Campbell DS, Mitzen P, et al. PlanYourLifeSpan.org - an intervention to help seniors make choices for their fourth quarter of life: Results from the randomized clinical trial. Patient Educ Couns 2017 Nov;100(11):1996-2004 [FREE Full text] [doi: 10.1016/j.pec.2017.06.028] [Medline: 28689855]
62. Ramirez-Zohfeld V, Seltzer A, Ramirez A, Muhammad R, Lindquist LA. Longitudinal follow-up of long-term care planning using PlanYourLifespan.org. J Appl Gerontol 2021 May;40(5):536-540 [FREE Full text] [doi: 10.1177/0733464820943066] [Medline: 32686545]
63. Sudore RL, Knight SJ, McMahan RD, Feuz M, Farrell D, Miao Y, et al. A novel website to prepare diverse older adults for decision making and advance care planning: a pilot study. J Pain Symptom Manage 2014 Apr;47(4):674-686 [FREE Full text] [doi: 10.1016/j.jpainsymman.2013.05.023] [Medline: 23972574]
64. Ouchi K, Knabben V, Rivera-Reyes L, Ganta N, Gelfman LP, Sudore R, et al. Preparing older adults with serious illness to formulate their goals for medical care in the emergency department. J Palliat Med 2017 Apr 01;20(4):404-408 [FREE Full text] [doi: 10.1089/jpm.2016.0109] [Medline: 27797638]
65. Cresswell MA, Robinson CA, Fyles G, Bottorff JL, Sudore R. Evaluation of an advance care planning web-based resource: applicability for cancer treatment patients. Support Care Cancer 2018 Mar;26(3):853-860. [doi: 10.1007/s00520-017-3901-7] [Medline: 28929291]
66. Sudore RL, Boscardin J, Feuz MA, McMahan RD, Katen MT, Barnes DE. Effect of the PREPARE website vs an easy-to-read advance directive on advance care planning documentation and engagement among veterans: a randomized clinical trial. JAMA Intern Med 2017 Aug 01;177(8):1102-1109 [FREE Full text] [doi: 10.1001/jamainternmed.2017.1607] [Medline: 28520838]
67. Sudore RL, Schillinger D, Katen MT, Shi Y, Boscardin WJ, Osua S, et al. Engaging diverse English- and Spanish-speaking older adults in advance care planning: the PREPARE randomized clinical trial. JAMA Intern Med 2018 Dec 01;178(12):1616-1625 [FREE Full text] [doi: 10.1001/jamainternmed.2018.4657] [Medline: 30383086]
68. Zapata C, Lum HD, Wistar E, Horton C, Sudore RL. Feasibility of a video-based advance care planning website to facilitate group visits among diverse adults from a safety-net health system. J Palliat Med 2018 Jun;21(6):853-856. [doi: 10.1089/jpm.2017.0476] [Medline: 29461918]
69. Lum HD, Barnes DE, Katen MT, Shi Y, Boscardin J, Sudore RL. Improving a full range of advance care planning behavior change and action domains: the PREPARE randomized trial. J Pain Symptom Manage 2018 Oct;56(4):575-81.e7 [FREE Full text] [doi: 10.1016/j.jpainsymman.2018.06.007] [Medline: 29940209]
70. Freytag J, Street Jr RL, Barnes DE, Shi Y, Volow AM, Shim JK, et al. Empowering older adults to discuss advance care planning during clinical visits: the PREPARE randomized trial. J Am Geriatr Soc 2020 Jun;68(6):1210-1217 [FREE Full text] [doi: 10.1111/jgs.16405] [Medline: 32157684]
71. Howard M, Robinson CA, McKenzie M, Fyles G, Sudore RL, Andersen E, et al. Effect of an interactive website to engage patients in advance care planning in outpatient settings. Ann Fam Med 2020 Mar;18(2):110-117 [FREE Full text] [doi: 10.1370/afm.2471] [Medline: 32152014]
72. Periyakoil VS, Neri E, Kraemer H. A randomized controlled trial comparing the letter project advance directive to traditional advance directive. J Palliat Med 2017 Sep;20(9):954-965. [doi: 10.1089/jpm.2017.0066] [Medline: 28471713]
73. Fischer B, Peine A, Östlund B. The importance of user involvement: a systematic review of involving older users in technology design. Gerontologist 2020 Sep 15;60(7):e513-e523 [FREE Full text] [doi: 10.1093/geront/gnz163] [Medline: 31773145]
74. Shah SG, Robinson I. Benefits of and barriers to involving users in medical device technology development and evaluation. Int J Technol Assess Health Care 2007;23(1):131-137. [doi: 10.1017/S0266462307051677] [Medline: 17234027]
75. de Beurs D, van Bruinessen I, Noordman J, Friele R, van Dulmen S. Active involvement of end users when developing web-based mental health interventions. Front Psychiatry 2017;8:72 [FREE Full text] [doi: 10.3389/fpsyt.2017.00072] [Medline: 28515699]
76. Zwakman M, Jabbarian LJ, van Delden J, van der Heide A, Korfage IJ, Pollock K, et al. Advance care planning: a systematic review about experiences of patients with a life-threatening or life-limiting illness. Palliat Med 2018 Sep;32(8):1305-1321 [FREE Full text] [doi: 10.1177/0269216318784474] [Medline: 29956558]
77. Piers R, Albers G, Gilissen J, De Lepeleire J, Steyaert J, Van Mechelen W, et al. Advance care planning in dementia: recommendations for healthcare professionals. BMC Palliat Care 2018 Jun 21;17(1):88 [FREE Full text] [doi: 10.1186/s12904-018-0332-2] [Medline: 29933758]
78. Slyer J, Archibald E, Moyo F, Truglio-Londrigan M. Advance care planning and anticipatory decision making in patients with Alzheimer disease. Nurse Pract 2018 Jun 11;43(6):23-31. [doi: 10.1097/01.NPR.0000532763.68509.e4] [Medline: 29757832]

79. 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]
80. Lenhart A. Pew Research Center. 2007 Jun 27. URL: https://www.pewresearch.org/internet/2007/06/27/about-the-pew-internet-american-life-project/ [accessed 2021-03-30]
81. Kaufman D, Roberts WD, Merrill J, Lai TY, Bakken S. Applying an evaluation framework for health information system design, development, and implementation. Nurs Res 2006;55(2 Suppl):S37-S42. [doi: 10.1097/00006199-200603001-00007] [Medline: 16601633]
82. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]

Abbreviations

ACP: advance care planning
CEFR: Common European Framework of Reference for Languages
PREPARE: PREPARE for Your Care
PROSPERO: International Prospective Register of Systematic Reviews
QUEST: Quality Evaluation Scoring Tool

Edited by T Leung; submitted 02.09.21; peer-reviewed by C Sinclair, K Reuter; comments to author 19.12.21; revised version received 01.02.22; accepted 23.02.22; published 20.04.22.

Please cite as:
Dupont C, Smets T, Monnet F, Pivodic L, De Vleminck A, Van Audenhove C, Van den Block L
Publicly Available, Interactive Web-Based Tools to Support Advance Care Planning: Systematic Review
J Med Internet Res 2022;24(4):e33320
URL: https://www.jmir.org/2022/4/e33320
doi:10.2196/33320
PMID:35442207

Review

Unguided Computer-Assisted Self-Help Interventions Without Human Contact in Patients With Obsessive-Compulsive Disorder: Systematic Review and Meta-analysis

Hissei Imai[1], MD, PhD; Aran Tajika[1], MD, PhD; Hisashi Narita[2,3], MD, PhD; Naoki Yoshinaga[4], PhD; Kenichi Kimura[5], MD; Hideki Nakamura[6], MA; Nozomi Takeshima[1], MD, PhD; Yu Hayasaka[1], MD, PhD; Yusuke Ogawa[7], MD, PhD; Toshi Furukawa[1], MD, PhD

[1]Department of Health Promotion and Human Behavior, Graduate School of Medicine, School of Public Health, Kyoto University, Kyoto, Japan

[2]Department of Psychiatry, Graduate School of Medicine, Hokkaido University, Sapporo, Japan

[3]Department of Neurology, Graduate School of Medicine, Hokkaido University, Sapporo, Japan

[4]School of Nursing, Faculty of Medicine, University of Miyazaki, Miyazaki, Japan

[5]Department of Psychiatry, Hakodate Watanabe Hospital, Hakodate, Japan

[6]Department of Cognitive Behavioral Physiology, Graduate School of Medicine, Chiba University, Chiba, Japan

[7]Department of Healthcare Epidemiology, Graduate School of Medicine, School of Public Health, Kyoto University, Kyoto, Japan

Corresponding Author:
Hissei Imai, MD, PhD
Department of Health Promotion and Human Behavior
Graduate School of Medicine, School of Public Health
Kyoto University
Yoshida Konoe-cho, Sakyo-ku
Kyoto, 606-8501
Japan
Phone: 81 75 753 9491
Fax: 81 75 753 4641
Email: ihits@hotmail.com

Abstract

Background: Computer-assisted treatment may reduce therapist contact and costs and promote client participation. This meta-analysis examined the efficacy and acceptability of an unguided computer-assisted therapy in patients with obsessive-compulsive disorder (OCD) compared with a waiting list or attention placebo.

Objective: This study aimed to evaluate the effectiveness and adherence of computer-assisted self-help treatment without human contact in patients with OCD using a systematic review and meta-analysis approach.

Methods: Randomized controlled trials with participants primarily diagnosed with OCD by health professionals with clinically significant OCD symptoms as measured with validated scales were included. The interventions included self-help treatment through the internet, computers, and smartphones. We excluded interventions that used human contact. We conducted a search on PubMed, Cochrane Central Register of Controlled Trials, EMBASE, World Health Organization International Clinical Trials Registry Platform, and ClinicalTrials.gov, as well as the reference lists of the included studies. The risk of bias was evaluated using version 2 of the Cochrane risk-of-bias tool for randomized trials. We calculated the standardized mean differences for continuous outcomes and risk ratios for dichotomous outcomes. The primary outcomes were short-term improvement of OCD symptoms measured by validated scales and dropout for any reason.

Results: We included 11 randomized controlled trials with a total of 983 participants. The results indicated that unguided computer-assisted self-help therapy was significantly more effective than a waiting list or psychological placebo (standard mean difference −0.47, 95% CI −0.73 to −0.22). Unguided computer-assisted self-help therapy had more dropouts for any reason than waiting list or psychological placebo (risk ratio 1.98, 95% CI 1.21 to 3.23). However, the quality of evidence was very low because of the risk of bias and inconsistent results among the included studies. The subgroup analysis showed that exposure response and prevention and an intervention duration of more than 4 weeks strengthen the efficacy without worsening acceptability. Only a few studies have examined the interaction between participants and systems, and no study has used gamification. Most

researchers only used text-based interventions, and no study has used a mobile device. The overall risk of bias of the included studies was high and the heterogeneity of results was moderate to considerable.

Conclusions: Unguided computer-assisted self-help therapy for OCD is effective compared with waiting lists or psychological placebo. An exposure response and prevention component and intervention duration of more than 4 weeks may strengthen the efficacy without worsening the acceptability of the therapy.

Trial Registration: PROSPERO (International Prospective Register of Systematic Reviews) CRD42021264644; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=264644

(J Med Internet Res 2022;24(4):e35940) doi:10.2196/35940

KEYWORDS

randomized controlled trial; RCT; information technology; psychotherapy; treatment adherence; anxiety disorder; anxiety; OCD; obsessive-compulsive disorder; systematic review; meta-analysis; mental health; computer-assisted; therapy; efficacy; acceptability; eHealth; mental illness

Introduction

Obsessive-compulsive disorder (OCD) is characterized by intrusive and unwanted thoughts, urges, or images and repetitive behavior or mental acts [1]. Affected patients try to ignore or suppress OCD symptoms; however, it impairs their ability to carry out daily life activities and deteriorates their quality of life (QOL). The median prevalence of OCD in 1 year was 1.0% (IQR 0.6% to 2.0%), and the cost associated with OCD was estimated as $10.6 billion per year in the United States alone [2].

The treatment of OCD involves psychotherapy and pharmacotherapy; however, psychotherapy may be a better treatment for OCD than pharmacotherapy [3]. Patients with psychiatric disorders prefer psychotherapy over pharmacotherapy [4]. Therefore, guidelines such as the National Institute for Health and Care Excellence recommend cognitive behavioral therapy (CBT) as the initial treatment for OCD [5].

Despite the presence of guidelines for the treatment of OCD, there are hindrances to therapy such as poor help-seeking behavior and inaccessible treatment. A study showed that more than half of patients with OCD have not received treatment [6]. Barriers to seeking treatment include shame about the symptoms or about asking for treatment, lack of knowledge regarding resources, and treatment-related inconveniences [7].

Computer and internet-based treatment is a promising way to overcome these barriers. It can reduce therapist contact and costs and promote client participation in therapies conducted in a nonclinical setting [8]. Successful internet-based interventions include engagement by the user for weeks to months. Examples are interactive elements such as prompted personalized feedback, self-monitoring, and assignment [9]. All the interventions contain educational materials and frequently use cognitive behavioral elements [9]. More specifically, computerized therapy for OCD often includes psychoeducation, cognitive elements, and exposure and response prevention (ERP) [10].

Systematic reviews were conducted on studies including computer-assisted treatment for OCD, but there were limitations; Pearcy et al [11] examined self-help intervention against OCD, but they included quasi-randomized controlled trials (RCTs); Firth et al [12] examined smartphone interventions, but the focus was on anxiety disorder as a whole; and Tumur et al [10] examined computer-assisted CBT for OCD, but it included only one substantial program. These studies need to be updated since the research was conducted in 2015, 2016, and 2004, respectively, and particularly because the rate of publication on digital health has been increasing rapidly since 2015 [13].

Excluding therapist contact and therapy using information technology will improve access to treatment. However, the effectiveness and adherence of computer-assisted interventions without human contact has not been examined through systematic review and meta-analysis. Additionally, the influence of several variables should be examined. For example, ERP is an effective and widely used component for OCD treatment, but therapist assistance is suggested to increase its effectiveness [14,15], the duration or number of sessions attended may be influential factors in psychotherapy [16,17], and the effect of device characteristics and their contents, such as gamification and interaction, have not yet been established [18,19].

Therefore, this study aimed to evaluate the effectiveness and adherence of computer-assisted self-help treatment without human contact in patients with OCD using a systematic review and meta-analysis approach.

Methods

Selection Criteria

RCTs with participants primarily diagnosed with OCD according to the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision* (DSM-IV-TR), *and Fifth Edition* (DSM-5) and *International Classification of Diseases, Tenth Revision*, and those who were diagnosed by health professionals and had clinically significant OCD symptoms as measured with validated scales were included. Patients of any age and comorbidities were included. The interventions included self-help treatment through the internet, computers, and smartphones. Sending a digital treatment manual by email was also included because it uses the computer and internet. We excluded interventions that used human contact (except for technical support). We defined human contact as interventions with face-to-face support or interaction with humans on the internet or telephone; self-help means that participants conduct treatment without human contact. Comparisons were made with

respect to a placebo condition, including a psychological placebo and a waiting list. Any cotreatment was allowed if it was provided equally to both groups.

Ethics Approval

We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [20; Multimedia Appendix 1]. The protocol for this systematic review was registered at PROSPERO (International Prospective Register of Systematic Reviews) [CRD42021264644].

Outcomes

The primary outcomes were short-term subjective improvement of OCD symptoms as measured by validated scales such as the Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) and the Obsessive-Compulsive Inventory–Revised (OCI-R) and dropout for any reason at posttreatment. We defined short term as a period of 6 months.

Secondary outcomes were short-term response rate defined by validated scales and anxiety, depression, and QOL measured by validated scales. These outcomes measured at long term were also included in the secondary outcomes. We defined long term as a period greater than 6 months and gave priority to the longest end point.

Search Methods

We conducted a search on July 28, 2021, in PubMed, Cochrane Central Register of Controlled Trials, EMBASE, World Health Organization International Clinical Trials Registry Platform, and ClinicalTrials.gov, as well as the reference lists of the included studies (Multimedia Appendix 2). We conducted a grey literature search in devices@FDA, a catalog of cleared and approved medical device information. We applied no search restrictions on date, language, or publication status.

Selection of Studies and Data Extraction

Two authors independently examined the titles and abstracts of the references identified in the search and included them in the second screening if at least 1 author judged them to be included. We then obtained and examined the full text of the included studies using the first screening process. Finally, we included the studies that both reviewers felt should be included. If the 2 authors disagreed after a discussion, a third author was consulted to make a decision. We conducted data extraction in the same way as in the second screening process. We contacted the authors of the studies to obtain additional data or further clarification if needed.

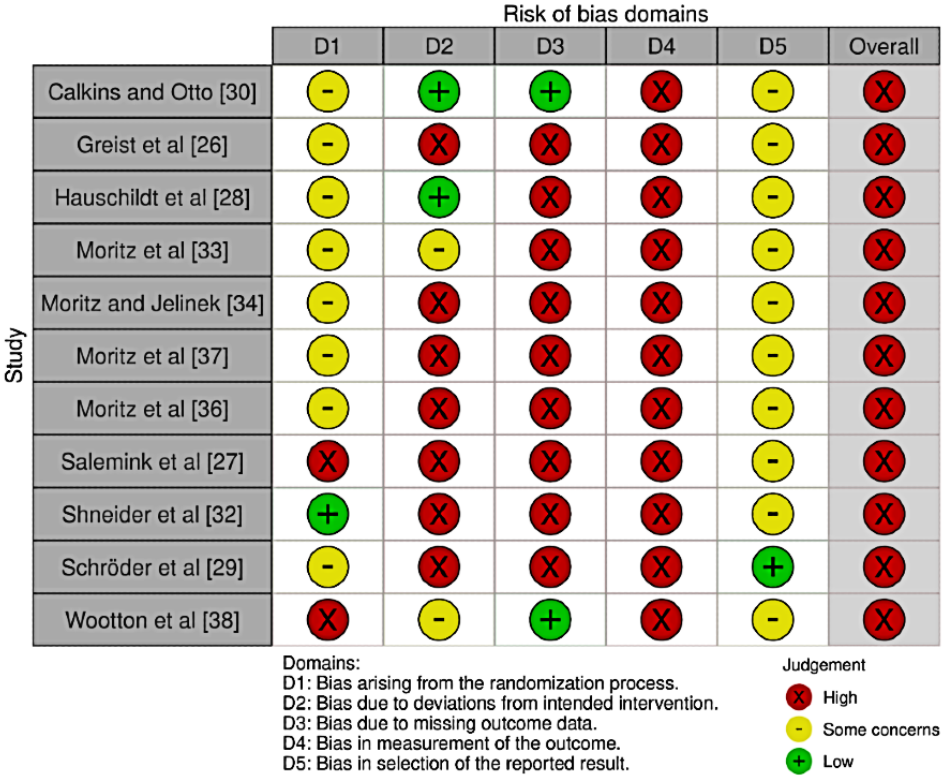
Measurement of Outcomes

We calculated the standardized mean differences (SMDs) and their 95% confidence intervals for continuous outcomes and risk ratios and their 95% confidence intervals for dichotomous outcomes. We used a random effects model.

Assessment of Risk of Bias

The risk of bias was evaluated using version 2 of the Cochrane risk-of-bias tool for randomized trials (Figure 1) [21]. The risk-of-bias tool assesses the following domains: bias arising from the randomization process, bias due to deviations from intended interventions, bias due to missing outcome data, bias in the measurement of the outcome, and bias in the selection of the reported results. Each bias was assigned 1 of 3 levels: low risk of bias, some concerns, or high risk of bias. The risk of bias of each studies was presented in traffic light plots.

Figure 1. Risk of bias in included studies.



Analysis

We assessed heterogeneity using the I^2 statistic. We interpreted the I^2 value as in the Cochrane Handbook for Systematic Review of Interventions (0%-40% might not be important, 30%-60% may represent moderate heterogeneity, 50%-90% may represent substantial heterogeneity, and 75%-100% may represent considerable heterogeneity). The source was investigated if significant heterogeneity was observed. Publication bias was evaluated by visual inspection of the funnel plot if at least 10 studies were included in the analysis. We calculated a pooled standard deviation for studies where standard deviations were not reported [22]. The results were compared using a sensitivity analysis with or without studies of imputed standard deviations and study targeted to children and adolescent. All analyses were conducted with Review Manager (version 5.4, The Cochrane Collaboration) software.

We performed the following subgroup analyses:

- By type of psychotherapy included in the intervention (with or without ERP), as a systematic review showed the efficacy of ERP against OCD [14]
- By intervention devices, as we hypothesized that device characteristics would influence the results. We planned to include portability with mobile phone, interaction with computer, and gamification. Portability may make it easy for participants to conduct ERP. Interaction and gamification may motivate participants to continue the intervention. However, no study included in this review used a mobile phone or gamification. As a result, we conducted a subgroup analysis with and without interaction with the system and intervention using a treatment manual via email or computer display. Interaction with the system means that participants can automatically get responses from a computer system without human contact
- By study duration or number of sessions to examine the influence of duration. We conducted an analysis on studies with a duration equal to or less than 4 weeks and studies over 4 weeks, as the median and mode of the included study duration was 4 weeks. We could not conduct subgroup analysis by session because no studies reported the number of sessions conducted
- By type of control arm, conducted as post hoc analysis, as a recent study showed that effect size may differ according to the control condition [23]. The subgroup differences were interpreted as suggestive when $P<.10$, in consideration of the small number of included studies and difficulty finding subgroup interactions.

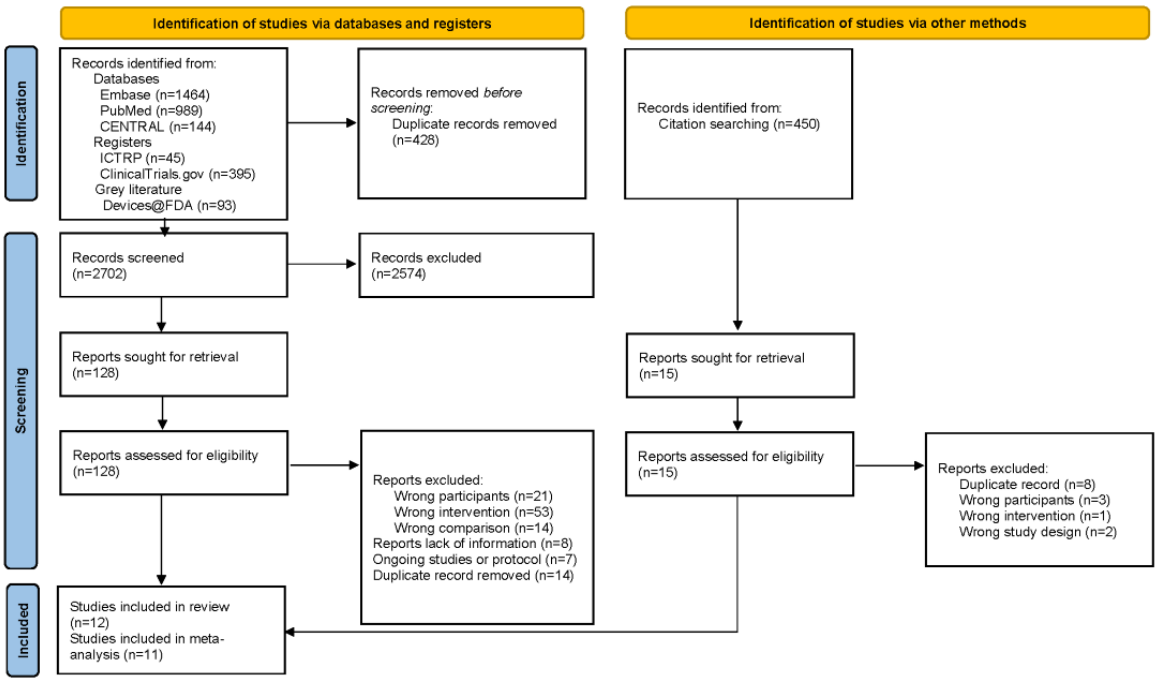
The quality of evidence for primary outcomes was evaluated according to the GRADE (Grading of Recommendations, Assessment, Development, and Evaluations) rating [24].

Results

Search Results

We identified 3130 references and excluded 2574 studies after assessing the title and abstracts. We retrieved 128 full-text papers, excluded 117 studies, and included 11 studies. We inspected the citations of the 11 studies and found 1 study to include. Finally, a total of 12 studies were included in the review, but we could not obtain additional data from the author of 1 study [25]. As a result, 11 studies with a total of 983 participants were included in the meta-analysis (Figure 2).

Figure 2. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow chart. CENTRAL: Cochrane Central Register of Controlled Trials; ICTRP: International Clinical Trials Registry Platform.



Characteristics of Included Studies

As seen in Table 1, all included studies were parallel group, individually RCTs. One was a 3-armed study [26], while others were 2-armed. The mean sample size per arm was 45 (range 9-100).

Participants were recruited in European countries in 3 studies [27-29], in North America in 2 studies [26,30], and cross-continental in 1 study [31], but recruitment method was unclear in the other studies because it was done through the internet [32-36]. Diagnosis was based on DSM-IV-TR in 2 studies [27,28], DSM-IV in 1 study [26], health professional diagnosis using unclear diagnostic criteria in 5 studies [32-34,36,37], and the OCD symptom scale in 3 studies [29,30,38]. The proportion of women ranged from 42% to 83%. The mean age in a study targeting adolescents was 15 years [27], while others ranged from 28 to 41 years.

Interventions included computer-assisted cognitive training [30], behavioral therapy [26], metacognitive training [28,33,36], association splitting [39], inference-based therapy [37], competitive memory training [32], and CBT [29,38]. Of the included studies, Moritz et al [36] conducted 7 of them; however, only 3 used the same or a revised version of the intervention among them [28,33]. As for the component of therapy, 6 studies included exposure therapy [26,27,29,32,36,38]. ERP was used in 5 studies [26,27,29,36,38], and interoceptive exposure was used in 1 study [32]. Two studies explicitly examined the interaction between the system and the participants [26,29]. Three studies used a computer display that presented text-based online slides [38], text, video, audio elements, photos, illustrations [29], and a scenario with missing words that patients filled in [27]. These studies did not use gamification and did not include mobile devices. No studies used combination therapy; all but 1 study [30] allowed adjunctive medication.

Table 1. Characteristics of included studies.

Author; year; citation; country; study design	Participants		Interventions	Outcomes
	Diagnosis; sex; medications	Age (years), mean (SD); baseline severity, mean (SD)		
Calkins et al [30]; North America; RCT[a]	Dx: OCI-R[b]>15; Sex: CCT[c] arm: 54.2% women; PVT[d] arm 62.5%; meds[e]: unclear	Age: CCT arm 27.9 (SD 14.1); PVT arm 30 (SD 13.8); Severity: OCI-R CCT arm 28.9 (SD 11.1); PVT arm 30.8 (SD 0.9)	CCT n=24; duration 2 weeks; exposure: no; cognitive modification: no; device: computer; interaction: no; gamification: no but a kind of task; PVT n=24; duration 2 weeks	OCI-R; BDI-II[f]; PANAS[g]; PSWQ[h]
Greist et al [26]; North America; RCT	Dx: DSM-IV[i]; Sex: 42% women; meds: yes	Age: 39 (SD 12); Severity (Y-BOCS[j]): BT STEPS[k] arm 24.6 (SD 4.3); systematic relaxation arm 25.8 (SD 5.1)	BT STEPS n=74; duration 10 weeks; exposure: yes (ERP[l]); cognitive modification: unclear; device: computer-driven interactive voice response system and workbook; interaction: yes; gamification: no; Systematic relaxation n=75; duration 10 weeks; Clinician-guided behavior therapy	Y-BOCS; PGI-I[m]; CGI scale[n]; WSAS[o]; HAM-D[p]; SRI[q] medication status; treatment expectations; treatment satisfaction
Haushildt et al 2016 [28]; Europe; RCT	Dx: DSM-IV; Sex: myMCT arm: 67.2% women; psychoeducation arm: 67.2%; Meds: yes	Age: myMCT arm 38.41 (SD 11.61); psychoeducation arm 39.64 (SD 9.88); Severity (Y-BOCS): myMCT arm 22.56 (SD 6.58); psychoeducation arm 21.45 (SD 6.42)	myMCT n=64; duration 4 weeks; Device: pdf file through email; exposure: no; cognitive modification: yes (metacognitive training, association splitting); interaction: no; gamification: no; Psychoeducation n=64; duration 4 weeks	Y-BOCS, BDI[r], OBQ[s]
Moritz et al 2010 [33]; recruited from internet forums; RCT	Dx: OCD[t] diagnosis made by health care professionals; Sex: myMCT arm: 62.8% women; waiting list arm: 72.1%; Meds: yes	Age: myMCT arm 34.95 (SD 11.87); waiting list arm 34.09 (SD 9.41); Severity (Y-BOCS): myMCT arm 18.6 (SD 6.86); waiting list arm 19.98 (SD 5.9)	myMCT n=43; duration 4 weeks; device: pdf file through email; exposure: no; cognitive modification: yes (metacognitive training); Waiting list n=43; duration 4 weeks.	Y-BOCS, OCI-R, BDI-SF[u]
Moritz & Jelinek 2011 [34]; recruited from internet forums; RCT	Dx: OCD diagnosis made by health care professionals; Sex: AS[v] arm: 56.5% women; waiting list arm: 78.3%; Meds: yes	Age: AS arm 36.0 (SD 9.81); waiting list arm 36.3 (SD 9.66); Severity (Y-BOCS): AS arm 21.96 (SD 8.17); waiting list arm 22.83 (SD 6.66)	AS n=43; duration 4 weeks; exposure: no; cognitive modification: yes (association splitting); device: treatment manual through email; interaction: no; gamification: no; Waiting list n=43; duration 4 weeks	Y-BOCS, OCI-R, BDI
Moritz et al 2015 [37]; English-speaking self-help groups and institutions devoted to research and treatment of OCD; RCT	Dx: externally verified diagnosis of OCD; Sex: IBT[w] arm: 64% women; waiting list arm: 60%; Meds: yes	Age: IBT arm 36.88 (SD 13.14); waiting list arm 34.32 (SD 10.79); Severity (Y-BOCS): IBT arm 22.64 (SD 7.56); waiting list arm 21.48 (SD 7.38)	IBT n=25; duration 4 weeks; exposure: no; cognitive modification: yes (association splitting); device: treatment manual through email; interaction: no; gamification: no; Waiting list n=25; duration 4 weeks	Y-BOCS, OCI-R, ICQ[x], WHOQOL-BREF[y]
Moritz et al 2018 [36]; online forum on OCD, Facebook OCD group, Yahoo newsgroups devoted to OCD; RCT	Dx: diagnosis by a mental health specialist; Sex: myMCT arm: 71.4% women; waiting list arm: 82.9%; Meds: yes	Age: myMCT arm 38.17 (SD 11.96); waiting list arm 39.34 (SD 14.52); Severity (Y-BOCS): myMCT arm 23.09 (SD 5.93); waiting list arm 21.74 (SD 6.23)	myMCT n=36; duration 6 weeks; exposure: yes (ERP); cognitive modification: yes; other: mindfulness; device: treatment manual through email; interaction: no; gamification: no; Waiting list n=36; duration 6 weeks	Y-BOCS, OCI-R, PHQ-9[z], Maladaptive and Adaptive Coping Scale, PSQ*[aa]
Salemink et al 2015 [27]; Europe; RCT	Dx: DSM-IV-TR [bb]; Sex: CBM-I[cc] arm: 55.6% women; psychological placebo arm: 71.4%; Meds: yes	Age: CBM-I arm 15.6 (SD 2.4); psychological placebo arm 9 (SD 15.1); Severity (Children's Y-BOCS): CBM-I arm 23.9 (SD 7.6); psychological placebo arm, 20.4 (SD 4.3)	CBM-I n=12; duration 1.6 weeks; exposure: yes (interoceptive exposure); cognitive modification: yes (cognitive bias modification training); device: computer; interaction: yes; gamification: no; Waiting list n=9; duration 1.6 weeks	Y-BOCS, OBQ-CV[dd], RCADS[ee], CDI[ff]
Schneider et al 2015 [32]; recruited from self-help forum through internet; RCT	Dx: diagnosis by a health care professional; Sex: COMET[gg] arm: 55.9% women; waiting list arm: 61.3%; Meds: yes	Age: COMET arm 37.47 (SD 10); psychological placebo arm 37.06 (SD 10.3); Severity (Y-BOCS): COMET arm 18.5 (SD 5.95); waiting list arm 19.84 (SD 5.99)	COMET n=34; duration 4 weeks; exposure: yes (interoceptive exposure); cognitive modification: yes (competitive memory training); device: pdf manual through email; interaction: no; gamification: no; Waiting list n=34; duration 4 weeks	Y-BOCS, OCI-R, BDI-SF, RSES[hh]

Author; year; citation; country; study design	Participants		Interventions	Outcomes
	Diagnosis; sex; medications	Age (years), mean (SD); baseline severity, mean (SD)		
Schröder et al 2020 [29]; Europe; RCT	Dx: Y-BOCS >7; Sex: iCBT[ii] arm: 75% women; CAU[jj] arm: 78.13%; Meds: yes	Age: iCBT arm 41.45 (SD 12.15); CAU arm 38.98 (SD 11.55); Severity (Y-BOCS): iCBT arm 20.2 (SD 6.29); CAU arm 20.17 (SD 5.73)	iCBT n=64; duration 8 weeks; exposure: yes (ERP); cognitive modification: yes (metacognitive training); other, mindfulness; device: computer (text, video, audio, photo, illustration); interaction: yes; gamification: no; CAU n=64; duration 8 weeks	Y-BOCS, OCI-R, OBQ-44[kk], WHO-QOL-BREF
Wootton et al 2019 [38]; cross-continental; RCT	Dx: Y-BOCS ≥14; Sex: ICBT[kk] arm: 81.5% women; waiting list arm: 81.3%; Meds: yes	Age: ICBT arm 34.03 (SD 10.8); waiting list arm 33.39 (SD 10.25); Severity (Y-BOCS): ICBT arm 22.52 (SD 4.91); waiting list arm 22.44 (SD 5.55)	ICBT n=90; duration 8 weeks; exposure: yes (ERP); cognitive modification: no; device: text-based online slides; interaction: no; gamification: no; Waiting list n=100; duration 8 weeks	Y-BOCS, DOCS[ll], PHQ-9

[a]RCT: randomized controlled trial.
[b]OCI-R: Obsessive-Compulsive Inventory–Revised.
[c]CCT: computerized cognitive control.
[d]PVT: peripheral vision training.
[e]meds: adjunctive medications.
[f]BDI-II: Beck Depression Inventory–Second Edition.
[g]PANAS: Positive and Negative Affectivity Scale.
[h]PSWQ: Penn State Worry Questionnaire.
[i]DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition.
[j]Y-BOCS: Yale-Brown Obsessive-Compulsive Scale.
[k]BT STEPS: Behavior Therapy Self-Help System.
[l]ERP: exposure and response prevention.
[m]PGI-I: Patient Global Impression of Improvement.
[n]CGI scale: Clinical Global Impression scale.
[o]WSAS: Work and Social Adjustment Scale.
[p]HAM-D: Hamilton Depression Rating Scale.
[q]SRI: serotonin reuptake inhibitor.
[r]BDI: Beck Depression Inventory.
[s]OBQ: Obsessive Belief Questionnaire.
[t]OCD: obsessive-compulsive disorder.
[u]BDI-SF: Beck Depression Inventory–Short Form.
[v]AS: association splitting.
[w]IBT: inference-based therapy.
[x]ICQ: Inferential Confusion Questionnaire.
[y]WHOQOL-BREF: Brief Quality of Life Questionnaire of the World Health Organization.
[z]PHQ-9: Patient Health Questionnaire.
[aa]PSQ: Patient Satisfaction Questionnaire.
DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision
[cc]CBM-I: Cognitive Bias Modification of Interpretation training.
[dd]OBQ-CV: Obsessive Belief Questionnaire–Child Version.
[ee]RCADS: Revised Child Anxiety and Depression Scale.
[ff]CDI: Children's Depression Inventory.
[gg]COMET: Competitive Memory Training.
[hh]RSES: Rosenberg Self-Esteem Scale.
[ii]iCBT: internet-based cognitive-behavioral therapy.
[jj]CAU: care-as-usual.
[kk]OBQ-44: Obsessive Belief Questionnaire–44 item.
[ll]DOCS: Dimensional Obsessive-Compulsive Scale.

Bias Arising From Randomization Process

Most of the studies did not provide information on allocation sequence concealment. Of those who provided details, Schneider et al [32] used an online randomization and allocation system, Salemink et al [27] suspected baseline imbalance, and Wootton et al [38] did not blind allocation to the clinician assessing participants.

Bias Due to Deviations From Intended Interventions

Six studies used waiting lists [25-27,29-31], and the other 5 used psychological placebo or treatment as control arms [19-23]. The percentage of dropouts was unbalanced between the arms and probably affected the results except for 2 studies, where the authors conducted analyses to confirm the deviations did not affect the outcome [33,38].

Bias Due to Missing Outcome Data

Most of the studies were missing more than 5% of the data, were unbalanced, and neglected to provide reasons for dropouts [19-22,25-27,29,30]. One study, however, had no missing data [30], and another conducted an analysis to prove that missingness did not affect the true value [38].

Bias in Measurement of Outcome

Primary efficacy outcome was measured by the self-rated Y-BOCS. It was unclear if knowledge of the intervention influenced the results.

Bias in Selection of Reported Results

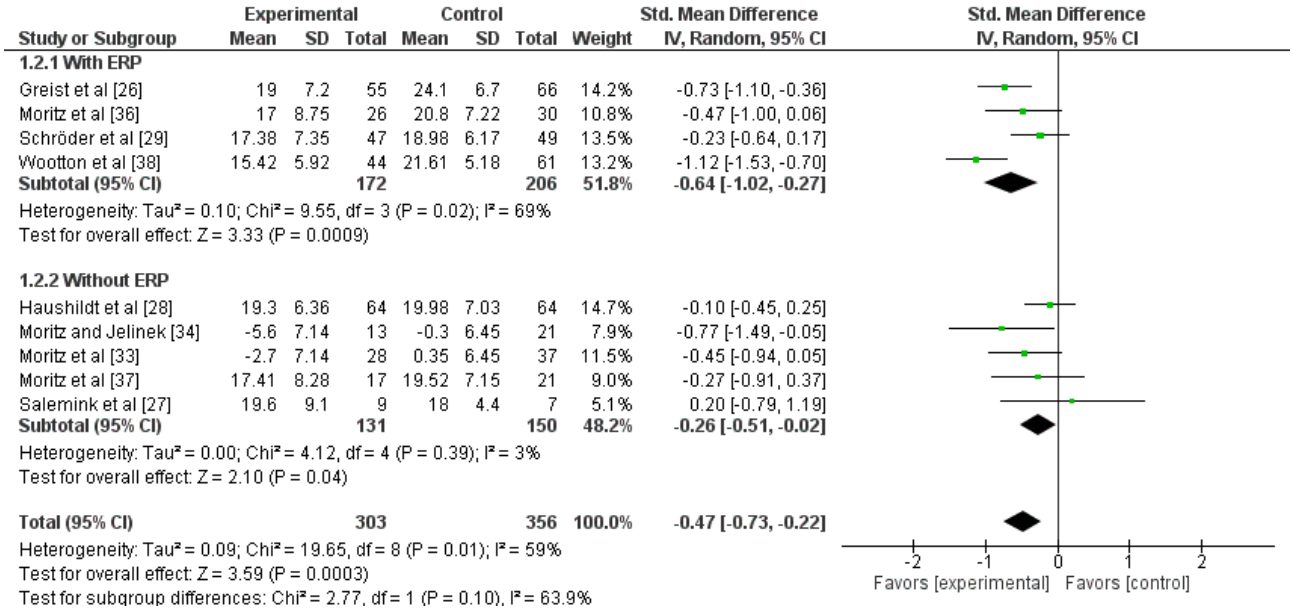
We found the protocol for the RCT by Schröder et al [29] but no others; therefore, selection of the reported results was unclear.

Primary Outcomes

Short-term Subjective Improvement of OCD Symptoms

Unguided computer-assisted self-help therapy was more effective than the waiting list and psychological placebo in terms of short-term subjective improvement of OCD symptoms (SMD −0.47, 95% CI −0.73 to −0.22; 9 studies; 659 participants). There was moderate heterogeneity (I^2=59%; Tau^2=0.09; Figure 3). The quality of evidence was very low due to the risk of bias of the included studies and inconsistency of the results.

Figure 3. Forest plot of short-term improvement of obsessive-compulsive disorder symptoms. ERP: exposure and response prevention.



Heterogeneity decreased to 24% after we excluded a study that reported a large number of dropouts before the intervention began [38]. This may have caused participants with high motivation to start the intervention, exaggerating the therapeutic effect. The improvement in OCD symptoms in the intervention arms remained larger than that in the control arms after exclusion of the study (SMD −0.38, 95% CI −0.58 to −0.18; 8 studies; 554 participants).

We imputed standard deviations in 2 studies [33,34]. The exclusion of these studies did not substantially change the result (SMD −0.44, 95% CI −0.76 to −0.12; 7 studies; 560 participants). The sensitivity analysis without studies targeted to child and adolescent [27] did not substantially change the result (SMD –0.51; 95% CI, –0.56 to –0.04; 8 studies; 643 participants; I^2=61%).

The subgroup analysis limited to those with ERP tended to strengthen the efficacy of unguided computer-assisted self-help therapy (SMD −0.64, 95% CI −1.02 to −0.27; 4 studies; 378 participants; I^2=69%) [26,29,36,38,40] compared with those without ERP (SMD −0.26, 95% CI −0.51 to −0.02; 5 studies; 281 participants; I^2=3%) [27,28,33,34,37]. The test for subgroup difference suggested a subgroup difference (P=.10).

The majority of the studies only sent the treatment manual via email. The subgroup analysis limited to those studies with treatment manual showed that the intervention was more effective than the control arm (SMD −0.44, 95% CI −0.68 to −0.20; 6 studies; 442 participants; I^2=66%) [26,28,33,34,36,37]. In comparison, the analysis limited to those using computer display showed no significant difference between the intervention and control groups (SMD −0.46, 95% CI −1.21 to

0.29; 3 studies; 217 participants; I^2=83%; Multimedia Appendix 3) [27,29,38]. The test for subgroup difference showed no significant subgroup difference (*P*=.96). This tendency became more evident after excluding a study where dropouts were relatively large before the interventions started (SMD −0.17, 95% CI −0.54 to 0.20; 2 studies; 112 participants; I^2=0%) [27,29]. However, the test for subgroup difference still showed no significant subgroup difference (*P*=.24).

The subgroup analysis limited to studies with some kind of interaction with systems showed no significant difference between the intervention and control groups (SMD −0.38, 95% CI –0.84 to 0.09; 3 studies; 233 participants; I^2=60%) [26,27,29], whereas an analysis limited to studies without interaction showed the intervention was more effective than the control arm treatment (SMD −0.52, 95% CI −0.87 to −0.17; 6 studies; 426 participants; I^2=55%). The test for subgroup difference showed no significant subgroup difference (*P*=.63; Multimedia Appendix 4).

In terms of duration of the intervention, studies with 4 weeks or less of intervention showed no significant difference between the intervention and control groups (SMD −0.20, 95% CI −0.45 to 0.06; 4 studies; 247 participants; I^2=0%) [27,28,33,37], whereas those with a duration of over 4 weeks showed that interventions were more effective than control (SMD −0.64, 95% CI −1.02 to −0.27; 4 studies; 378 participants; I^2=69%; Multimedia Appendix 5) [26,29,36,38]. The test for subgroup difference suggested a subgroup difference (*P*=.05). The number of sessions conducted was unclear, as they were self-help interventions.

The subgroup analysis limited to studies with waiting list as control arm showed that the intervention was significantly more effective than the waiting list (SMD −0.56, 95% CI −0.91 to −0.22; 6 studies; 314 participants; I^2=51%) [27,33,34,36,38], whereas studies with psychological placebo as control arm showed no significant difference between the intervention and control groups (SMD −0.35, 95% CI −0.74 to 0.03; 3 studies; 345 participants; I^2=69%). The test for subgroup difference showed no significant subgroup difference (*P*=.43; Multimedia Appendix 6).

Dropout for Any Reason at Posttreatment

Unguided computer-assisted self-help therapy had more dropouts for any reason than waiting list or psychological placebo (risk ratio [RR] 1.98, 95% CI 1.21 to 3.23; 11 studies, 983 participants; Figure 4). The visual inspection of the funnel plot suggested publication bias (Figure 5). In fact, there was considerable heterogeneity (I^2=79%, Tau^2=0.41). The quality of evidence was very low due to the risk of bias, inconsistency of results, and suspected publication bias.

Heterogeneity decreased to 29% after excluding a study that had also been excluded from the sensitivity analysis of the short-term improvement of OCD symptoms [38]. The dropouts for any reason in the intervention arm were still larger than those in the control arms after exclusion of the study (RR 2.19, 95% CI 1.56 to 3.07; 11 studies; 793 participants).

The sensitivity analysis without studies targeted to child and adolescent [27] did not substantially change the result (RR 2.06, 95% CI 1.23 to 3.45; 10 studies; 962 participants; I^2=82%).

Figure 4. Forest plot of dropout for any reason at posttreatment.

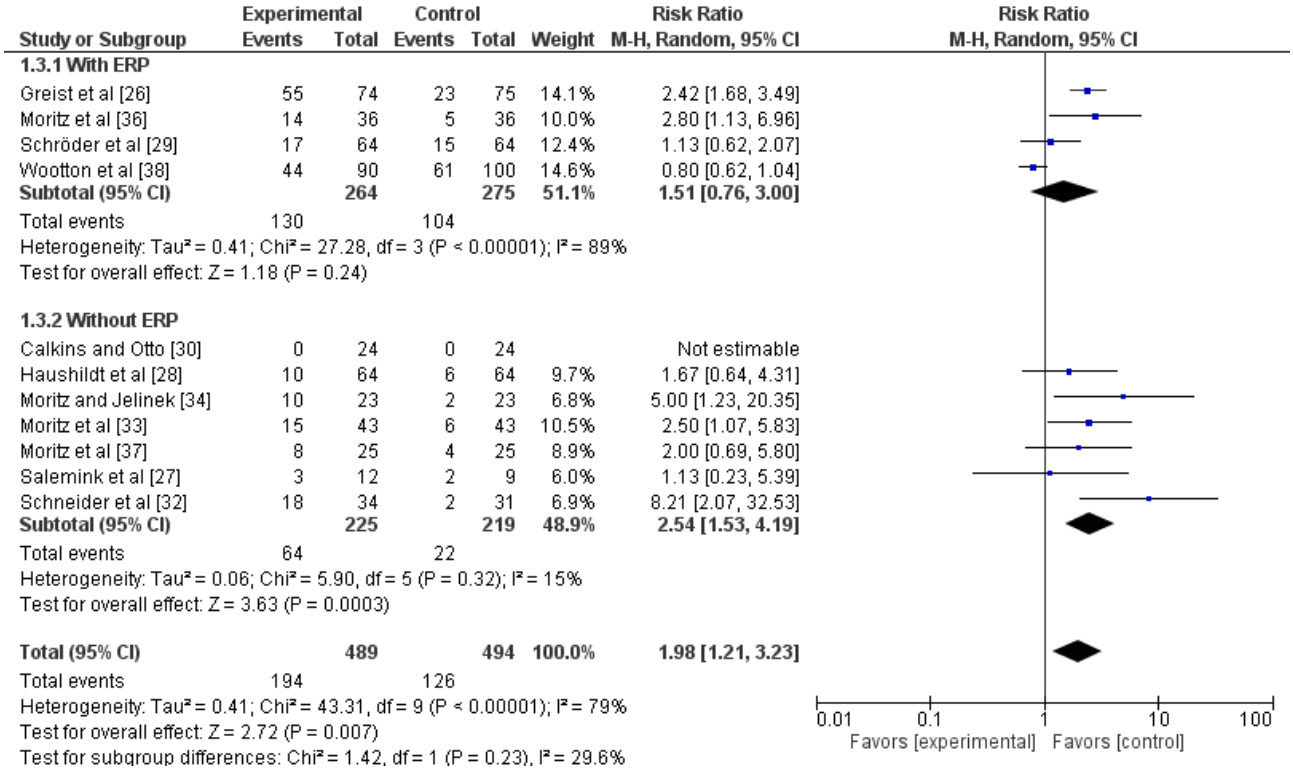
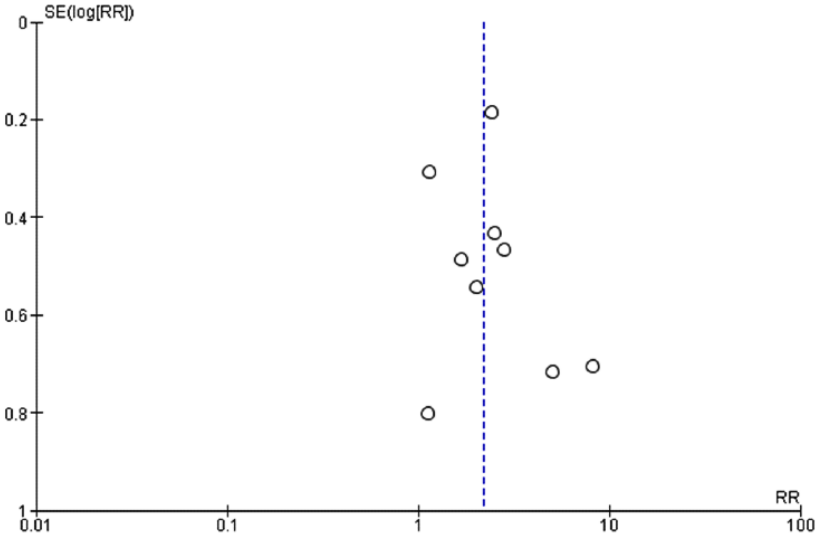


Figure 5. Funnel plot to assess for publication bias of dropout for any reason.



The subgroup analysis limited to those with ERP showed no significant difference between the intervention and control arms in dropout for any reason (RR 1.51, 95% CI 0.76 to 3.00; 4 studies; 539 participants; I^2=89%) [26,29,36,38]. Among these 4 studies, the 2 with manual-based treatment had significantly more dropouts in the intervention arm than in the control arm (RR 2.47, 95% CI 1.76 to 3.47; 221 participants; I^2=0%) [26,36], and the 2 with online slide or video, audio, photo, and illustration were not significantly different from the control arm in dropout for any reason (RR 0.86, 95% CI 0.65 to 1.13; 4 studies; 318 participants; I^2=10%) [29,38]. The studies with manual-based treatment [26,36] had intervention durations of 10 weeks and 6 weeks, respectively; the online interventions [29,38] were both 8 weeks in duration. The analysis limited to those without ERP showed that the intervention arm had significantly more dropout for any reason than the control arm (RR 2.54, 95% CI 1.53 to 4.19; 7 studies; 444 participants; I^2=15%) [27,28,30,32-34,37].

The subgroup analysis limited to those using computer display showed no significant difference between the intervention and control group with respect to dropout for any reason (RR 0.85, 95% CI 0.67 to 1.08; 3 studies; 339 participants; I^2=0%) [27,29,38]. The analysis of studies using a treatment manual showed that the intervention had significantly more dropout for any reason than the control arm (RR 2.55, 95% CI 1.93 to 3.36; 8 studies; 644 participants; I^2=0%; Multimedia Appendix 7) [26,28,30,32-34,36,37]. The test for subgroup difference suggested a subgroup difference (P=.05).

The subgroup analysis limited to studies with some kind of interaction with systems showed that the intervention arm had a significantly higher dropout for any reason than the control (RR 1.65, 95% CI 0.89 to 3.06; 3 studies; 233 participants; I^2=60%) [26,27,29]. This trend was the same as the analysis of those without interaction (RR 2.33, 95% CI 1.11 to 4.88; 8 studies; 685 participants; I^2=82%; Multimedia Appendix 8). The test for subgroup difference showed no significant subgroup difference (P=.48). There was substantial heterogeneity in both analyses.

In terms of duration of the intervention, studies with interventions of 4 weeks or less had significantly more dropouts than the control arm (RR 2.54, 95% CI 1.53 to 4.19; 7 studies; 444 participants; I^2=15%), whereas those with more than 4 weeks of intervention showed no significant difference in dropout for any reason between the intervention and control arms (RR 1.51, 95% CI 0.76 to 3.00; 4 studies; 539 participants; I^2=89%; Multimedia Appendix 9). The test for subgroup difference showed no significant subgroup difference (P=.23), The latter analysis included the same studies as the analysis of studies with ERP.

The subgroup analysis by control arm showed that the intervention group had significantly more dropouts for any reason than control groups (waiting list control RR 1.79, 95% CI 1.24 to 2.58; 7 studies; 530 participants; I^2=78%; others RR 2.76, 95% CI 1.73 to 4.38; 4 studies; 453 participants; I^2=81%; Multimedia Appendix 10).

Secondary Outcomes

Short-term Response Rate

The unguided computer-assisted self-help therapy had a more short-term response than the waiting list/psychological placebo (RR 1.93, 95% CI 1.16 to 3.21; 2 studies; 249 participants). Heterogeneity was negligible (I^2=18%, Tau^2=0.02).

Short-term Improvement of Anxiety

One study evaluated short-term improvements in anxiety [27]. There was no significant difference between the unguided computer-assisted self-help therapy and waiting list/psychological placebo in the improvement of anxiety (mean difference [MD] −6.20, 95% CI −20.38 to 7.98; 1 study; 16 participants).

Short-term Improvement of Depression

The improvement in depression was significantly greater in unguided computer-assisted self-help therapy than in the waiting list/psychological placebo (SMD −0.19, 95% CI −0.35 to −0.02; 7 studies; 560 participants). Heterogeneity was negligible (I^2=0%, Tau^2=0). The sensitivity analysis without studies

targeted to child and adolescent [27] did not substantially change the result (SMD –0.18, 95% CI –0.35 to –0.01; 6 studies; 544 participants; I^2=0%).

Quality of Life

There was no significant difference in short-term improvement of QOL between the unguided computer-assisted self-help therapy and waiting list/psychological placebo (MD 0.48, 95% CI −4.06 to 5.03; 2 studies; 134 participants). Heterogeneity was negligible (I^2=0%, Tau^2=0).

Other Outcomes

No study has evaluated outcomes longer than 6 months. One study evaluated the Y-BOCS and Beck Depression Inventory–Second Edition (BDI-II) at 6 months [28]. There was no difference in the improvement of Y-BOCS (MD 0.46, 95% CI −2.02 to 2.94; 128 participants) and BDI-II (MD 0.47, 95% CI −2.65 to 3.59; 128 participants) at 6 months between unguided computer-assisted self-help therapy and waiting list/psychological placebo.

Discussion

Summary of Main Outcomes

We included 11 studies with a total of 983 participants. The results indicated that unguided computer-assisted self-help therapy was moderately more effective than waiting lists or a attention placebo, which was confirmed by sensitivity analyses. In addition, there were no significant differences in acceptability as measured by dropout for any reason between the 2 arms.

Subgroup analysis limited to studies with ERP or interventions of 4 weeks or less tended to strengthen the efficacy of unguided computer-assisted self-help therapy, although the number of included studies in these analyses was small. Moreover, there was no significant difference in efficacy between the 2 groups when the analysis was limited to studies using computer display or studies with the interaction between participants and systems.

For the acceptability measured by dropout for any reason, subgroup analysis limited to studies with ERP did not change the result, but the intervention arm had more dropouts when the analysis was limited to studies using treatment manual via email. In terms of intervention duration, analysis limited to studies of 4 weeks or less showed that the intervention arm had a greater number of dropouts than the control arm.

Short-term responses for secondary outcomes supported the efficacy of unguided computer-assisted self-help therapy; however, only 2 studies were included in the analysis. The short-term improvement of depression was greater with unguided computer-assisted therapy, but 2 studies reported no significant difference in the improvement of QOL and 1 study reported no difference in level of anxiety. There are no studies with long-term outcomes.

Comparison With Other Systematic Reviews and Strengths of This Review

There were 3 systematic reviews and meta-analyses related with this study. All results favored the interventions. Firth et al [12] indicated a small-to-moderate effect (Hedges g=0.325) of a smartphone intervention on the total symptoms of anxiety in comparison with control conditions, which did not exclude face-to-face support. Tumur et al [10] showed that the effect size of Y-BOCS in a computer-assisted CBT intervention named BT Step was 0.84, which was the only intervention included in the analysis. The study conducted by Peacy et al [11], which was most similar to this study, showed that the effect size of self-administered self-help intervention was small (Hedges g=0.33).

In accordance with previous reviews, our review favored unguided computer-assisted self-help therapy against control arms, and the effect size was moderate (SMD −0.47). Although Pearcy et al [11] showed a small effect size of the intervention, they included quasi-experimental studies, and the RCT conducted by Greist et al [26] was misclassified to predominantly self-help; the study author confirmed was a self-administered therapy upon our inquiry.

This review reveals the acceptability of self-guided computer-assisted therapy for OCD measured by dropout for any reason. Future systematic reviews on self-guided OCD therapy should include the analysis of acceptability as one of the problems of self-guided therapy [41,42].

Importance of ERP and Comparison to Intervention With Human Contact

This study reconfirmed the importance of ERP in the treatment of OCD. The results of the meta-analysis showed that interventions with ERP were significantly more effective than those without ERP. However, human contact may strengthen the effect of ERP. The past meta-analyses on intervention with ERP compared with control condition showed that the SMDs of obsessive-compulsive symptoms were 1.16 and 0.74, respectively [14,15]. The former did not include computer-assisted interventions and the latter did. Our results showed that the effect of unguided computer-assisted self-help interventions without human contact expressed as SMD was 0.64. These facts suggest the importance of human contact in ERP. In fact, one of the meta-analyses listed above showed that the SMD of therapist-controlled ERP (SMD 1.58) was greater than that of self-controlled ERP (SMD 0.81) [15]. Unguided computer-assisted self-help interventions without human contact should include ERP, and future studies should examine what factors of human contact strengthen the effect of ERP.

Duration of Intervention and Its Influence on Effect and Dropouts

Our results showed that interventions with a duration over 4 weeks were more effective and tended to have fewer dropouts than interventions of 4 weeks or less. Avoiding interventions shorter than 4 weeks is recommended, considering the negligible heterogeneity of the results. However, it is unclear how long the intervention should be.

Several studies indicated that increment of treatment effect would decrease as the number of sessions increases [16,43], and a study suggested that patients tend to end therapy when they are satisfied with their improvement [17]. An intervention

with a flexible number of sessions may be one option to determine the optimal number of sessions.

Comparison With Other Apps

One systematic review showed that highly rated anxiety apps contain gamification (32%) and social elements including chat and communication with others (46%) [44]. The studies included in our systematic review did not use gamification or mobile devices, and only 2 studies used interaction. Future studies of self-guided computer-assisted therapy for OCD should include these elements to increase efficacy and acceptability.

Limitations

This study has several limitations. First, this study did not include active interventions as a comparison. While this would increase the number of included studies and precision, such an analysis may lead to an underestimation of the target intervention's efficacy. Second, the overall risk of bias of the included studies was high, which led to downgrading the quality of evidence. However, this was unavoidable since a waiting list was the comparison arm, and the primary efficacy outcome was measured using a self-administered questionnaire. Future studies should use a psychological placebo to keep participants blinded to the intervention and the objective outcomes. Third, we did not consider sponsorship bias, which may favor the results of the intervention. However, as it seems that all authors developed the intervention, the results of this review may have overestimated the effect. The test of sponsorship bias should be initiated at the study design level. Fourth, the heterogeneity of results was moderate to considerable. This suggests that various factors are related to the effect of the computer-assisted self-help interventions in patients with OCD, such as the module, duration, modality of presenting intervention, gamification, and intervention. The number of studies on the computer-assisted self-help interventions in patients with OCD is still small, as shown in this study. More studies to explore and optimize the effect of the intervention should be conducted.

Conclusions

Our study suggests that unguided computer-assisted self-help therapy for OCD is effective compared to waiting lists or psychological placebo. An ERP component and intervention duration of more than 4 weeks may strengthen the efficacy without worsening the acceptability of the therapy.

However, the included studies did not effectively use the merits of computerization. Few studies have examined the interaction between participants and systems, and none of the studies used gamification. Furthermore, most studies only used text-based interventions. No study used a mobile device. Portability seems to be useful for intervention components, such as self-monitoring and in vivo exposure; therefore, future studies should examine these factors.

Acknowledgments

We would like to thank Dr Ethan Sahker and Editage for English language editing. This work was supported by grant 21K157294 from the Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research to HI. The funder had no role in study design, data collection, and analysis; decision to publish; or preparation of the manuscript.

Authors' Contributions

HI, AT, and TAF were involved in study design and data interpretation. HI, AT, HN, NY, KK, HN, NT, YH, and YO were involved in data selection and extraction. All authors critically revised the report, commented on drafts of the manuscript, and approved the final report.

Conflicts of Interest

HI received consulting fees from Mitsubishi-Tanabe Pharma and honoraria for lectures from Mochida Pharmaceutical, Otsuka Pharmaceutical, and Kyowa Pharmaceutical. AT received lecture fees from Dainippon-Sumitomo, Janssen, Meiji-seika, Mitsubishi-Tanabe, and Otsuka. NY received a book royalty from Medical Friend; honoraria for lectures from Gakken Medical Support, Eisai, Meiji Seika Pharma, Mitsubishi-Tanabe Pharma, and Mochida Pharmaceutical; and honoraria for writings from Igaku-Shoin, Nikkei Business Publications, Maruzen Publishing, and Elsevier. He is a diplomate of the Academy of Cognitive and Behavioral Therapies, Secretary Board Member of the Japanese Association for Cognitive Therapy, and a member of the Japan Clinical Guideline Development Group for Anxiety Disorders and Obsessive-Compulsive Disorder. TAF reports grants and personal fees from Mitsubishi-Tanabe, personal fees from SONY, and grants and personal fees from Shionogi outside the submitted work. In addition, TAF has a patent 2020-548587 concerning smartphone CBT apps pending and intellectual properties for Kokoro-app licensed to Mitsubishi-Tanabe. All other authors declare no conflicts of interest.

Multimedia Appendix 1
PRISMA checklist.
[DOCX File , 32 KB - jmir_v24i4e35940_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOCX File , 15 KB - jmir_v24i4e35940_app2.docx]

Multimedia Appendix 3
Forest plot of short-term improvement of obsessive-compulsive disorder symptoms: subgroup analysis by treatment type.
[PNG File , 20 KB - jmir_v24i4e35940_app3.png]

Multimedia Appendix 4
Forest plot of short-term improvement of obsessive-compulsive disorder symptoms: subgroup analysis with and without interaction.
[PNG File , 20 KB - jmir_v24i4e35940_app4.png]

Multimedia Appendix 5
Forest plot of short-term improvement of obsessive-compulsive disorder symptoms: subgroup analysis by treatment duration.
[PNG File , 20 KB - jmir_v24i4e35940_app5.png]

Multimedia Appendix 6
Forest plot of short-term improvement of obsessive-compulsive disorder symptoms: subgroup analysis by control condition.
[PNG File , 20 KB - jmir_v24i4e35940_app6.png]

Multimedia Appendix 7
Forest plot of dropout for any reason at posttreatment: subgroup analysis by treatment type.
[PNG File , 19 KB - jmir_v24i4e35940_app7.png]

Multimedia Appendix 8
Forest plot of dropout for any reason at posttreatment: subgroup analysis with or without interaction.
[PNG File , 20 KB - jmir_v24i4e35940_app8.png]

Multimedia Appendix 9
Forest plot of dropout for any reason at posttreatment: subgroup analysis by treatment duration.
[PNG File , 20 KB - jmir_v24i4e35940_app9.png]

Multimedia Appendix 10
Forest plot of dropout for any reason at posttreatment: subgroup analysis by control condition.
[PNG File , 20 KB - jmir_v24i4e35940_app10.png]

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th Edition. Washington: American Psychiatric Publishing; 2013.
2. Eaton WW, Martins SS, Nestadt G, Bienvenu OJ, Clarke D, Alexandre P. The burden of mental disorders. Epidemiol Rev 2008;30:1-14 [FREE Full text] [doi: 10.1093/epirev/mxn011] [Medline: 18806255]
3. Skapinakis P, Caldwell DM, Hollingworth W, Bryden P, Fineberg NA, Salkovskis P, et al. Pharmacological and psychotherapeutic interventions for management of obsessive-compulsive disorder in adults: a systematic review and network meta-analysis. Lancet Psychiatry 2016 Aug;3(8):730-739 [FREE Full text] [doi: 10.1016/S2215-0366(16)30069-4] [Medline: 27318812]
4. McHugh RK, Whitton SW, Peckham AD, Welge JA, Otto MW. Patient preference for psychological vs pharmacologic treatment of psychiatric disorders: a meta-analytic review. J Clin Psychiatry 2013 Jun;74(6):595-602 [FREE Full text] [doi: 10.4088/JCP.12r07757] [Medline: 23842011]
5. Obsessive-compulsive disorder and body dysmorphic disorder: treatment. National Institute for Health and Care Excellence. 2005. URL: https://www.nice.org.uk/guidance/CG31/chapter/1-Guidance#steps-35-treatment-options-for-people-with-ocd-or-bdd [accessed 2022-04-07]
6. Goodwin R, Koenen KC, Hellman F, Guardino M, Struening E. Helpseeking and access to mental health treatment for obsessive-compulsive disorder. Acta Psychiatr Scand 2002 Aug;106(2):143-149. [doi: 10.1034/j.1600-0447.2002.01221.x] [Medline: 12121213]
7. García-Soriano G, Rufer M, Delsignore A, Weidt S. Factors associated with non-treatment or delayed treatment seeking in OCD sufferers: a review of the literature. Psychiatry Res 2014 Dec 15;220(1-2):1-10. [doi: 10.1016/j.psychres.2014.07.009] [Medline: 25108591]
8. Lind C, Boschen MJ, Morrissey S. Technological advances in psychotherapy: implications for the assessment and treatment of obsessive compulsive disorder. J Anxiety Disord 2013 Jan;27(1):47-55. [doi: 10.1016/j.janxdis.2012.09.004] [Medline: 23247201]

9. Rogers MA, Lemmen K, Kramer R, Mann J, Chopra V. Internet-delivered health interventions that work: systematic review of meta-analyses and evaluation of website availability. J Med Internet Res 2017 Mar 24;19(3):e90 [FREE Full text] [doi: 10.2196/jmir.7111] [Medline: 28341617]
10. Tumur I, Kaltenthaler E, Ferriter M, Beverley C, Parry G. Computerised cognitive behaviour therapy for obsessive-compulsive disorder: a systematic review. Psychother Psychosom 2007;76(4):196-202. [doi: 10.1159/000101497] [Medline: 17570957]
11. Pearcy CP, Anderson RA, Egan SJ, Rees CS. A systematic review and meta-analysis of self-help therapeutic interventions for obsessive-compulsive disorder: is therapeutic contact key to overall improvement? J Behav Ther Exp Psychiatry 2016 Jun;51:74-83. [doi: 10.1016/j.jbtep.2015.12.007] [Medline: 26794856]
12. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
13. 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]
14. Reid JE, Laws KR, Drummond L, Vismara M, Grancini B, Mpavaenda D, et al. Cognitive behavioural therapy with exposure and response prevention in the treatment of obsessive-compulsive disorder: a systematic review and meta-analysis of randomised controlled trials. Compr Psychiatry 2021 Apr;106:152223 [FREE Full text] [doi: 10.1016/j.comppsych.2021.152223] [Medline: 33618297]
15. Abramowitz JS. Variants of exposure and response prevention in the treatment of obsessive-compulsive disorder: a meta-analysis. Behavior Therapy 1996;27(4):583-600. [doi: 10.1016/s0005-7894(96)80045-1]
16. Kopta SM. The dose-effect relationship in psychotherapy: a defining achievement for Dr. Kenneth Howard. J Clin Psychol 2003 Jul;59(7):727-733. [doi: 10.1002/jclp.10167] [Medline: 12808579]
17. Barkham M, Connell J, Stiles WB, Miles JNV, Margison F, Evans C, et al. Dose-effect relations and responsive regulation of treatment duration: the good enough level. J Consult Clin Psychol 2006 Feb;74(1):160-167. [doi: 10.1037/0022-006X.74.1.160] [Medline: 16551153]
18. Cuijpers P, Marks IM, 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]
19. 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]
20. 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]
21. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
22. Furukawa TA, Barbui C, Cipriani A, Brambilla P, Watanabe N. Imputing missing standard deviations in meta-analyses can provide accurate results. J Clin Epidemiol 2006 Jan;59(1):7-10. [doi: 10.1016/j.jclinepi.2005.06.006] [Medline: 16360555]
23. Michopoulos I, Furukawa TA, Noma H, Kishimoto S, Onishi A, Ostinelli EG, et al. Different control conditions can produce different effect estimates in psychotherapy trials for depression. J Clin Epidemiol 2021 Apr;132:59-70. [doi: 10.1016/j.jclinepi.2020.12.012] [Medline: 33338564]
24. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. 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]
25. Nieman D, Domen A, Kumar R, Harrison J, De Haan L, Denys D. Cognitive remediation in psychiatric patients with an online cognitive game and assessment tool. Eur Neuropsychopharmacol 2015 Sep;25:S344-S345 [FREE Full text] [doi: 10.1016/s0924-977x(15)30433-8]
26. Greist JH, Marks IM, Baer L, Kobak KA, Wenzel KW, Hirsch MJ, et al. Behavior therapy for obsessive-compulsive disorder guided by a computer or by a clinician compared with relaxation as a control. J Clin Psychiatry 2002 Feb;63(2):138-145. [doi: 10.4088/jcp.v63n0209] [Medline: 11874215]
27. Salemink E, Wolters L, de Haan E. Augmentation of treatment as usual with online cognitive bias modification of interpretation training in adolescents with obsessive compulsive disorder: a pilot study. J Behav Ther Exp Psychiatry 2015 Dec;49(Pt A):112-119. [doi: 10.1016/j.jbtep.2015.02.003] [Medline: 25724385]
28. Hauschildt M, Schröder J, Moritz S. Randomized-controlled trial on a novel (meta-)cognitive self-help approach for obsessive-compulsive disorder ("myMCT"). J Obsessive Compuls Rel Dis 2016 Jul;10:26-34. [doi: 10.1016/j.jocrd.2016.04.010]

29. Schröder J, Werkle N, Cludius B, Jelinek L, Moritz S, Westermann S. Unguided Internet-based cognitive-behavioral therapy for obsessive-compulsive disorder: a randomized controlled trial. Depress Anxiety 2020 Dec 10;37(12):1208-1220. [doi: 10.1002/da.23105] [Medline: 33169490]
30. Calkins AW, Otto MW. Testing the boundaries of computerized cognitive control training on symptoms of obsessive compulsive disorder. Cogn Ther Res 2012 Nov 9;37(3):587-594. [doi: 10.1007/s10608-012-9496-x]
31. Wootton BM. Remote cognitive-behavior therapy for obsessive-compulsive symptoms: a meta-analysis. Clin Psychol Rev 2016 Feb;43:103-113. [doi: 10.1016/j.cpr.2015.10.001] [Medline: 26494179]
32. Schneider BC, Wittekind CE, Talhof A, Korrelboom K, Moritz S. Competitive Memory Training (COMET) for OCD: a self-treatment approach to obsessions. Cogn Behav Ther 2015;44(2):142-152. [doi: 10.1080/16506073.2014.981758] [Medline: 25421606]
33. Moritz S, Jelinek L, Hauschildt M, Naber D. How to treat the untreated: effectiveness of a self-help metacognitive training program (myMCT) for obsessive-compulsive disorder. Dialogues Clin Neurosci 2010;12(2):209-220 [FREE Full text] [Medline: 20623925]
34. Moritz S, Jelinek L. Further evidence for the efficacy of association splitting as a self-help technique for reducing obsessive thoughts. Depress Anxiety 2011 Jul;28(7):574-581. [doi: 10.1002/da.20843] [Medline: 21681869]
35. O'Connor KP, Aardema F, Bouthillier D, Fournier S, Guay S, Robillard S, et al. Evaluation of an inference-based approach to treating obsessive-compulsive disorder. Cogn Behav Ther 2005;34(3):148-163. [doi: 10.1080/16506070510041211] [Medline: 16195054]
36. Moritz S, Hauschildt M, Murray SC, Pedersen A, Krausz M, Jelinek L. New wine in an old bottle? Evaluation of myMCT as an integrative bibliotherapy for obsessive-compulsive disorder. J Obsessive Compulsive Rel Dis 2018 Jan;16:88-97. [doi: 10.1016/j.jocrd.2017.12.009]
37. Moritz S, Dietl C, Kersten JF, Aardema F, O'Connor K. Evaluation of inference-based therapy (doubt therapy) as a self-help tool for obsessive-compulsive disorder. J Cogn Psychother 2015;29(4):315-330. [doi: 10.1891/0889-8391.29.4.315] [Medline: 32755941]
38. Wootton BM, Karin E, Titov N, Dear BF. Self-guided internet-delivered cognitive behavior therapy (ICBT) for obsessive-compulsive symptoms: a randomized controlled trial. J Anxiety Disord 2019 Aug;66:102111. [doi: 10.1016/j.janxdis.2019.102111] [Medline: 31301476]
39. Moritz S, Russu R. Further evidence for the efficacy of association splitting in obsessive-compulsive disorder: an internet study in a Russian-speaking sample. J Obsessive Compulsive Rel Dis 2013 Apr;2(2):91-98. [doi: 10.1016/j.jocrd.2012.12.002]
40. Wolters L, Salemink E, De Beek V, De Haan E. Improving treatment: supplementing cognitive behavioral therapy with a cognitive bias modification training for children and adolescents with OCD. Eur Child Adolesc Psychiatry 2015;24 Suppl 1(S1):S120-S121.
41. Cuijpers P, Noma H, Karyotaki E, Cipriani A, Furukawa TA. Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression: a network meta-analysis. JAMA Psychiatry 2019 Apr 17;76(7):700-707. [doi: 10.1001/jamapsychiatry.2019.0268] [Medline: 30994877]
42. Yeung W, Chung K, Ho FY, Ho L. Predictors of dropout from internet-based self-help cognitive behavioral therapy for insomnia. Behav Res Ther 2015 Oct;73:19-24. [doi: 10.1016/j.brat.2015.07.008] [Medline: 26226091]
43. Howard KI, Kopta SM, Krause MS, Orlinsky DE. The dose-effect relationship in psychotherapy. Am Psychol 1986 Feb;41(2):159-164. [Medline: 3516036]
44. Drissi N, Ouhbi S, Janati Idrissi MA, Ghogho M. An analysis on self-management and treatment-related functionality and characteristics of highly rated anxiety apps. Int J Med Inform 2020 Jul 30;141:104243 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104243] [Medline: 32768994]

Abbreviations

BDI-II: Beck Depression Inventory–Second Edition
BT STEPS: Behavior Therapy Self-Help System
CBT: cognitive behavioral therapy
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision
ERP: exposure and response prevention
GRADE: Grading of Recommendations, Assessment, Development, and Evaluations
MD: mean difference
OCD: obsessive-compulsive disorder
OCI-R: Obsessive-Compulsive Inventory–Revised
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
QOL: quality of life
RCT: randomized controlled trial

RR: risk ratio
SMD: standardized mean difference
Y-BOCS: Yale-Brown Obsessive-Compulsive Scale

Edited by A Mavragani; submitted 22.12.21; peer-reviewed by Y Kanai, M Wu; comments to author 01.03.22; revised version received 07.03.22; accepted 11.03.22; published 21.04.22.

Please cite as:
Imai H, Tajika A, Narita H, Yoshinaga N, Kimura K, Nakamura H, Takeshima N, Hayasaka Y, Ogawa Y, Furukawa T
Unguided Computer-Assisted Self-Help Interventions Without Human Contact in Patients With Obsessive-Compulsive Disorder: Systematic Review and Meta-analysis
J Med Internet Res 2022;24(4):e35940
URL: https://www.jmir.org/2022/4/e35940
doi:10.2196/35940
PMID:35451993

Review

Effects of mHealth Interventions on Improving Antenatal Care Visits and Skilled Delivery Care in Low- and Middle-Income Countries: Systematic Review and Meta-analysis

Md Obaidur Rahman[1,2], MSc, PhD; Noyuri Yamaji[1], MNSc, PhD; Yasuko Nagamatsu[1], PhD; Erika Ota[1,3], PhD

[1]Department of Global Health Nursing, Graduate School of Nursing Science, St. Luke's International University, Tokyo, Japan

[2]Center for Surveillance, Immunization, and Epidemiologic Research, National Institute of Infectious Diseases, Tokyo, Japan

[3]Tokyo Foundation for Policy Research, Tokyo, Japan

Corresponding Author:
Md Obaidur Rahman, MSc, PhD
Department of Global Health Nursing
Graduate School of Nursing Science
St. Luke's International University
10-1 Akashi-cho, Chuo-ku
Tokyo, 104-0044
Japan
Phone: 81 3 3543 6391 ext 4990
Fax: 81 3 3543 1626
Email: obaidur006@gmail.com

Abstract

Background: The poor coverage of essential maternal services, such as antenatal care (ANC) and skilled delivery care utilization, accounts for higher maternal and infant mortality in low- and middle-income countries (LMICs). Although mobile health (mHealth) interventions could potentially improve the service utilization in resource-limited settings, their effectiveness remains unclear.

Objective: This review aimed to summarize the effect of mHealth interventions on improving the uptake of ANC visits, skilled birth attendance at the time of delivery, and facility delivery among pregnant women in LMICs.

Methods: We conducted a comprehensive search on 9 electronic databases and other resources from inception to October 2020. We included individual randomized controlled trials and cluster randomized controlled trials that assessed the effectiveness of mHealth interventions for improving perinatal health care utilization among healthy pregnant women in LMICs. We performed a random-effects meta-analysis and estimated the pooled effect size by using risk ratios (RRs) with 95% CIs. In addition, 2 reviewers independently assessed the risk of bias of the included studies by using the Cochrane risk of bias tool and the certainty of the evidence by using the Grading of Recommendation, Assessment, Development and Evaluation approach.

Results: A total of 9 studies (10 articles) that randomized 10,348 pregnant women (n=6254, 60.44% in the intervention group; n=4094, 39.56% in the control group) were included in this synthesis. The pooled estimates showed a positive effect of mHealth interventions on improving 4 or more ANC visit utilizations among pregnant women in LMICs, irrespective of the direction of interventions (1-way communications: RR 2.14, 95% CI 1.76-2.60, I^2=36%, 2 studies, moderate certainty; 2-way communications: RR 1.17, 95% CI 1.08-1.27, I^2=59%, 3 studies, low certainty). Only 2-way mHealth interventions were effective in improving the use of skilled birth attendance during delivery (RR 1.23, 95% CI 1.14-1.33, I^2=0%, 2 studies, moderate certainty), but the effects were unclear for 1-way mHealth interventions (RR 1.04, 95% CI 0.97-1.10, I^2=73%, 3 studies, very low certainty) when compared with standard care. For facility delivery, the interventions were effective in settings where fewer pregnant women used facility delivery (RR 1.68, 95% CI 1.30-2.19, I^2=36%, 2 studies, moderate certainty); however, the effects were unclear in settings where most pregnant women already used facility delivery (RR 1.01, 95% CI 0.97-1.04, I^2=0%, 1 study, low certainty).

Conclusions: mHealth interventions may contribute to improving ANC and skilled delivery care utilization among pregnant women in LMICs. However, more studies are required to improve their reproducibility and efficiency or strengthen the evidence of different forms of mHealth interventions because of the considerable heterogeneity observed in the meta-analyses.

Trial Registration: PROSPERO CRD42020210813; https://tinyurl.com/2n7ny9a7

(J Med Internet Res 2022;24(4):e34061) doi:10.2196/34061

KEYWORDS

mobile health; ANC; skilled delivery care; LMICs; systematic review and meta-analysis

Introduction

Background

Despite progress in improving global maternal mortality, it remains unacceptably high, particularly in low- and middle-income countries (LMICs) [1]. Reducing complications during and following pregnancy and childbirth, when most complications occur, can reduce or prevent maternal mortality. Skilled care during and following pregnancy and childbirth could reduce complications and may result in preventing maternal deaths. Studies conducted in Tanzania and Ethiopia have confirmed the ability of antenatal care (ANC) and skilled birth attendance (SBA) during labor and delivery to reduce maternal mortality [2-4].

Globally, most pregnant women have access to ANC with a skilled health professional (eg, physician, nurse, or midwife) at least once, but only 65% receive the World Health Organization–recommended number of at least four ANC visits, and 81% of births occur with the assistance of skilled health personnel [5]. Although there has been a significant improvement in the coverage for SBA and facility delivery in the last decade, millions of births still occur annually without any assistance from a skilled health professional [6]. Several factors prevent pregnant women from receiving the care provided by skilled health personnel during pregnancy and childbirth, such as lack of information, limited preventive health education, limited access to maternal health services owing to poverty or distance factors, poor administration, shortage of health care professionals, and inadequate or poor-quality services [7,8].

In LMICs, only approximately half of pregnant women receive 4 or more ANC visits, and the rate of skilled delivery care including SBA and facility delivery is relatively poor [5]. Moreover, the lowest levels of ANC and skilled delivery coverage are observed in regions where maternal mortality remains excessively high. For instance, the coverage for 4 or more ANC visits was 49% in South Asia and 52% in sub-Saharan Africa [5]. In terms of SBA at the time of delivery, the coverage was 60% in sub-Saharan Africa and 77% in Southern Asia, whereas other World Health Organization regions have achieved universal coverage [5]. The poor coverage of ANC and SBA accounted for the higher maternal mortality in these regions. Hence, a faster pace of progress is required to improve the coverage of ANC and skilled delivery in these high-burden regions.

In the last decade, mobile phone coverage has rapidly increased worldwide. The International Telecommunication Union reported that global mobile phone subscriptions crossed >7 billion in 2015, and mobile phone penetration reached over 90% in LMICs [9]. Thus, mobile phone penetration has the potential to strengthen existing health care service utilization, particularly ANC, SBA, and facility delivery services, in resource-limited settings in a cost-effective manner. Mobile health (mHealth) interventions are becoming more widespread in LMICs because the technology involved is more rapid and accessible than internet access. Labrique et al [10] reported on 12 mHealth applications that could respond to various health issues. Most mHealth interventions are designed to promote behavior change in patients or health personnel by providing health care reminders, health advice, health education, health information and facilitating referral, or access to health facilities or point-of-care remote consultation. Numerous models of mHealth interventions have been used to support pregnant women during and following pregnancy and childbirth in LMICs [11]. Previous studies have reported that mHealth interventions may be capable of and effective in improving essential maternal health care service utilization in LMICs [12-21]. Most of the systematic reviews narratively synthesized the available literature and reported a great potential for mHealth interventions to change maternal health care–seeking behaviors and showed a positive effect on improving ANC, SBA, postnatal care, or childhood immunization [12,14,15,18,22]. However, most evidence comes from observational studies and pilot or small-scale mHealth intervention studies, and researchers have expressed concerns about the study quality [12,14,15,18,22].

Although mHealth interventions have shown a great potential for behavior change more broadly, there are relatively few rigorous evaluations assessing the effectiveness of mHealth interventions on the uptake of ANC, SBA, and facility delivery utilization among pregnant women [19,21]. A recent systematic review reported a significantly higher mean of ANC attendance in mHealth interventions than standard care, but the result remains inconclusive because of the higher statistical heterogeneity among the studies; they also summarized randomized controlled trials (RCTs) and non-RCTs together into the meta-analysis [21]. Another systematic review considered RCTs to rigorously evaluate the effectiveness of mHealth interventions, and a meta-analysis based on a limited number of studies found a positive effect on improving 4 or more ANC visits and SBA outcomes [19]. Owing to higher statistical heterogeneity and the limited number of studies for each outcome, the evidence remains inconclusive [19]. To the best of our knowledge, no systematic review has reported on the effectiveness of mHealth interventions in improving facility delivery outcomes. Therefore, a rigorous evaluation of high-level studies is required to assess the effects of mHealth interventions on ANC and skilled delivery care utilization (SBA and facility delivery) in LMICs. This review further summarizes the findings of high-level studies (such as RCTs) and provides a clear direction to health practitioners, researchers, and policy makers.

Objectives

The objective of this review is to explore and synthesize the effects of mHealth interventions on improving the uptake of ANC visits, SBA at the time of delivery, and facility delivery among healthy pregnant women in resource-limited settings.

Methods

Protocol Registration and Review Guideline

The review protocol was registered in the PROSPERO database (CRD42020210813) [23]. The guideline of PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) was followed for reporting this systematic review and meta-analysis [24].

Search Methods for Study Identification

Using a highly sensitive search strategy, we conducted a comprehensive search from inception to October 2020 to identify RCTs, including cluster RCTs, in the following electronic databases: APA PsycINFO, British Nursing Index, CINAHL Plus, Embase, MEDLINE, POPLINE, PubMed, The Cochrane Library, and Web of Science. We used controlled vocabulary and text words for each database. The search terms were grouped into three major categories of interest: participants (pregnant women), interventions (mHealth interventions), and study designs (RCTs). We did not limit our search to language, date, or publication type to include all published studies. Moreover, we checked the reference lists of all the included studies and relevant systematic reviews to identify additional potential studies for inclusion. The details of the search strategies for each database are provided in Multimedia Appendix 1.

Study Eligibility Criteria

Overview

The study eligibility criteria were defined using the following PICOS (participants, interventions, comparisons, outcomes, and study designs or settings) framework. A study was included if it met all the following criteria.

Participants

We included the study if it was conducted on healthy pregnant women aged 15 to 49 years. If the study included a high-risk population, such as pregnant women with HIV/AIDS, cancer, preeclampsia, or other severe diseases at baseline, it was excluded because of higher medical adherence before intervention implementation among these groups. We considered low- or average-risk pregnant women, not high-risk pregnant women, as healthy pregnant women.

Interventions

We included all types of mHealth interventions that focused on improving perinatal health care utilization, including SBA and facility delivery.

Comparisons

We included studies that compared the effectiveness of any form of mHealth intervention (eg, voice calling, SMS text messaging, mobile apps, and videos) with standard care.

Outcomes

We included studies that reported ANC visits and skilled delivery care utilization, such as facility delivery and SBA during delivery.

Study Designs

We considered only RCTs and cluster RCTs in this review. Qualitative studies, case studies, cross-sectional studies, quasi-RCTs, quasi-experimental studies (controlled before and after studies), review studies, discussion papers, case reports, commentaries, editorials, expert opinions, and ongoing research with insufficient PICOS information were excluded.

Settings

We included studies conducted in LMICs based on the World Bank categories at the time of study implementation [25].

Study Selection Process

Two reviewers (MOR and NY) independently screened the titles and abstracts of all retrieved studies and identified potentially relevant studies using the predefined study eligibility criteria. To assess their eligibility in detail, they independently critiqued all potentially relevant studies during the full-text screening stage. In both stages, disagreements were resolved through discussion or by a third reviewer (YN or EO), when required. We recorded the reasons for exclusion of all studies in the full-text screening stage and reported them in a PRISMA study flow diagram. We used EndNote (Niles Software) reference management software and the Rayyan Qatar Computing Research Institute tool in the study selection process [26].

Study Quality Assessment

Two reviewers (MOR and NY) independently assessed the risk of bias of the included studies using the Cochrane risk of bias tool. The tool consists of the following domains: random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective outcome reporting, and other bias. For the domain of blinding of participants and personnel, we considered blinding of personnel only as blinding, because the nature of mHealth interventions, may not be possible to study participants. We classified the studies with high, low, and unclear risks of bias based on the Cochrane Handbook [27]. Any discrepancies were solved through discussion or by a third reviewer (YN or EO).

Certainty of Evidence Assessment

We evaluated certainty of evidence for ANC, SBA, and facility delivery outcome using the Grading of Recommendation, Assessment, Development and Evaluation (GRADE) system [28]. The GRADE system considered the judgment on the following factors while assessing the confidence in the evidence based on RCT studies: study limitations (risk of bias), inconsistency (statistical heterogeneity), indirectness (PICO [participants, interventions, comparisons, and outcomes] and applicability), imprecision (number of events and CIs), and publication bias. On the basis of the judgment of each factor, we classified our evidence as follows: (1) high-certainty evidence (further research is very unlikely to change the confidence of the pooled results), (2) moderate-certainty evidence (further research is likely to have an important impact on the confidence of the pooled results and may change the estimate), (3) low-certainty evidence (further research is extremely likely to have an important impact on the confidence

of the pooled results and likely to change the estimate), and (4) very low-certainty evidence (the pooled results have extreme uncertainty) [28]. We used the GRADEpro web-based platform to make a summary of findings table, considering the certainty of evidence.

Data Extraction

Two independent reviewers (MOR and NY) extracted a standard set of data, including study characteristics, participant characteristics, description of interventions, and outcome results from each of the selected studies and were cross-checked. As in the study selection process, disagreements were resolved through discussion or by a third reviewer (YN or EO). We reported key characteristics of the included studies in a separate table. The data characteristics included, but were not limited to, author information, year of publication, study location, study setting, study design, study name, number of participants, study year, age of participants, gestational age at recruitment, comparator, types, function, mode and duration of interventions, intervention provider, and reported outcomes with their results.

Data Synthesis and Analysis

We narratively synthesized study characteristics, participant characteristics, intervention characteristics and key findings among all included studies. To summarize the effect size of mHealth interventions, we used pairwise inverse-variance random-effects meta-analysis separately for each outcome. While pooling the effect size, we used risk ratios (RRs) because our outcome was dichotomous in nature. If the study provided odds ratios (ORs) and the risk of events in the control group (assumed control risk [ACR]), we converted ORs into RRs by using the formula (OR/[1 – ACR × (1 − OR)]) described in the Cochrane Handbook for Systematic Reviews of Intervention [29]. The results of the meta-analysis are presented in forest plots. In the meta-analysis, we used the estimated effective samples for cluster RCT studies by adjusting their design effect if the studies reported unadjusted data. Heterogeneity was assessed by visual inspection of forest plots or tested using the I^2 statistic, and we considered an I^2 value >50% to indicate substantial heterogeneity [29]. If substantial heterogeneity was found in the meta-analyses, we conducted a subgroup analysis based on the direction of interventions (1-way vs 2-way communication) and high baseline coverage of outcomes (80% or more vs <80%) and reported the subgroup-wise pooled estimates for each outcome separately. We used funnel plots and the Egger test to assess publication bias if a meta-analysis includes 10 or more studies. Statistical significance was defined as a P value <.05 for all analyses.

Ethics Approval and Consent to Participate

This study did not require ethical approval or consent to participate, as it used data from published studies.

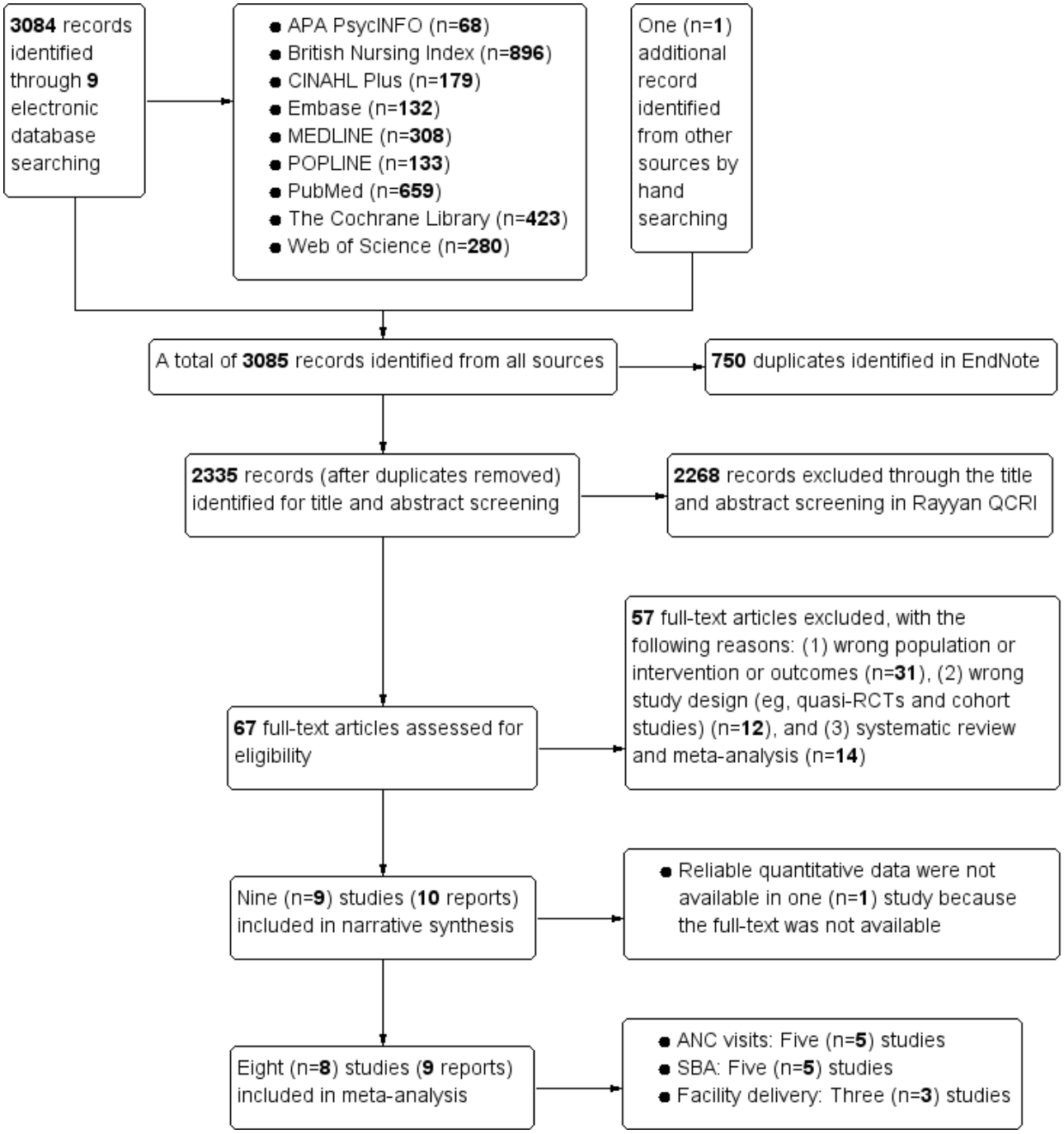
Results

Study Inclusion

A total of 3085 potentially relevant articles were retrieved from all targeted electronic databases and other resources. After removing duplicates in EndNote, 2335 unique articles underwent initial title and abstract screening. As a result, 67 articles were retained for a detailed assessment of study eligibility. After full-text screening, 57 articles that failed to meet the study eligibility criteria were excluded. The reasons for exclusion are reported in the PRISMA study flow diagram (Figure 1). Finally, 10 articles (9 studies) from all resources were found to be suitable for narrative synthesis. We identified 1 gray article by checking the reference lists of all included studies and relevant systematic reviews, but the full text was not available [30]. As we could not check its quantitative information, we narratively synthesized its results and excluded it from our meta-analysis.

Of all 9 included studies, 6 (67%) assessed the effect of mHealth interventions on improving the uptake of ANC visits [17,30-34], 5 (56%) on SBA during delivery [16,31-33,35], and 3 (33%) on facility delivery outcome [31,36,37]. We performed a meta-analysis for ANC visits and SBA outcomes based on the direction of interventions (1-way or 2-way communication) and for facility delivery based on high coverage of outcomes at baseline (80% or more; Figure 1).

Figure 1. PRISMA study flow diagram. ANC: antenatal care; PRISMA: Preferred Reporting Item for Systematic Reviews and Meta-Analyses; QRCI: Qatar Computing Research Institute; RCT: randomized controlled trial; SBA: skilled birth attendance.



Characteristics of the Included Studies

Of all 9 included studies, 6 (67%) were individual RCTs, and 3 (33%) were cluster RCTs (Table 1). Studies were conducted in Brazil [34], China [30], Ethiopia [31], India [32,35], Kenya [33,37], Nigeria [36], and Tanzania [16,17] and were published between 2012 and 2018. A total of 10,348 pregnant women (n=6254, 60.44% in the intervention groups and n=4094, 39.56% in the control groups) participated in the studies. The number of study participants ranged from 116 to 2160 in the intervention groups and from 100 to 1239 in the control groups.

Interventions were carried out through SMS text messaging (5/9, 56% studies), voice calling (1/9, 11% studies), SMS text messaging and mobile voucher (1/9, 11% studies), voice calling and SMS text messaging (1/9, 11% studies), and voice messaging and animation film clips (1/9, 11% studies). Approximately half of the studies used interventions for 1-way communication (4/9, 44% studies) or 2-way communications (5/9, 56% studies). The functions of interventions were categorized into appointment reminder (1/9, 11% studies), health education or advice and appointment reminder (5/9, 56% studies), and health education or preventive health information (3/9, 33% studies). The detailed characteristics of the interventions and the results of the included studies are presented in Table 2.

Table 1. Characteristics of included studies.

Characteristics	Studies, n (%)
Study design (n=9)	
Individual RCTs[a]	6 (67)
Cluster RCTs	3 (33)
Countries (n=9)	
Brazil	1 (11)
China	1 (11)
Ethiopia	1 (11)
India	2 (22)
Kenya	2 (22)
Nigeria	1 (11)
Tanzania	1 (11)
Publication year (n=9)	
2012	1 (11)
2013	1 (11)
2014	1 (11)
2015	1 (11)
2017	3 (33)
2018	2 (22)
Outcomes reported (n=9)	
4 or more ANC[b]	6 (67)
SBA[c]	5 (56)
Facility delivery	3 (33)
Participants (excluded 1 study; n=10,348)	
Intervention group	6254 (60.44)
Control group	4094 (39.56)
Medium of interventions (n=9)	
SMS text messaging	5 (56)
SMS text messaging and mobile voucher	1 (11)
Voice calling	1 (11)
Voice calling and SMS text messaging	1 (11)
Voice messaging and animation film clips	1 (11)
Direction of interventions (n=9)	
1-way communication	4 (44)
2-way communication	5 (56)
Function of interventions (n=9)	
Appointment reminder	1 (11)
Health education or advice and appointment reminder	5 (56)
Health education or preventive health information	3 (33)

[a]RCT: randomized controlled trial.

[b]ANC: antenatal care.

[c]SBA: skilled birth attendance.

Table 2. Characteristics of interventions and results of the included studies.

Authors	Country, participants, study design, sample size	Form of mHealth[a] interventions (medium, direction, and function)	Control group	Reported outcomes	Key findings
Lund et al, 2012 [16]	Tanzania; pregnant women; cluster RCT[b]; intervention: n=1311; control: n=1239	Mobile phone SMS text messaging (twice a week) and mobile voucher; 2-way communication; health education and appointment reminder	Routine ANC[c] and advice	SBA[d] at delivery	Significantly increased skilled delivery attendance among pregnant women (OR[e] 1.69, 95% CI 1.44-1.98)
Luo, 2013 [30]	China; pregnant women; individual RCT; intervention: not available; control: not available	Health education intervention through mobile phone SMS text messaging; 1-way communication	Usual care	4 or more ANC visits	Showed positive effect of health education intervention through mobile phone SMS text messaging on 4 or more ANC visits
Fedha, 2014 [33]	Kenya; pregnant women; individual RCT; intervention: n=191; control: n=206	Mobile phone reminder, updates, and advice: every fortnightly of the next visit to the clinic and given advice and updates on pregnancy; 2-way communication	Routine care with no mobile advice or update support	4 or more ANC visits, SBA, other birth outcomes	Mobile phone services for pregnant women enhanced 4 or more ANC visits (OR: 2.89, 95% CI 1.51-5.53) and SBA (OR 2.73, 95% CI 1.60-4.65)
Lund et al, 2014 [17]	Tanzania; pregnant women; cluster RCT; intervention: n=1311; control: n=1239	Mobile phone SMS text messaging (twice a week) and mobile voucher; 2-way communication; health education and appointment reminder	Routine ANC and advice	4 or more ANC visits; tetanus vaccination and other preventive services	44% of the women received 4 or more ANC visits in the intervention group versus 31% in the control group
Joshi et al, 2015 [35]	India; pregnant women; individual RCT (where most women already use a skilled birth attendant); intervention: n=1162; control: n=581	Preventive health information via voice messages and animation film clips (the automated platform for voice messages); 1-way communication	Usual care (no voice messages and animation films)	SBA, iron and folic acid tablet intake, and knowledge on ANC and their satisfaction	mHealth initiative for promoting higher uptake of ANC services is highly impactful
Atnafu et al, 2017 [31]	Ethiopia; pregnant women; a community-based RCT; intervention: n=1080 (group T1), n=1080 (group T2); control: n=1080	SMS text messaging–based mobile phone reminder intervention; 1-way communication	No SMS text messaging	Role of mobile phone SMS text messaging on MCH[f] outcomes: 4 or more ANC visits, SBA, and facility delivery	Confirmed the positive contribution of SMS text messaging–based mobile phone intervention to most of the selected MCH service indicators, such as improvement in the percentage of recommended number of ANC visit and percentage of delivery attended by health workers
Bangal et al, 2017 [32]	India; pregnant women; individual RCT; intervention: n=200; control: n=200	Mobile phone calls, as reminders about next visit, and SMS text messaging on important aspects of ANC at regular intervals; 1-way communication	Routine ANC and advice as per hospital protocol	Percentage of pregnant women coming for at least four ANC visits and percentage of institutional delivery and postnatal checkups	Women in the intervention group had significantly higher number of ANC visits, consumption of iron tablets, tetanus toxoid immunization, institutional deliveries and postnatal checkups as compared with the control group
Oliveira-Ciabati et al, 2017 [34]	Brazil; pregnant women; cluster RCT; intervention: n=770 (PRENACEL group: n=116); control: n=440	PRENACEL group received a weekly set of SMS text messages with health education and health promotion content related to pregnancy and childbirth and were also able to clarify ANC queries through SMS text messaging; 2-way communication	Routine ANC	ANC, tetanus vaccination, influenza vaccination, and other preventive services	A bidirectional, mobile phone–based, SMS text messaging service is potentially useful for improving the coverage of recommended ANC practices, including syphilis and HIV testing

Authors	Country, participants, study design, sample size	Form of mHealth[a] interventions (medium, direction, and function)	Control group	Reported outcomes	Key findings
Omole et al, 2018 [36]	Nigeria; pregnant women; cluster RCT; intervention: n=260; control: n=248	Pregnant women in the intervention facilities received pregnancy - related health messages and reminders for their ANC appointments through SMS text messaging and also had the opportunity of sending SMS text messages to the project team to seek for health information; 2-way communication	Only received general health messages through SMS text messaging	Attendance of at least four ANC clinic visits and delivery in a health facility	Most of the pregnant women in the intervention group (96.6%) expressed support for the use of SMS text messaging for maternal health promotion. The SMS text messaging–based intervention has a positive effect on facility delivery. A 13% increase was recorded in the rate of facility - based delivery among the control group between the last and index degrees, a much higher 29% increase was recorded among the intervention group
Unger et al, 2018 [37]	Kenya; pregnant women; individual RCT; intervention: n=200 (n=100, 1-way; n=100, 2-way); control: n=100	An automated weekly gestational age-appropriate educational and counseling messaging, and SMS text messaging topic included ANC, pregnancy complications, family planning, infant health, EBF[g], infant immunization, and visit reminders; 1-and 2-way communications	Routine clinic–based counseling and care	Facility delivery, EBF, and contraceptive use	Facility delivery was very high in all 3 arms (98.6%). The mobile WACh SMS text messaging intervention had no effect on the uptake of facility delivery

[a]mHealth: mobile health.

[b]RCT: randomized controlled trial.

[c]ANC: antenatal care.

[d]SBA: skilled birth attendance.

[e]OR: odds ratio.

[f]MCH: maternal and child health.

[g]EBF: exclusive breastfeeding.

Overall Risk of Bias Assessment of the Included Studies

A summary of the overall risk of bias assessment is presented in Figure 2 and Figure 3. RCT studies generally performed well in their risk of bias for random sequence generation (75% low risk, with the remainder unclear), allocation concealment (100% low risk), blinding outcome assessment (37.5% low risk, with the remainder unclear), incomplete outcome data (87.5% low risk), and selective reporting (100% low risk). A total of 3 studies [32,34,37] had a high risk of bias regarding blinding participants and personnel, and another study [34] also had a high risk of bias for incomplete outcome data. In addition, 1 study [33] had an unclear risk for many items. Overall, most of the studies had a low risk of bias.

Figure 2. Risk of bias graph: review of authors' judgments about each risk of bias item presented as percentages across all included studies.

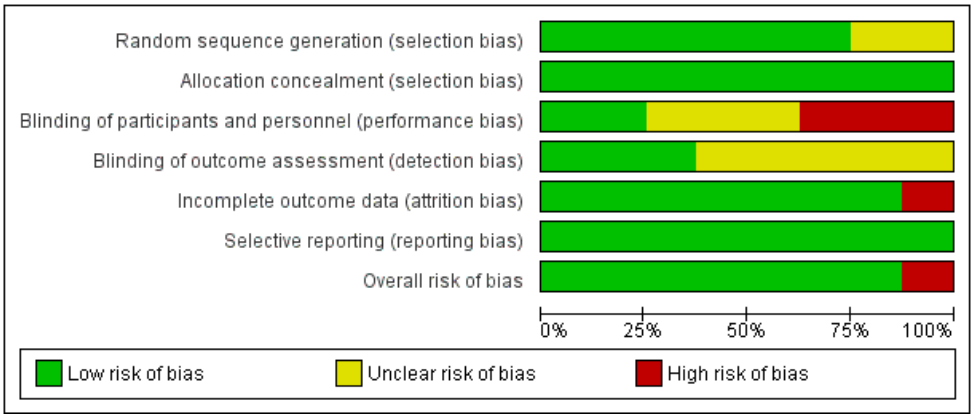
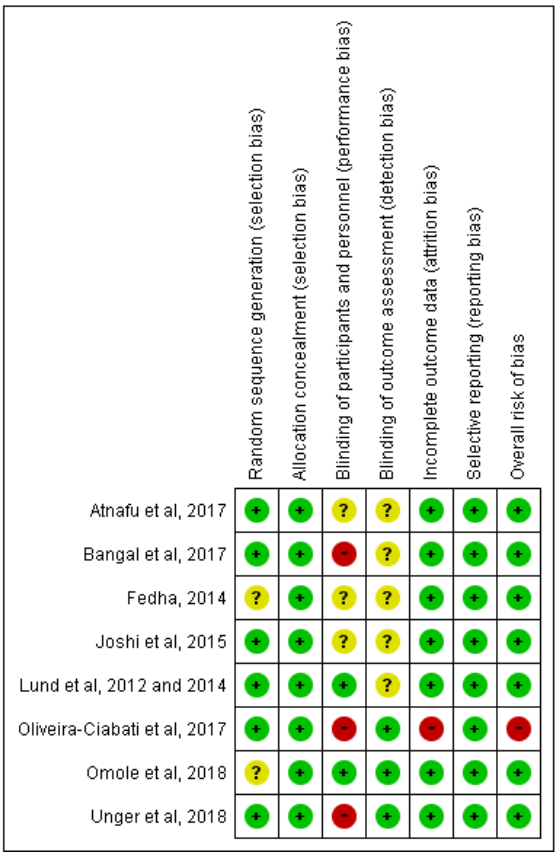


Figure 3. Risk of bias summary: review of authors' judgments about each risk of bias item for each included study [17,31-37].

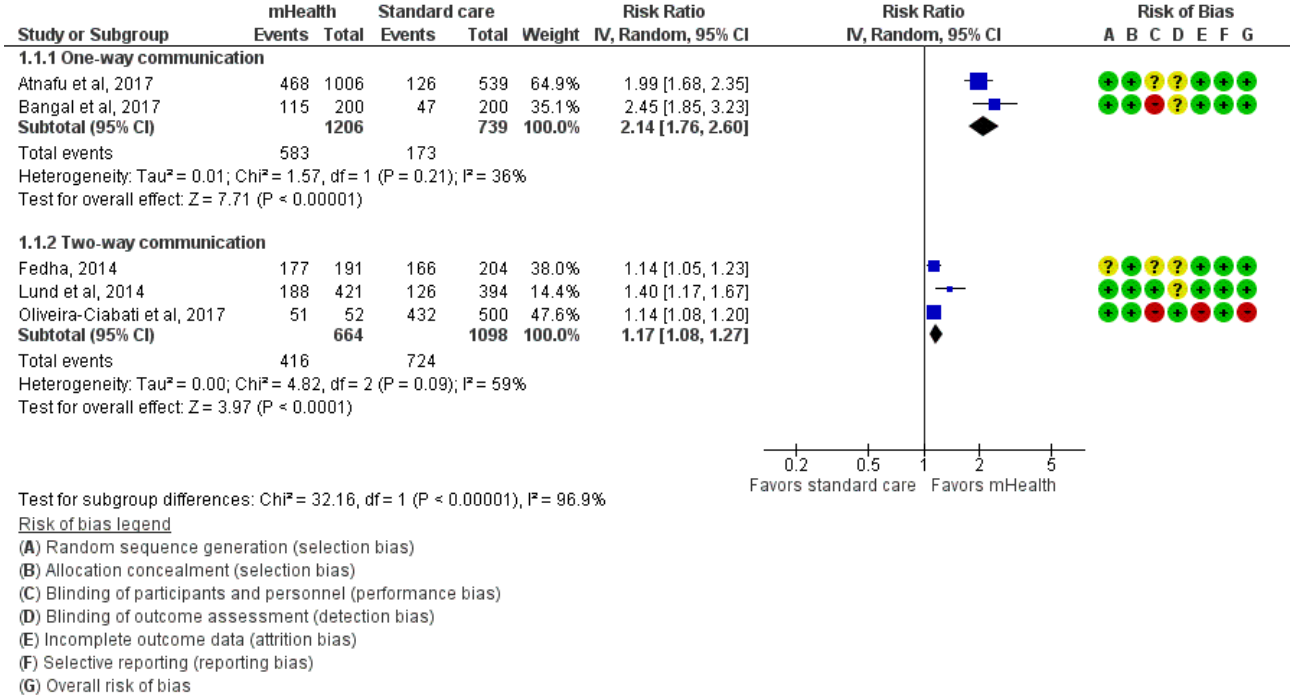


Pooled Effects of mHealth Interventions on the Uptake of ANC Utilization

We performed an inverse-variance random-effects meta-analysis to summarize the effects of 1-way and 2-way mHealth interventions versus standard care on the uptake of 4 or more ANC visits among pregnant women (Figure 4 and Table 3). A total of 2 studies [31,32] consisting of 1945 pregnant women (n=1206, 62.01% in the intervention group and n=739, 37.99% in the control group) implemented a 1-way mHealth intervention and were included in the meta-analysis. The pooled estimates showed a significantly large risk with a 114% increase in 4 or more ANC visits among pregnant women given a 1-way mHealth intervention, compared with the control group (RR 2.14, 95% CI 1.76-2.60, I^2=36%, moderate certainty of evidence).

In total, 3 studies [17,33,34] consisting of 1762 pregnant women (n=664, 37.68% in the intervention group and n=1098, 62.32% in the control group) reported 2-way mHealth intervention, and the pooled estimates from the meta-analysis showed that the risk of 4 or more ANC visits was 17% higher in the 2-way mHealth intervention group than in the control group (RR 1.17, 95% CI 1.08-1.27, I^2=59%, low certainty of evidence). Although all studies reported a positive effect of mHealth interventions, regardless of the direction of interventions, on improving the uptake of ANC visit utilization [17,31-34], we observed a significant difference between these subgroups (I^2=96.9%), which limited their combination.

Figure 4. Meta-analysis for the effect of mHealth interventions versus standard care on 4 or more ANC visits among pregnant women. ANC: antenatal care; mHealth: mobile health.



Test for subgroup differences: Chi² = 32.16, df = 1 (P < 0.00001), I² = 96.9%

Risk of bias legend

(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Blinding of participants and personnel (performance bias)
(**D**) Blinding of outcome assessment (detection bias)
(**E**) Incomplete outcome data (attrition bias)
(**F**) Selective reporting (reporting bias)
(**G**) Overall risk of bias

Table 3. Summary of findings.[a]

Outcomes	Anticipated absolute effects[b]		Relative effect, RR[c] (95% CI)	Number of participants (studies)	Certainty of the evidence (GRADE[d,e])	Comments
	Risk with standard care	Risk with mHealth[f] intervention (95% CI)				
4 or more ANC[g] visits (1-way communication)	234 per 1000	501 per 1000 (412-609)	2.14 (1.76-2.60)	1945 (2 RCTs[h])	Moderate[i]	One-way mHealth intervention likely results in large increase in 4 or more ANC visit utilizations among pregnant women in LMICs[j], and further research may change the estimate.
4 or more ANC visits (2-way communication)	659 per 1000	771 per 1000 (712-837)	1.17 (1.08-1.27)	1762 (3 RCTs)	Low[k,l]	Two-way mHealth intervention may result in an increase in 4 or more ANC visit utilizations among pregnant women in LIMCs and further research is likely to change the estimate.
SBA[m] (1-way communication)	771 per 1000	802 per 1000 (748-848)	1.04 (0.97-1.10)	3460 (3 RCTs)	Very low[i,l,n]	One-way mHealth intervention may not increase SBA during delivery in LMICs, but the evidence is very uncertain.
SBA (2-way communication)	557 per 1000	685 per 1000 (635-740)	1.23 (1.14-1.33)	1212 (2 RCTs)	Moderate[o]	Two-way mHealth intervention likely results in an increase in SBA during delivery among pregnant women in LMICs, and further research may change the estimate.
Facility delivery (<80% at baseline)	360 per 1000	604 per 1000 (467-787)	1.68 (1.30-2.19)	1819 (2 RCTs)	Moderate[o]	mHealth intervention likely results in an increase in facility delivery in LMICs where fewer pregnant women use facility delivery, and further research may change the estimate.
Facility delivery (80% or more at baseline)	990 per 1000	1000 per 1000 (960-1000)	1.01 (0.97-1.04)	300 (1 RCT)	Low[n,p]	mHealth intervention may not increase facility delivery in LMICs where most pregnant women already use facility delivery, and further research is likely to change the estimate.

[a]mHealth intervention compared with standard care for improving ANC utilization, SBA during delivery, and facility delivery among pregnant women. Population: pregnant women; setting: LMICs (Brazil, China, Ethiopia, India, Kenya, Nigeria, and Tanzania); intervention: mHealth intervention; comparison: standard care.

[b]The risk in the intervention group (and its 95% CI) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI).

[c]RR: risk ratio.

[d]GRADE: Grading of Recommendation, Assessment, Development and Evaluation.

[e]The GRADE Working Group grades of evidence. High certainty: we are very confident that the true effect lies close to that of the estimate of the effect. Moderate certainty: we are moderately confident in the effect estimate—the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different. Low certainty: our confidence in the effect estimate is limited—the true effect may be substantially different from the estimate of the effect. Very low certainty: we have very little confidence in the effect estimate—the true effect is likely to be substantially different from the estimate of effect.

[f]mHealth: mobile health.

[g]ANC: antenatal care.

[h]RCT: randomized controlled trial.

[i]Unclear or lack of blinding of participants and outcome assessors.

[j]LMICs: low- and middle-income countries.

[k]Unclear or lack of sequence generation, blinding of participants and outcome assessors, and incomplete outcome data.

[l]Statistical heterogeneity (I^2>50%).

[m]SBA: skilled birth attendance.

[n]CI crossed the threshold.

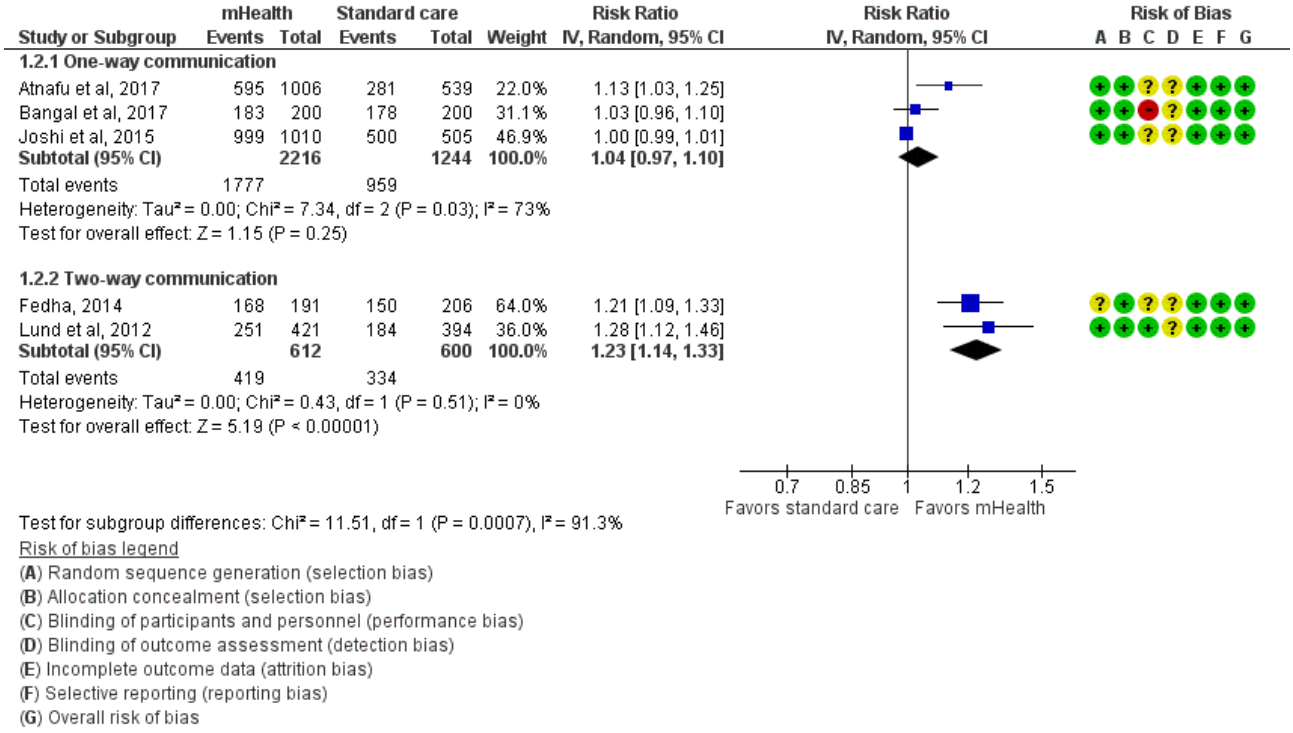
[o]Unclear sequence generation, blinding of participants, and outcome assessors.

[p]Lack of blinding of participants and personnel.

Pooled Effects of mHealth Interventions on SBA at the Time of Delivery

An inverse-variance random-effects meta-analysis was performed to pool the effects of 1-way and 2-way mHealth interventions versus standard care on improving the rate of SBA at the time of delivery (Figure 5 and Table 3). In total, 3 studies [31,32,35] comprising 3460 pregnant women (n=2216, 64.05% in the intervention group and n=1244, 35.85% in the control group) reported a 1-way mHealth intervention in which only 1 study [31] found a positive effect of the intervention on SBA during delivery. The effects of 1-way mHealth interventions on SBA during delivery were not clear; however, the effects were pooled in the meta-analysis (RR 1.04, 95% CI 0.97-1.10, I^2=73%, very low certainty of evidence).

Figure 5. Meta-analysis for the effect of mHealth interventions versus standard care on SBA during delivery. mHealth: mobile health; SBA: skilled birth attendance.

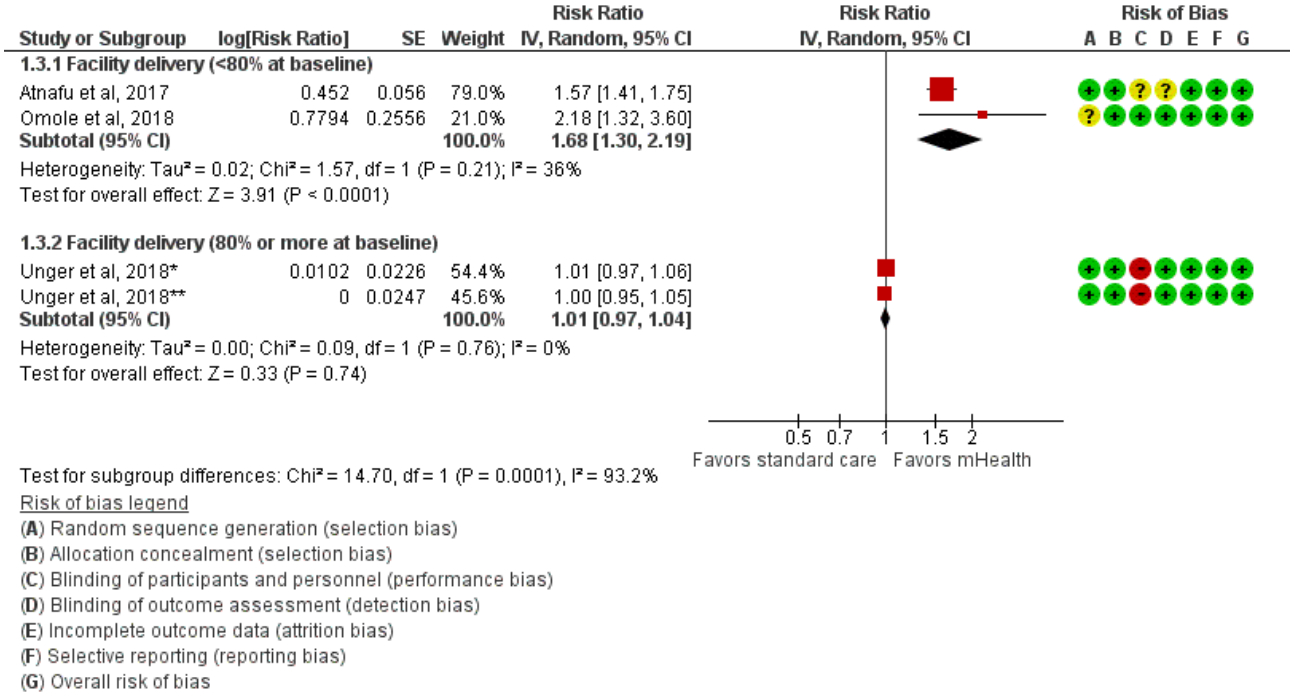


In total, 2 studies [16,33] involving 1212 pregnant women (n=612, 50.50% in the intervention group and n=600, 49.50% in the control group) implemented 2-way mHealth interventions, and their pooled estimates showed that the proportion of SBA during delivery was 23% higher in the 2-way mHealth intervention group than in the control group (RR 1.23, 95% CI 1.14-1.33, I^2=0%, moderate certainty of evidence). Owing to significant subgroup differences in the direction of mHealth interventions (I^2=91.3%), we could not combine their effects.

Pooled Effects of mHealth Interventions on Facility Delivery Among Pregnant Women

Overall, 3 studies assessed the effects of mHealth interventions on facility delivery outcome [31,36,37]. In addition, 1 study [31] found that the probability of facility delivery was 57% higher in the 1-way mHealth intervention group than in the control group (RR 1.57, 95% CI 1.41-1.75). Another study [36] identified a positive result supporting the effectiveness of 2-way mHealth interventions for improving facility delivery (RR 2.18, 95% CI 1.32-3.60). However, 1 study reported an unclear effect of the mHealth interventions, irrespective of direction, on the uptake of facility delivery [37]. We conducted a subgroup analysis considering high coverage of facility delivery at baseline (80% or more) and found a positive effect of mHealth interventions on improving the uptake of facility delivery in LMICs, where fewer pregnant women used facility delivery at baseline (RR 1.68, 95% CI 1.30-2.19, I^2=36%, moderate certainty of evidence); however, the effects were unclear where most pregnant women already used facility delivery (RR 1.01, 95% CI 0.97-1.04, I^2=0%, low certainty of evidence). Significant subgroup differences (I^2=93.2%) limited the combination of these effects (Figure 6 and Table 3).

Figure 6. Meta-analysis for the effect of mHealth interventions versus standard care on facility delivery among pregnant women. mHealth: mobile health; *1-way mHealth intervention; **2-way mHealth intervention.



Subgroup Differences

Significant subgroup differences (P<.001) were found in all meta-analyses, which limited the combination of the effects of all included studies (Figures 4-6). It included an estimated 96.9% subgroup variance for 4 or more ANC visits, 91.3% for SBA at the time of delivery, and 93.2% for facility delivery.

Publication Bias

Although we planned to assess the publication bias of our meta-analyses, we did not perform the analysis because of the limited number of studies.

Discussion

Principal Findings

This meta-analysis identified a statistically significant positive effect of mHealth interventions, regardless of the direction of interventions (1-way or 2-way communications), on improving ANC care utilization of healthy pregnant women in LMICs. Only 2-way mHealth interventions were effective in improving SBA during delivery, but the effects were unclear for 1-way mHealth interventions compared with standard care. For facility delivery among healthy pregnant women, the interventions were effective in settings where fewer pregnant women used facility delivery. Most studies that combined multiple mHealth interventions were implemented in Brazil, China, Ethiopia, India, Kenya, Nigeria, and Tanzania. The functions or directions of interventions varied among the included studies, such as 1-way or 2-way communication, appointment reminder, or health advice. All studies, except one [34], had low concerns of their methodological qualities. However, high statistical heterogeneity limited the combination of subgroups in all the meta-analyses.

These findings are consistent with a systematic review reporting that mHealth interventions had a positive effect and resulted in a 43% increase in the uptake of recommended ANC visits among pregnant women [13]. In another systematic review conducted by Wagnew et al [19], SMS text messaging had positive effects on the uptake of 4 or more ANC visits (OR 2.74, 95% CI 1.41-5.32) and SBA (OR 1.82, 95% CI 1.33-2.49) in LMICs [19], which strongly supports our findings. Although the availability of high-level evidence such as RCTs on mHealth interventions targeting healthy pregnant women's health care utilization is still very limited, our findings have generated promising results regarding the positive effects of mHealth interventions, regardless of their directions, on improving recommended ANC utilization, SBA during delivery, and facility delivery in resource-limited countries. The findings are also consistent with other systematic reviews that found that mHealth tools are effective in influencing maternal and child health service utilization by enhancing the uptake of recommended ANC and postnatal care services, including SBA, at the time of delivery and institutional delivery [12,14,18].

The effects of 1-way mHealth interventions on SBA at the time of delivery were not clear in the 2 included studies [32,35], which was reflected in our pooled results. Our meta-analysis identified a significant difference in the likelihood of SBA during delivery, which was higher in the 2-way mHealth intervention group than in the control group. Our findings support the effectiveness of mHealth interventions in improving facility delivery in settings where fewer pregnant women use the service. Consistent with our findings, a systematic review conducted by Colaci et al [38] reported that mHealth interventions offered an opportunity to increase the acceptability of prenatal and obstetric care, including SBA at the time of delivery. This is because mHealth interventions, either health care reminders or health advice, boost self-efficacy and access

to care among pregnant women and create closer interactions with health service providers. Furthermore, mHealth interventions have been used as appointment reminders and can provide basic health information, notably throughout the pregnancy period. As novel and more cost-effective systems are being sought to promote health care utilization in underserved areas, particularly in remote and rural areas, this intervention offers a potential cost-effective solution. This study strongly supports the use of mHealth interventions, either 1-way or 2-way communication, to enhance the uptake of maternal health care services such as ANC, SBA, or facility delivery, by changing health care behavior among pregnant women in similar settings.

Strengths

This review had several strengths. To the best of our knowledge, this is the first comprehensive review that conducted meta-analyses for different subgroups, even in similar settings, for each outcome, such as the direction of interventions (1-way or 2-way communication) and high coverage of outcomes at baseline, and identified some novel findings not seen in previous studies. For example, 2-way (not 1-way) mHealth interventions likely result in an increase in SBA during delivery in resource-limited settings. Likewise, mHealth interventions likely result in an increase in facility delivery in LMICs where fewer pregnant women use the service but may not increase in settings where most pregnant women already use facility delivery. Interestingly, the certainty of the evidence was moderate, indicating that our estimates were likely to be close to the true effect.

Second, we conducted a comprehensive search of electronic databases without any limit on language, date, or type of publication. We also checked the reference lists of the included studies and other relevant systematic reviews to avoid missing any potentially relevant studies, and identified 1 additional study. Third, this study considered high-level studies such as RCTs and cluster RCTs and performed meta-analyses well. In the meta-analyses, we used the estimated effective samples for cluster RCT studies; however, the studies did not adjust their design effect. In addition, with the larger sample size, we enhanced the statistical power to provide more precise and reliable effect estimates.

Limitations

Despite the positive effects of mHealth interventions reported in our meta-analyses, this review had several limitations. One of the important limitations is the limited number of RCT studies included in the meta-analyses (8 studies). All included studies were reported from only 7 LMICs, and some used small sample sizes that may compromise representativeness. In addition, approximately half of the studies tried to combine multiple mHealth interventions, making it difficult to understand the extent to which each intervention contributed to the observed results. For instance, a study assessed the effectiveness of a combined intervention of mobile SMS text messaging and mobile voucher [16,17], another studied a voice messaging and animation film clip intervention [35] and a mobile phone call and SMS text messaging intervention [32]. Likewise, some studies tried to combine multiple functions of mHealth interventions (eg, health education or advice and appointment reminder), which made it difficult to determine the extent to which each function of intervention contributed to the reported outcomes [16,17,32,33,36].

Although the mHealth interventions were effective in improving facility delivery in the settings where fewer pregnant women used the service, we need to interpret these results with caution because it is not clear if the benefits presented regarding facility delivery are a function of the mHealth interventions or greater health literacy and knowledge of navigating the local health care system among users and because it is also not clear if the increase in ANC and SBA are due to the mHealth technology or access to health care. With the mHealth interventions, we also need to consider other factors such as sociocultural norms and beliefs during perinatal periods, perinatal care availability and resources, or other structural factors to improve ANC, SBA, and facility delivery utilization among pregnant women.

We performed an inverse-variance random-effects meta-analysis to summarize the effect of mHealth interventions, irrespective of their directions or high coverage of outcomes at baseline, on the uptake of ANC visits, SBA, and facility delivery utilization among pregnant women in LMICs; however, we observed a significant difference between these subgroups that limited their combination. Within the subgroups, we did not observe considerable heterogeneity among the included studies.

Implications for Future Research

mHealth interventions (1-way or 2-way communication) may contribute toward improving maternal health care–seeking behavior throughout the pregnancy cycle. Public health researchers, practitioners, and policy makers should consider such interventions in resource-limited settings. It is reasonable to use mHealth interventions in resource-limited settings, as this study found a positive effect of different forms of intervention on the uptake of ANC visits, SBA, and facility delivery utilization. This study could play an important role in addressing the Sustainable Development Goal of Maternal and Child Health, as it provides insights and evidence-informed recommendations for the utilization of different forms of mHealth interventions in addressing maternal health care challenges in LMICs. However, owing to the limited number of RCT studies that met the study eligibility criteria in this review, further large-scale RCTs are suggested to be implemented in resource-limited countries, particularly where service utilization is quite low among pregnant women. In recent times, most LMICs have recognized the need for appropriate technology use strategies to promote the utilization of their existing health systems, which also justifies the necessity of further evidence of technology-based health care interventions such as mHealth interventions.

Conclusions

Although mHealth interventions (1-way or 2-way communication) can improve the uptake of ANC and skilled delivery care utilization, more rigorous evaluations are required to strengthen the evidence of different forms of mHealth interventions for improving the existing health care utilization among healthy pregnant women in LMICs. This systematic

review and meta-analysis will help public health researchers or policy makers in designing and implementing mHealth interventions in resource-limited settings, as this study identified some novel findings not found in previous reviews.

Acknowledgments

The authors would like to thank St. Luke's International University and The University of Tokyo for providing access to their library for the database search and acquisition of articles. The authors are thankful to the Japanese Ministry of Education, Culture, Sports, Science, and Technology for providing a Ministry of Education, Culture, Sports, Science, and Technology doctoral program scholarship to MOR, which allowed the authors to carry out this research as a partial fulfillment of the doctoral program. The authors would also like to express their gratitude to Sarah E Porter, PhD, RN, academic adviser, St. Luke's International University, for the final checking and editing of the manuscript.

Data Availability

No additional data available.

Authors' Contributions

MOR and EO conceived and designed the review and were responsible for the certainty of evidence assessment. MOR contributed to electronic database searching, data analysis, manuscript writing, and project management. MOR and NY contributed to study screening, data extraction, and study quality assessment. EO and YN contributed to the critical revision of the methodology and intellectual content. EO contributed to project supervision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy of electronic databases.
[DOCX File , 43 KB - jmir_v24i4e34061_app1.docx]

References

1. World Health Organization, United Nations Children's Fund, United Nations Population Fund, World Bank Group, United Nations Population Division. Trends in maternal mortality 2000 to 2017: estimates by WHO, UNICEF, UNFPA, World Bank Group and the United Nations Population Division. Geneva, Switzerland: World Health Organization; 2019.
2. Langlois EV, Miszkurka M, Zunzunegui MV, Ghaffar A, Ziegler D, Karp I. Inequities in postnatal care in low- and middle-income countries: a systematic review and meta-analysis. Bull World Health Organ 2015 Apr 01;93(4):259-70G [FREE Full text] [doi: 10.2471/BLT.14.140996] [Medline: 26229190]
3. Mrisho M, Obrist B, Schellenberg JA, Haws RA, Mushi AK, Mshinda H, et al. The use of antenatal and postnatal care: perspectives and experiences of women and health care providers in rural southern Tanzania. BMC Pregnancy Childbirth 2009 Mar 04;9:10 [FREE Full text] [doi: 10.1186/1471-2393-9-10] [Medline: 19261181]
4. Regassa N. Antenatal and postnatal care service utilization in southern Ethiopia: a population-based study. Afr Health Sci 2011 Sep;11(3):390-397 [FREE Full text] [Medline: 22275929]
5. United Nations Children's Fund. Healthy Mothers, Healthy Babies: taking stock of maternal health. New York, NY, USA: United Nations Children's Fund; Jun 2019.
6. United Nations Children's Fund. Delivery care. United Nations Children's Fund. 2021 Apr. URL: https://data.unicef.org/topic/maternal-health/delivery-care/ [accessed 2022-02-12]
7. Bazile J, Rigodon J, Berman L, Boulanger VM, Maistrellis E, Kausiwa P, et al. Intergenerational impacts of maternal mortality: qualitative findings from rural Malawi. Reprod Health 2015 May 06;12:S1 [FREE Full text] [doi: 10.1186/1742-4755-12-S1-S1] [Medline: 26000733]
8. Maternal mortality. World Health Organization. 2019 Sep 19. URL: https://www.who.int/news-room/fact-sheets/detail/maternal-mortality [accessed 2022-02-12]
9. Amoakoh-Coleman M, Borgstein AB, Sondaal SF, Grobbee DE, Miltenburg AS, Verwijs M, et al. Effectiveness of mHealth interventions targeting health care workers to improve pregnancy outcomes in low- and middle-income countries: a systematic review. J Med Internet Res 2016 Aug 19;18(8):e226 [FREE Full text] [doi: 10.2196/jmir.5533] [Medline: 27543152]
10. Labrique AB, Vasudevan L, Kochi E, Fabricant R, Mehl G. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. Glob Health Sci Pract 2013 Aug;1(2):160-171 [FREE Full text] [doi: 10.9745/GHSP-D-13-00031] [Medline: 25276529]

11. 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]
12. Atnafu A, Bisrat A, Kifle M, Taye B, Debebe T. Mobile health (mHealth) intervention in maternal and child health care: evidence from resource-constrained settings: a review. Ethiop J Health Dev 2015;29(3):139-196.
13. Chen H, Chai Y, Dong L, Niu W, Zhang P. Effectiveness and appropriateness of mHealth interventions for maternal and child health: systematic review. JMIR Mhealth Uhealth 2018 Jan 09;6(1):e7 [FREE Full text] [doi: 10.2196/mhealth.8998] [Medline: 29317380]
14. Feroz A, Perveen S, Aftab W. Role of mHealth applications for improving antenatal and postnatal care in low and middle income countries: a systematic review. BMC Health Serv Res 2017 Nov 07;17(1):704 [FREE Full text] [doi: 10.1186/s12913-017-2664-7] [Medline: 29115992]
15. Lee S, Nurmatov U, Nwaru B, Mukherjee M, Grant L, Pagliari C. Effectiveness of mHealth interventions for maternal, newborn and child health in low- and middle-income countries: systematic review and meta-analysis. J Glob Health 2016 Jun;6(1):010401 [FREE Full text] [doi: 10.7189/jogh.06.010401] [Medline: 26649177]
16. Lund S, Hemed M, Nielsen BB, Said A, Said K, Makungu MH, et al. Mobile phones as a health communication tool to improve skilled attendance at delivery in Zanzibar: a cluster-randomised controlled trial. BJOG 2012 Sep;119(10):1256-1264. [doi: 10.1111/j.1471-0528.2012.03413.x] [Medline: 22805598]
17. Lund S, Nielsen BB, Hemed M, Boas IM, Said A, Said K, et al. Mobile phones improve antenatal care attendance in Zanzibar: a cluster randomized controlled trial. BMC Pregnancy Childbirth 2014 Jan 17;14:29 [FREE Full text] [doi: 10.1186/1471-2393-14-29] [Medline: 24438517]
18. Sondaal SF, Browne JL, Amoakoh-Coleman M, Borgstein A, Miltenburg AS, Verwijs M, et al. Assessing the effect of mhealth interventions in improving maternal and neonatal care in low- and middle-income countries: a systematic review. PLoS One 2016;11(5):e0154664 [FREE Full text] [doi: 10.1371/journal.pone.0154664] [Medline: 27144393]
19. Wagnew F, Dessie G, Alebel A, Mulugeta H, Belay YA, Abajobir AA. Does short message service improve focused antenatal care visit and skilled birth attendance? A systematic review and meta-analysis of randomized clinical trials. Reprod Health 2018 Nov 22;15(1):191 [FREE Full text] [doi: 10.1186/s12978-018-0635-z] [Medline: 30466453]
20. World Health Organization. Mhealth: New Horizons for Health Through Mobile Technologies: Second Global Survey on eHealth (Global Observatory for eHealth). Geneva, Switzerland: World Health Organization; 2011.
21. Dol J, Richardson B, Tomblin Murphy G, Aston M, McMillan D, Campbell-Yeo ML. Impact of mobile health (mHealth) interventions during the perinatal period for mothers in low- and middle-income countries: a systematic review. JBI Database System Rev Implement Rep 2019 Aug;17(8):1634-1667. [doi: 10.11124/JBISRIR-2017-004022] [Medline: 31404051]
22. Watterson JL, Walsh J, Madeka I. Using mHealth to improve usage of antenatal care, postnatal care, and immunization: a systematic review of the literature. Biomed Res Int 2015;2015:153402 [FREE Full text] [doi: 10.1155/2015/153402] [Medline: 26380263]
23. Schiavo JH. PROSPERO: an international register of systematic review protocols. Med Ref Serv Q 2019;38(2):171-180. [doi: 10.1080/02763869.2019.1588072] [Medline: 31173570]
24. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 01;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
25. World Bank Country and Lending Groups. The World Bank. 2019. URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups [accessed 2022-03-30]
26. 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]
27. 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 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
28. 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]
29. Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane handbook for systematic reviews of interventions. 2nd ed. Hoboken, NJ, USA: John Wiley & Sons; Sep 20, 2019.
30. Luo M. Effectiveness of health education intervention through mobile telephone message in pregnancy. Changsha, China: Central South University; 2013.
31. Atnafu A, Otto K, Herbst CH. The role of mHealth intervention on maternal and child health service delivery: findings from a randomized controlled field trial in rural Ethiopia. Mhealth 2017;3:39 [FREE Full text] [doi: 10.21037/mhealth.2017.08.04] [Medline: 29184891]
32. Bangal VB, Borawake SK, Gavhane SP, Aher KH. Use of mobile phone for improvement in maternal health: a randomized control trial. Int J Reprod Contracept Obstet Gynecol 2017 Nov 23;6(12):5458. [doi: 10.18203/2320-1770.ijrcog20175260]

33. Fedha T. Impact of mobile telephone on maternal health service care: a case of Njoro division. Open J Prev Med 2014;04(05):365-376. [doi: 10.4236/ojpm.2014.45044]
34. Oliveira-Ciabati L, Vieira CS, Franzon AC, Alves D, Zaratini FS, Braga GC, et al. PRENACEL - a mHealth messaging system to complement antenatal care: a cluster randomized trial. Reprod Health 2017 Nov 07;14(1):146 [FREE Full text] [doi: 10.1186/s12978-017-0407-1] [Medline: 29116028]
35. Joshi S, Patil N, Hegde A. Impact of mhealth initiative on utilization of ante natal care services in rural Maharashtra, India. Indian J Matern Child Health 2015;17(2):1-7.
36. Omole O, Ijadunola MY, Olotu E, Omotoso O, Bello B, Awoniran O, et al. The effect of mobile phone short message service on maternal health in south-west Nigeria. Int J Health Plann Manage 2018 Jan;33(1):155-170. [doi: 10.1002/hpm.2404] [Medline: 28332259]
37. Unger JA, Ronen K, Perrier T, DeRenzi B, Slyker J, Drake AL, et al. Short message service communication improves exclusive breastfeeding and early postpartum contraception in a low- to middle-income country setting: a randomised trial. BJOG 2018 Nov;125(12):1620-1629 [FREE Full text] [doi: 10.1111/1471-0528.15337] [Medline: 29924912]
38. Colaci D, Chaudhri S, Vasan A. mHealth interventions in low-income countries to address maternal health: a systematic review. Ann Glob Health 2016;82(5):922-935 [FREE Full text] [doi: 10.1016/j.aogh.2016.09.001] [Medline: 28283147]

Abbreviations

ANC: antenatal care
GRADE: Grading of Recommendation, Assessment, Development and Evaluation
LMIC: low- and middle-income country
mHealth: mobile health
OR: odds ratio
PICO: participants, interventions, comparisons, and outcomes
PICOS: participants, interventions, comparisons, outcomes, and study designs or settings
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RR: risk ratio
SBA: skilled birth attendance

Edited by A Mavragani; submitted 05.10.21; peer-reviewed by R Farrell, J Stingone; comments to author 30.12.21; revised version received 15.02.22; accepted 15.03.22; published 22.04.22.

Please cite as:
Rahman MO, Yamaji N, Nagamatsu Y, Ota E
Effects of mHealth Interventions on Improving Antenatal Care Visits and Skilled Delivery Care in Low- and Middle-Income Countries: Systematic Review and Meta-analysis
J Med Internet Res 2022;24(4):e34061
URL: https://www.jmir.org/2022/4/e34061
doi:10.2196/34061
PMID:35451987

Review

Challenges in Participant Engagement and Retention Using Mobile Health Apps: Literature Review

Saki Amagai[1], BA; Sarah Pila[1], BS, MA, PhD; Aaron J Kaat[1], BA, MA, PhD; Cindy J Nowinski[1], BS, MA, MD, PhD; Richard C Gershon[1], BS, PhD

Northwestern University Feinberg School of Medicine, Chicago, IL, United States

Corresponding Author:
Saki Amagai, BA
Northwestern University Feinberg School of Medicine
625 North Michigan Avenue, 2700 suite
Chicago, IL, 60613
United States
Phone: 1 312 503 1725
Email: saki.amagai@northwestern.edu

Abstract

Background: Mobile health (mHealth) apps are revolutionizing the way clinicians and researchers monitor and manage the health of their participants. However, many studies using mHealth apps are hampered by substantial participant dropout or attrition, which may impact the representativeness of the sample and the effectiveness of the study. Therefore, it is imperative for researchers to understand what makes participants stay with mHealth apps or studies using mHealth apps.

Objective: This study aimed to review the current peer-reviewed research literature to identify the notable factors and strategies used in adult participant engagement and retention.

Methods: We conducted a systematic search of PubMed, MEDLINE, and PsycINFO databases for mHealth studies that evaluated and assessed issues or strategies to improve the engagement and retention of adults from 2015 to 2020. We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. Notable themes were identified and narratively compared among different studies. A binomial regression model was generated to examine the factors affecting retention.

Results: Of the 389 identified studies, 62 (15.9%) were included in this review. Overall, most studies were partially successful in maintaining participant engagement. Factors related to particular elements of the app (eg, feedback, appropriate reminders, and in-app support from peers or coaches) and research strategies (eg, compensation and niche samples) that promote retention were identified. Factors that obstructed retention were also identified (eg, lack of support features, technical difficulties, and usefulness of the app). The regression model results showed that a participant is more likely to drop out than to be retained.

Conclusions: Retaining participants is an omnipresent challenge in mHealth studies. The insights from this review can help inform future studies about the factors and strategies to improve participant retention.

(J Med Internet Res 2022;24(4):e35120) doi:10.2196/35120

KEYWORDS

mobile phone; mHealth; retention; engagement

Introduction

Background

Today, 85% of the US population owns a smartphone device and daily use averages 4.5 hours [1]. With the rise in smartphone ownership and use, smartphones have become one of the most accessible and cost-effective platforms in health care and research. Smartphones are also pervasive across age, race, and socioeconomic status, allowing researchers to inexpensively reach out to myriad of population-level samples with ease. Specifically, the adoption of mobile health (mHealth) apps—mobile apps that help monitor and manage health of participants through smartphone devices, tablets, and other wireless network devices—has been increasing in the research sphere. The mHealth market is expected to grow at a compound annual growth rate of 17.6% from 2021 to 2028 [2]. In addition, the recent COVID-19 pandemic has led to a rise in the downloads and use of various mHealth apps, highlighting the importance of technology-based remote monitoring and

diagnosis for continued advancement in modern health care (eg, [3]).

The greatest advantage of mHealth apps is their convenience. Unlike traditional in-person study settings, mHealth apps can be easily accessed from anywhere at the participant's convenience. Using apps for remote assessment allows participants to make fewer site visits, substantially reducing the burden of travel and the time needed to participate in laboratory studies. With lowered barriers, it becomes easier for participants to conduct repeated testing and share real-time data based on their daily life experiences. Some mHealth research apps also allow participants to directly communicate with their providers via the app, which may enhance both the effectiveness of the app in its goals (eg, in disease management) and adherence in research studies. Given the ubiquity of smartphones among US adults, mHealth apps for research stand to better meet participants where they are at.

For researchers, the convenience of mHealth apps allows them to reach out to large and diverse participant populations more inexpensively and efficiently than traditional in-person studies. Recently, several large-scale studies were able to recruit thousands of participants within a span of a few months using Apple's ResearchKit framework (eg, [4-7]). Using these apps, researchers can monitor day-to-day fluctuations of a wide range of real-time data. For example, self-reported emotional outcomes can be assessed together with passive location data to then infer many other real-time variables, such as physical activity, weather, and air quality, that could potentially affect mood throughout the day.

Despite these overwhelming advantages, many mHealth studies experience high participant attrition rooted in the fundamental challenges of keeping participants engaged. For example, consistent with other large-scale mHealth studies, the notable Stanford-led MyHeart Counts study experienced substantial dropout rates; mean engagement with the app was only 4.1 days [8]. It is a ubiquitous problem across all app uses; approximately 71% of app users are estimated to disengage within 90 days of a new activity [9].

It is imperative for mHealth studies to minimize participant dropout, as substantial attrition may reduce study power and threaten the representativeness of the sample. A potential benefit of mHealth research studies should include easier access to well-balanced, representative samples in terms of race, ethnicity, gender, age, education status, etc. However, given that many studies systematically lose participants, systematic differences between participants who are not completing the studies and those who complete the studies, may introduce bias to the sample. Differential retention makes it difficult to conclude whether any observed effects were caused by the intervention itself, retention bias, or inherent differences between groups. Participant dropout also precludes the conduct of longitudinal research.

In an effort to understand the various factors affecting participant retention, recent studies have evaluated recruitment and retention in several remotely conducted mHealth studies. In their cross-study evaluation of 100,000 participants, Pratap et al [10] analyzed individual-level study app use data from 8 studies that accumulated nearly 3.5 million remote health evaluations. Their study identified 4 factors that were significantly associated with increased participant retention: clinician referral, compensation, having the clinical condition of interest, and older age. However, the study only focused on large-scale observational studies led by the Sage or Research Kit, with especially low barriers to entry and exit, thus questioning the appropriateness of applying these findings to other small-scale studies with varying levels of participation. To our knowledge, other published systematic reviews and meta-analyses on engagement and retention are narrowly focused on one subfield of mHealth research, such as depression or smoking, or are only based on a few studies. Thus, it is impossible to extrapolate their findings to other mHealth apps that are not in the same subfield [11-13].

Retention strategies could be incorporated as app features to prevent participant dropout. For example, gamifying mHealth apps by incorporating badges, competitions, and rankings should make the experience more enjoyable and provide better incentives for participants. The addition of reminders, such as push notifications and SMS text messages, and enabling communication with clinicians are also expected to increase participant retention. However, the extent of their effectiveness in successfully engaging and retaining participants is not yet well defined.

This Study

One fundamental challenge for many mHealth app studies is the rapid and substantial participant dropout. This study aimed to better understand how mHealth studies conducted in the past 5 years have addressed the challenges of participant engagement and retention. We conducted a systematic review of the literature to identify notable factors and strategies used in participant engagement and retention. We hypothesize that participant attrition will be high overall and that there will be shared challenges across different studies that researchers should be cognizant of in future research.

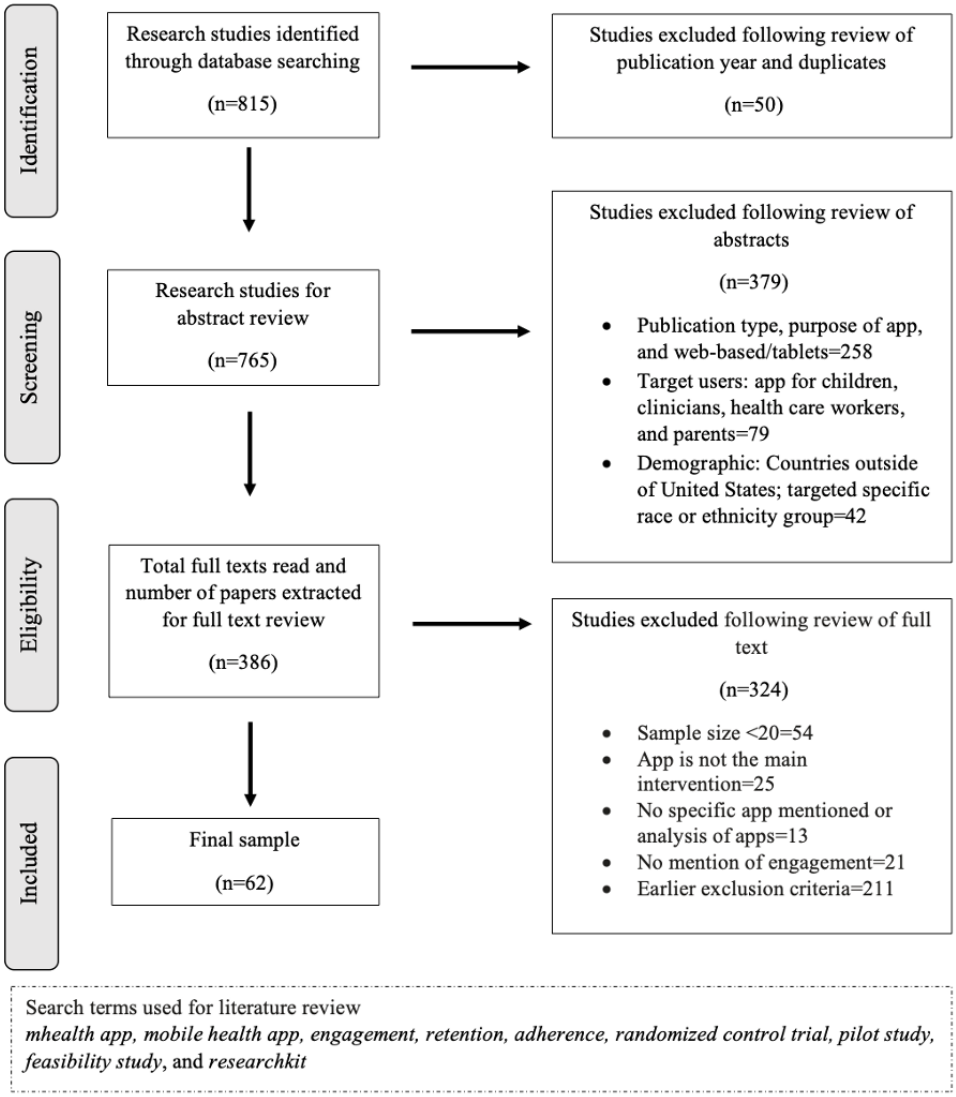
Methods

Search Criteria and Eligibility

Our methodology was guided by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [14]. We identified 3 main databases for this search: PubMed, MEDLINE, and PsycInfo. This review aimed to evaluate the engagement and retention of adults in evaluation research on mHealth apps. Study inclusion criteria were peer-reviewed publications within the last 5 years (January 2015 to October 2020), conducted within the United States, with a minimum of 20 adults. Refer to Figure 1 for more details on the search strategy and exact search terms. Although mHealth takes many forms, we were exclusively interested in mobile-based apps rather than SMS text messaging, tablets, or web-based interventions. We used a variety of research methods and designs, including qualitative, quantitative, or mixed methods. To conduct a comprehensive analysis, we also included mHealth apps in various research areas, ranging from smoking cessation to cardiovascular health research. Articles that were written purely as study protocols or design pieces were excluded. As we were primarily interested in mHealth for intervention

purposes, we excluded studies that used fitness app data exclusively (eg, Fitbit and digital pedometers), unless they were specifically geared toward a particular health population (eg, breast cancer survivors and patients with other chronic illnesses). We also excluded evaluations of mHealth apps that focused solely on participant education or where the clinician was the focus of the intervention.

Figure 1. Study selection flowchart.



Data Extraction and Analysis

We initially extracted basic information from each study: title, year, author, target population, operating system, definition of engagement, sample size and type (clinical vs nonclinical), participant age, study duration, main findings, possible implications, and whether participants were compensated. Most of these data were analyzed in a quantitative manner and are described as descriptive statistics in the Results section (ie, app system, sample size, sample type, compensation, and participant age). These data were also used to develop a binomial regression model to determine the factors affecting retention. For the remaining variables, such as the definition of engagement, findings, and implications, we extracted whole sentences or paragraphs that mentioned these items. Following the narrative approach described by Mays et al [15], the first (SA) and second author (SP) analyzed the findings and implications of the initial sample extraction to determine potential themes around retention and engagement. At this point, codes were applied to individual considerations of retention and engagement (or lack thereof) within the articles. After several readings of all extracted findings and implications, the second author initially determined approximately 5 themes related to support and barriers to engagement. These themes were developed from sets of codes, and these sets of codes were considered a *theme* once they were identified in 2 unique articles. After discussion and agreement with the first author, the second author reread the full-text articles to continue to refine these themes and consolidate the findings. We reached saturation when we could no longer identify new themes during the analysis, a process Saunders et al [16] considered *inductive thematic saturation*. Descriptions of these themes are presented in the qualitative findings of the Results section. The definitions of engagement themes and success rates were also processed in a similar way, and they are described in the quantitative findings of the Results section.

Results

Final Sample of Studies

After locating all studies that met our search criteria (N=389) and downloading the full text, the first and second authors briefly reviewed the abstracts and full text to determine whether the selected studies met our inclusion criteria. In this process, we confirmed that all the studies that should have been excluded were, in fact, excluded. In a random sample of 100 articles, the authors agreed on 91% (91/100) of these decisions. After reaching an agreement about the remaining 9% (9/100), the first and second author divided the remaining articles for a more detailed review. The final sample comprised 62 articles. Refer to Figure 1 for the study selection flowchart and Multimedia Appendix 1 [4-8,17-73] for characteristics of the studies.

Descriptive Findings

The mean age across the users of mHealth apps among the 62 studies was 44.14 years (range 32-64.9 years), and the majority were of a clinical population (48/62, 77%). The sample size ranged from 20 (our predetermined minimum) to 101,108 participants, with most studies reporting a sample size of <100 (34/62, 55%). Most studies reported compensating participants (34/62, 55%). Most articles described apps that were available for both iPhone and Android users (29/62, 47%). Refer to Table 1 for more information about the descriptive statistics.

Table 1. Descriptive statistics of the 62 studies.

	Values
Age of users (years), mean (range)	44.14[a] (32-64.9)
Sample size[b] (n=108), n (%)	
20-49	17 (27)
50-99	17 (27)
100-499	12 (19)
>500	16 (25)
Platform, n (%)	
Android	11 (17)
iPhone	11 (17)
Both	29 (46)
Not reported	11 (17)
Clinical vs nonclinical, n (%)	
Clinical	48 (77)
Nonclinical	14 (22)
Compensation, n (%)	
Provided compensation	34 (54)
No compensation	28 (45)
Success code, n (%)	
Not successful	13 (21)
Partially successful	42 (67)
Successful	6 (9)
Not able to calculate	1 (1)

[a]Adequate information to calculate the mean age was not provided for 13 out of the 62 studies. We excluded these studies from the mean age calculation.

[b]The sample size ranged from 20 to 101. The median was 90.5 (IQR 436).

Definitions

Engagement, Retention, Adherence, Compliance, Completion, etc

We identified 2 main themes regarding the use of definitions in the literature sample. Our initial finding was that there is no clear agreement on the definition of engagement. This was likely because the literature in this space varies widely across questions, motivations, and perspectives. The second finding was that engagement was often captured by many different terms. In our sample, we saw terms such as *retention*, *adherence*, *compliance*, *completion*, and others sometimes used interchangeably. Despite this lack of clarity, we categorized our final sample into 3 distinct areas of engagement. Almost all studies (59/62, 95%) described or measured some form of engagement around the opening or using a specific app. Depending on the interface of the app being studied, this open

or use definition encompasses nearly any type of app interaction. In some cases, the number of app opens and duration of time spent were collected via a backend system, whereas for others, the data that users logged within the app were part of this definition. The 3 articles that did not fit our open or use category relied on self-reported use of the app or measuring the completion of intervention activities from the app.

Success

We asked about the extent to which the research was successful in maintaining participant engagement. Regardless of the term used for engagement or retention, we defined success as the percentage of participants with complete data from the initial sample after an intervention. We defined a *Success Code* variable with 3 categories based on information from the mean and SD. Percentages below the mean minus one SD were considered *not successful* and percentages above the mean plus one SD were considered *successful*. Everything else in between was considered *partially successful* (42/62, 68%). Only 19% (12/62) were considered not successful and 3% (2/62) could not be calculated because they relied on self-reported app use.

Simultaneously, we developed a binomial regression model to examine the factors that could affect retention. The outcome of our binomial regression model was the proportion of complete data from the final sample. The model was weighted on the sample size of the studies. Table 2 shows the odds ratio estimates and CIs from the binomial regression model. For any given participant, it is more likely that they will not be retained than they will be retained. Furthermore, participants with a clinical condition of interest were 4 times more likely to stay in the study than those who did not. Moreover, participants who were compensated were 10 times more likely to stay in the study than those who were not compensated.

Table 2. Results from the binomial regression model.

	Odds ratio (95% CI)
Intercept	0.09[a] (0.093-0.094)
Clinical	4.34[a] (4.16-4.52)
Compensation	10.32[a] (9.48-11.25)

[a]*P*<.001.

Qualitative Findings

Our qualitative findings represent the recurring themes around engagement listed in the findings, discussion, limitations, or conclusions sections of the articles. To be considered a stand-alone theme, the concept must have appeared in at least two independent studies in our sample.

Support Themes

We identified 3 major themes (ie, app affordances, successful recruitment, and low barriers to entry) that researchers mentioned that might have kept the participants engaged in their mHealth apps. Even if the article in the sample did not specifically use these supports, we noted where researchers recommended more work to address these supports in future research.

App Affordances

Affordances are "the quality or property of an object that defines its possible uses or makes clear how it can or should be used" [74]. In the technology space, this word is often used to describe the possibilities of specific actions that software or hardware allows. On the level of the app being studied (either compared with business-as-usual, another app, or something else altogether), there were several affordances that made research participants more engaged or more likely to stay engaged across the study span. One such factor was gamification. According to Fernandez et al [57], "Gamification or the use of game design elements (badges, leaderboards, rewards, and avatars) can help maintain user engagement." Very few studies have actually implemented gamification, but this theme was often mentioned as a possibility for future research to evaluate. Approximately one-quarter of our sample mentioned gamification as a future tool for promoting or sustaining engagement in a given mHealth app.

Although it was sometimes an area of interest in its own right, most articles mentioned some level of app reminders, feedback, or notifications that promoted engagement. Indeed, Bidargaddi et al [22] tested the effect of timing on weekends versus weekdays and found that users were most likely to engage with the app within 24 hours if prompted midday on the weekend. It is clear that reminders or other feedback through notifications was a supportive element for producing more engagement and less retention.

Approximately half of the articles mentioned some form of social support provided by coaches or peers within the app. Apps that included a coaching element, either from paraprofessionals, other participants, or the research study team, reported that this social support was critical for maintaining engagement throughout. One specific study by Mao et al [64] reported that 90% of participants who downloaded the app completed 4 months of coaching. This finding was likely because of a combination of participant-selected professional coaches who provided accountability and the social nature of the coaching relationship. In addition to social support, apps featuring tailored and personalized content were more likely to support engagement and adherence to the study.

Successful Recruitment

A total of 2 subthemes were drawn from the discussion of recruitment as support for engagement: recruiting highly motivated niche groups and providing some type of motivator in the form of either an incentive or a compensation. mHealth apps that were focused on a niche or highly motivated group of

users tended to be more successful in engaging participants over the course of the study. For example, mHealth apps created to support smoking cessation for adult smokers were more likely to be successful when participants were already highly motivated to stop smoking (eg, [18,40,72]). Most studies also mentioned either some form of compensation or other incentives or motivators that could engage more study participants for a longer period. More than half of the studies mentioned providing some type of compensation. Several articles mentioned that there was also a necessary balance needed to use compensation effectively. Providing *too little* incentive might make participants less compelled to continue in the research, but at the same time, providing *too much* incentive could also backfire by reducing their intrinsic motivation to continue. This balance continues to be important for researchers to consider moving forward.

Low Barriers to Entry

Related to both app affordances and recruitment strategies, another subtheme that emerged was apps with extremely low barriers to entry. This theme was best described by McConnell et al [7] in the MyHeart Counts Study. Their app was based on Apple's ResearchKit and enabled nearly 50,000 participants to register and provide consent for research. By launching a free app on smartphones, the authors stated "...the bar for entry to this study was much lower than that for equivalent studies performed using in-person visits. This lowering has the demonstrated advantage that many people consented..." Several other large-scale studies developed using ResearchKit had the advantage of recruiting and enrolling several thousands of participants [75]. This initial engagement was noted as a benefit, but as we learned later, such a low barrier to entry also often meant a low barrier to exit.

Barrier Themes

Researchers have also mentioned barriers that might diminish participant engagement. Here, we also noted barriers that were addressed in the discussion or limitations section of the articles, even though they were not actively described in the measures or results. These themes were described as (1) the lack of support codes; (2) low barriers to exit; (3) technical difficulties in using the app; and (4) somewhat counterintuitively, the usefulness of an app.

Lack of Support Features

Most barriers, either explicitly described or implied, were those that counteracted the support features. Articles routinely mentioned the lack of app affordances and recruitment success. Research involving apps without gamification, notifications of some sort, or support from peers or coaches was more likely to mention these as potential rationales for poor engagement and areas that could be improved in the future. A similar phenomenon was found in terms of recruitment strategies, where lack of compensation or having a niche group for the app were regularly noted as barriers to retention.

Low Barriers to Exit

In the same manner that large smartphone-based studies using the ResearchKit format provided a low barrier to entry, they also provided an equally low barrier to exit. For example, the MyHeart Counts Study further noted that when there is a low barrier to entry, there is a "notable disadvantage that those individuals are by definition less invested in the study and thus less likely to complete all portions" [7]. Almost all the apps available from ResearchKit in our sample represent the highest end of the sample size; however, none of the studies received even a partially successful code in our analysis.

Technical Difficulties

Articles that mentioned occasional glitches or *bugs* in the use of their apps were also likely to describe technical difficulties as a reason for lack of engagement. One study explicitly mentioned the use of the research support team to troubleshoot any technical difficulties for users [35], but most articles did not mention how they handled technology support requests. It is likely that some of the technical difficulties could have been on the app side, especially when the apps tested were in a pilot or beta form, but it is also possible that the participants had their own technical difficulties. None of the studies we evaluated performed any kind of pretest to measure participant comfort or familiarity with apps in general or apps similar to the one being studied. Generally, participants who were young adults or middle-aged were assumed to be good with technology overall. In addition, despite nearly a third of the articles mentioning usability and feasibility as a main investigation, only 5 studies mentioned participant results from the System Usability Scale [76], a standardized measure of usability frequently included in the human-computer interaction research space. Otherwise, usability and feasibility analyses were conducted on a study-by-study basis.

Usefulness of App

Although it may seem counterintuitive, apps that were extremely useful for participants were also some of the apps anecdotally deemed poor at engagement. For example, participants who successfully quit smoking while using a smoking cessation app generally had poor engagement in the long term. Indeed, if an app *works*, or achieves what it is meant to achieve, and does not offer some kind of regular check-in or maintenance program, it may be reasonable that participants taper the use of the app. In these cases, reduced engagement is a sign of success rather than failure and could actually be considered the goal of the app.

Discussion

Principal Findings

This study synthesizes the literature on mHealth apps and the engagement strategies. As mHealth apps continue to grow in popularity and research in this space follows that trend, researchers need to identify what made participants *stay* engaged in the app or studies with the app.

Our review found that most (48/62, 77%) studies were at least partially successful in maintaining participant engagement throughout. Many of these successes were because of the support features of the research or app and the lack of barriers to entry. We determined the categories of strategies that support or detract from engagement. We identified particular elements of the app (eg, feedback, appropriate reminders, and in-app support from peers or coaches) and strategies for research that promote

retention (eg, compensation and niche samples) as well as those that do not support retention (eg, lack of support features, technical difficulties, and usefulness of app). Research on the massive population-level ResearchKit apps appeared in both cases, using both successful and unsuccessful engagement and retention techniques. Although low barriers to initial entry could allow thousands of participants to be recruited, the same features also functioned as low barriers to exit. Recruiting a large number of participants is certainly beneficial, but that benefit may be substantially reduced if retention is poor. Future research should consider how to better balance these needs and incorporate factors such as clinical status, referral from providers, and compensation into recruitment plans for population-level apps.

This study used a binomial regression model to assess whether having a clinical condition of interest or receiving compensation affects retention rates. The empirical outcomes of the binomial regression model revealed that (1) any participant is more likely to not be retained than to be retained, (2) participants who have the same clinical condition targeted by the study are 4.33 times more likely to stay in the study than participants who do not have the same clinical condition targeted by the study, and (3) participants receiving compensation are 10.32 times more likely to stay in the study than participants who do not receive compensation. These findings, in line with previous research [10], demonstrate that retaining participants is a true challenge for studies using mHealth apps. Unlike that study [10], we were unable to incorporate clinician referral and age as part of our model because of inconsistent reporting in the articles. Although we planned to include other factors of interest, such as participant gender, income level, years of education, and smartphone platform type, the inconsistent reporting across studies made it challenging to accurately compare these variables. We also recognize that our definition of *success* relies on a normal distribution rather than some other indicator, which might be more appropriate for research with mobile apps that are still in their infancy. To summarize, scientists and researchers must consider different strategies to incentivize and encourage participant retention.

Of course, there is a balance when it comes to successful recruitment strategies, specifically compensation and niche groups. Strong participant engagement or retention may not accurately demonstrate the effectiveness of an app if the participants are overly compensated. Likewise, recruiting a niche group that is highly motivated to use a particular app presents a selection bias and leads to a lack of generalizability of the evaluation findings. Researchers and industry alike would do well to consider this balance when implementing studies using mHealth apps.

Limitations

Although we offer new insights into mHealth apps and participant engagement, this study has some limitations. First, as a systematic review, we were unable to make claims about all studies on apps. Owing to the file drawer phenomenon and our use of only peer-reviewed published articles, we do not report any studies that might have found null results, even though they might have described *different* interesting supports and barriers for engagement. Therefore, we encourage readers to refrain from generalizations about research on all mHealth apps. Second, we initially extracted information about the diversity of the sample; however, not all articles were clear about the diversity and the possible limitations of their own samples. Unfortunately, we were unable to describe these features in detail, as it is a critical area for more scholarship. Future research should consider the diversity in the demographics of published articles on mHealth apps and provide guidance about that.

Implications

We recommend that future mHealth apps consider potential support and barriers to participant engagement. Although the promise of moving health experiences onto the devices that people are currently using is great, many of the same barriers to participant engagement still exist and should be considered before moving research onto smartphone administration exclusively.

Conclusions

Retaining participants is a ubiquitous challenge for studies using mHealth apps. Despite the continued success of mHealth apps in the research sphere, there are many barriers to participant retention and long-term engagement. The insights from this review will help inform future studies about the potential different strategies and factors to consider and improve mHealth app engagement and retention.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study characteristics of the 62 articles.
[DOCX File , 29 KB - jmir_v24i4e35120_app1.docx]

References

1. Mobile fact sheet. Pew Research Center. 2019. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2022-04-07]
2. Grand View Research. 2021. URL: https://www.marketresearch.com/Grand-View-Research-v4060/mHealth-Size-Share-Trends-Component-14163316/ [accessed 2022-04-07]

3. Almalki M, Giannicchi A. Health apps for combating COVID-19: descriptive review and taxonomy. JMIR Mhealth Uhealth 2021 Mar 02;9(3):e24322 [FREE Full text] [doi: 10.2196/24322] [Medline: 33626017]
4. Chan YF, Bot BM, Zweig M, Tignor N, Ma W, Suver C, et al. The asthma mobile health study, smartphone data collected using ResearchKit. Sci Data 2018 May 22;5:180096 [FREE Full text] [doi: 10.1038/sdata.2018.96] [Medline: 29786695]
5. Chan YF, 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]
6. Doerr M, Maguire Truong A, Bot BM, Wilbanks J, Suver C, Mangravite LM. Formative evaluation of participant experience with mobile eConsent in the app-mediated Parkinson mPower study: a mixed methods study. JMIR Mhealth Uhealth 2017 Feb 16;5(2):e14 [FREE Full text] [doi: 10.2196/mhealth.6521] [Medline: 28209557]
7. McConnell MV, Shcherbina A, Pavlovic A, Homburger JR, Goldfeder RL, Waggot D, et al. Feasibility of obtaining measures of lifestyle from a smartphone app: the MyHeart counts cardiovascular health study. JAMA Cardiol 2017 Jan 01;2(1):67-76. [doi: 10.1001/jamacardio.2016.4395] [Medline: 27973671]
8. Hershman SG, Bot BM, Shcherbina A, Doerr M, Moayedi Y, Pavlovic A, et al. Physical activity, sleep and cardiovascular health data for 50,000 individuals from the MyHeart Counts Study. Sci Data 2019 Apr 11;6(1):24 [FREE Full text] [doi: 10.1038/s41597-019-0016-7] [Medline: 30975992]
9. Team Localytics. Mobile App Retention Rate: What's a Good Retention Rate? Upland Software. 2021. URL: https://uplandsoftware.com/localytics/resources/blog/mobile-apps-whats-a-good-retention-rate/ [accessed 2022-04-07]
10. 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 Feb;3:21 [FREE Full text] [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
11. Chu KH, Matheny SJ, Escobar-Viera CG, Wessel C, Notier AE, Davis EM. Smartphone health apps for tobacco Cessation: a systematic review. Addict Behav 2021 Jan;112:106616 [FREE Full text] [doi: 10.1016/j.addbeh.2020.106616] [Medline: 32932102]
12. Druce KL, Dixon WG, McBeth J. Maximizing engagement in mobile health studies: lessons learned and future directions. Rheum Dis Clin North Am 2019 May;45(2):159-172 [FREE Full text] [doi: 10.1016/j.rdc.2019.01.004] [Medline: 30952390]
13. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. J Affect Disord 2020 Feb 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
14. 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]
15. Mays N, Pope C, Popay J. Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field. J Health Serv Res Policy 2005 Jul;10 Suppl 1:6-20. [doi: 10.1258/1355819054308576] [Medline: 16053580]
16. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
17. Purkayastha S, Addepally SA, Bucher S. Engagement and usability of a cognitive behavioral therapy mobile app compared with web-based cognitive behavioral therapy among college students: randomized heuristic trial. JMIR Hum Factors 2020 Feb 03;7(1):e14146 [FREE Full text] [doi: 10.2196/14146] [Medline: 32012043]
18. Marler JD, Fujii CA, Utley DS, Tesfamariam LJ, Galanko JA, Patrick H. Initial assessment of a comprehensive digital smoking cessation program that incorporates a mobile app, breath sensor, and coaching: cohort study. JMIR Mhealth Uhealth 2019 Feb 04;7(2):e12609 [FREE Full text] [doi: 10.2196/12609] [Medline: 30670372]
19. Patrick H, Fujii CA, Glaser DB, Utley DS, Marler JD. A comprehensive digital program for smoking cessation: assessing feasibility in a single-group cohort study. JMIR Mhealth Uhealth 2018 Dec 18;6(12):e11708 [FREE Full text] [doi: 10.2196/11708] [Medline: 30563807]
20. Huberty J, Green J, Glissmann C, Larkey L, Puzia M, Lee C. Efficacy of the mindfulness meditation mobile app "Calm" to reduce stress among college students: randomized controlled trial. JMIR Mhealth Uhealth 2019 Jun 25;7(6):e14273 [FREE Full text] [doi: 10.2196/14273] [Medline: 31237569]
21. Sittig S, Wang J, Iyengar S, Myneni S, Franklin A. Incorporating behavioral trigger messages into a mobile health app for chronic disease management: randomized clinical feasibility trial in diabetes. JMIR Mhealth Uhealth 2020 Mar 16;8(3):e15927 [FREE Full text] [doi: 10.2196/15927] [Medline: 32175908]
22. 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]
23. Selter A, Tsangouri C, Ali SB, Freed D, Vatchinsky A, Kizer J, et al. An mHealth app for self-management of chronic lower back pain (Limbr): pilot study. JMIR Mhealth Uhealth 2018 Sep 17;6(9):e179 [FREE Full text] [doi: 10.2196/mhealth.8256] [Medline: 30224333]

24. Dahne J, Lejuez CW, Diaz VA, Player MS, Kustanowitz J, Felton JW, et al. Pilot randomized trial of a self-help behavioral activation mobile app for utilization in primary care. Behav Ther 2019 Jul;50(4):817-827 [FREE Full text] [doi: 10.1016/j.beth.2018.12.003] [Medline: 31208690]
25. Shah S, Kemp JM, Kvedar JC, Gracey LE. A feasibility study of the burden of disease of atopic dermatitis using a smartphone research application, myEczema. Int J Womens Dermatol 2020 Dec;6(5):424-428 [FREE Full text] [doi: 10.1016/j.ijwd.2020.08.001] [Medline: 33898712]
26. Shcherbina A, Hershman SG, Lazzeroni L, King AC, O'Sullivan JW, Hekler E, et al. The effect of digital physical activity interventions on daily step count: a randomised controlled crossover substudy of the MyHeart Counts Cardiovascular Health Study. Lancet Digit Health 2019 Nov;1(7):e344-e352 [FREE Full text] [doi: 10.1016/S2589-7500(19)30129-3] [Medline: 33323209]
27. Iacoviello BM, Steinerman JR, Klein DB, Silver TL, Berger AG, Luo SX, et al. Clickotine, a personalized smartphone app for smoking cessation: initial evaluation. JMIR Mhealth Uhealth 2017 Apr 25;5(4):e56 [FREE Full text] [doi: 10.2196/mhealth.7226] [Medline: 28442453]
28. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The use and effectiveness of mobile apps for depression: results from a fully remote clinical trial. J Med Internet Res 2016 Dec 20;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]
29. Bricker JB, Watson NL, Heffner JL, Sullivan B, Mull K, Kwon D, et al. A smartphone app designed to help cancer patients stop smoking: results from a pilot randomized trial on feasibility, acceptability, and effectiveness. JMIR Form Res 2020 Jan 17;4(1):e16652 [FREE Full text] [doi: 10.2196/16652] [Medline: 31951215]
30. Ben-Zeev D, Scherer EA, Gottlieb JD, Rotondi AJ, Brunette MF, Achtyes ED, et al. mHealth for schizophrenia: patient engagement with a mobile phone intervention following hospital discharge. JMIR Ment Health 2016 Jul 27;3(3):e34 [FREE Full text] [doi: 10.2196/mental.6348] [Medline: 27465803]
31. Serrano KJ, Coa KI, Yu M, Wolff-Hughes DL, Atienza AA. Characterizing user engagement with health app data: a data mining approach. Transl Behav Med 2017 Jun;7(2):277-285 [FREE Full text] [doi: 10.1007/s13142-017-0508-y] [Medline: 28616846]
32. Zhang R, Nicholas J, Knapp AA, Graham AK, Gray E, Kwasny MJ, et al. Clinically meaningful use of mental health apps and its effects on depression: mixed methods study. J Med Internet Res 2019 Dec 20;21(12):e15644 [FREE Full text] [doi: 10.2196/15644] [Medline: 31859682]
33. Moberg C, Niles A, Beermann D. Guided self-help works: randomized waitlist controlled trial of Pacifica, a mobile app integrating cognitive behavioral therapy and mindfulness for stress, anxiety, and depression. J Med Internet Res 2019 Jun 08;21(6):e12556 [FREE Full text] [doi: 10.2196/12556] [Medline: 31199319]
34. 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]
35. Ross EL, Jamison RN, Nicholls L, Perry BM, Nolen KD. Clinical integration of a smartphone app for patients with chronic pain: retrospective analysis of predictors of benefits and patient engagement between clinic visits. J Med Internet Res 2020 Apr 16;22(4):e16939 [FREE Full text] [doi: 10.2196/16939] [Medline: 32297871]
36. Scott AR, Alore EA, Naik AD, Berger DH, Suliburk JW. Mixed-methods analysis of factors impacting use of a postoperative mHealth app. JMIR Mhealth Uhealth 2017 Feb 08;5(2):e11 [FREE Full text] [doi: 10.2196/mhealth.6728] [Medline: 28179215]
37. Gordon JS, Armin J, D Hingle M, Giacobbi Jr P, Cunningham JK, Johnson T, et al. Development and evaluation of the See Me Smoke-Free multi-behavioral mHealth app for women smokers. Transl Behav Med 2017 Jun;7(2):172-184 [FREE Full text] [doi: 10.1007/s13142-017-0463-7] [Medline: 28155107]
38. Berman MA, Guthrie NL, Edwards KL, Appelbaum KJ, Njike VY, Eisenberg DM, et al. Change in glycemic control with use of a digital therapeutic in adults with type 2 diabetes: cohort study. JMIR Diabetes 2018 Feb 14;3(1):e4 [FREE Full text] [doi: 10.2196/diabetes.9591] [Medline: 30291074]
39. Michaelides A, Major J, Pienkosz Jr E, Wood M, Kim Y, Toro-Ramos T. Usefulness of a novel mobile diabetes prevention program delivery platform with human coaching: 65-week observational follow-up. JMIR Mhealth Uhealth 2018 May 03;6(5):e93 [FREE Full text] [doi: 10.2196/mhealth.9161] [Medline: 29724709]
40. Bruno M, Wright M, Baker CL, Emir B, Carda E, Clausen M, et al. Mobile app usage patterns of patients prescribed a smoking cessation medicine: prospective observational study. JMIR Mhealth Uhealth 2018 Apr 17;6(4):e97 [FREE Full text] [doi: 10.2196/mhealth.9115] [Medline: 29666043]
41. Petersen CL, Weeks WB, Norin O, Weinstein JN. Development and implementation of a person-centered, technology-enhanced care model for managing chronic conditions: cohort study. JMIR Mhealth Uhealth 2019 Mar 20;7(3):e11082 [FREE Full text] [doi: 10.2196/11082] [Medline: 30892274]
42. Bush J, Barlow DE, Echols J, Wilkerson J, Bellevin K. Impact of a mobile health application on user engagement and pregnancy outcomes among Wyoming Medicaid members. Telemed J E Health 2017 Nov;23(11):891-898 [FREE Full text] [doi: 10.1089/tmj.2016.0242] [Medline: 28481167]

43. Ryan P, Brown RL, Csuka ME, Papanek P. Efficacy of osteoporosis prevention smartphone app. Nurs Res 2020;69(1):31-41 [FREE Full text] [doi: 10.1097/NNR.0000000000000392] [Medline: 31568199]
44. Zeng EY, Heffner JL, Copeland WK, Mull KE, Bricker JB. Get with the program: adherence to a smartphone app for smoking cessation. Addict Behav 2016 Dec;63:120-124 [FREE Full text] [doi: 10.1016/j.addbeh.2016.07.007] [Medline: 27454354]
45. Sridharan V, Shoda Y, Heffner J, Bricker J. A pilot randomized controlled trial of a web-based growth mindset intervention to enhance the effectiveness of a smartphone app for smoking cessation. JMIR Mhealth Uhealth 2019 Jul 09;7(7):e14602 [FREE Full text] [doi: 10.2196/14602] [Medline: 31290404]
46. Hales S, Turner-McGrievy GM, Wilcox S, Davis RE, Fahim A, Huhns M, et al. Trading pounds for points: engagement and weight loss in a mobile health intervention. Digit Health 2017 Apr 24;3:2055207617702252 [FREE Full text] [doi: 10.1177/2055207617702252] [Medline: 29942590]
47. Mundi MS, Lorentz PA, Grothe K, Kellogg TA, Collazo-Clavell ML. Feasibility of smartphone-based education modules and ecological momentary assessment/intervention in pre-bariatric surgery patients. Obes Surg 2015 Oct;25(10):1875-1881. [doi: 10.1007/s11695-015-1617-7] [Medline: 25702141]
48. Betthauser LM, Stearns-Yoder KA, McGarity S, Smith V, Place S, Brenner LA. Mobile app for mental health monitoring and clinical outreach in veterans: mixed methods feasibility and acceptability study. J Med Internet Res 2020 Aug 11;22(8):e15506 [FREE Full text] [doi: 10.2196/15506] [Medline: 32779572]
49. Escobar-Viera C, Zhou Z, Morano JP, Lucero R, Lieb S, McIntosh S, et al. The Florida mobile health adherence project for people living with HIV (FL-mAPP): longitudinal assessment of feasibility, acceptability, and clinical outcomes. JMIR Mhealth Uhealth 2020 Jan 08;8(1):e14557 [FREE Full text] [doi: 10.2196/14557] [Medline: 31913127]
50. Patel ML, Hopkins CM, Bennett GG. Early weight loss in a standalone mHealth intervention predicting treatment success. Obes Sci Pract 2019 Jun;5(3):231-237 [FREE Full text] [doi: 10.1002/osp4.329] [Medline: 31275596]
51. Forman EM, Goldstein SP, Zhang F, Evans BC, Manasse SM, Butryn ML, et al. OnTrack: development and feasibility of a smartphone app designed to predict and prevent dietary lapses. Transl Behav Med 2019 Mar 01;9(2):236-245 [FREE Full text] [doi: 10.1093/tbm/iby016] [Medline: 29617911]
52. Du H, Venkatakrishnan A, Youngblood GM, Ram A, Pirolli P. A group-based mobile application to increase adherence in exercise and nutrition programs: a factorial design feasibility study. JMIR Mhealth Uhealth 2016 Jan 15;4(1):e4 [FREE Full text] [doi: 10.2196/mhealth.4900] [Medline: 26772910]
53. Krzyzanowski MC, Kizakevich PN, Duren-Winfield V, Eckhoff R, Hampton J, Blackman Carr LT, et al. Rams Have Heart, a mobile app tracking activity and fruit and vegetable consumption to support the cardiovascular health of college students: development and usability study. JMIR Mhealth Uhealth 2020 Aug 05;8(8):e15156 [FREE Full text] [doi: 10.2196/15156] [Medline: 32755883]
54. Michaelides A, Raby C, Wood M, Farr K, Toro-Ramos T. Weight loss efficacy of a novel mobile Diabetes Prevention Program delivery platform with human coaching. BMJ Open Diabetes Res Care 2016 Sep 5;4(1):e000264 [FREE Full text] [doi: 10.1136/bmjdrc-2016-000264] [Medline: 27651911]
55. Kelechi TJ, Prentice MA, Mueller M, Madisetti M, Vertegel A. A lower leg physical activity intervention for individuals with chronic venous leg ulcers: randomized controlled trial. JMIR Mhealth Uhealth 2020 May 15;8(5):e15015 [FREE Full text] [doi: 10.2196/15015] [Medline: 32412419]
56. 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]
57. Fernandez MP, Bron GM, Kache PA, Larson SR, Maus A, Gustafson Jr D, et al. Usability and feasibility of a smartphone app to assess human behavioral factors associated with tick exposure (The Tick App): quantitative and qualitative study. JMIR Mhealth Uhealth 2019 Oct 24;7(10):e14769 [FREE Full text] [doi: 10.2196/14769] [Medline: 31651409]
58. Hébert ET, Ra CK, Alexander AC, Helt A, Moisiuc R, Kendzor DE, et al. A mobile just-in-time adaptive intervention for smoking cessation: pilot randomized controlled trial. J Med Internet Res 2020 Mar 09;22(3):e16907 [FREE Full text] [doi: 10.2196/16907] [Medline: 32149716]
59. Huberty J, Vranceanu AM, Carney C, Breus M, Gordon M, Puzia ME. Characteristics and usage patterns among 12,151 paid subscribers of the calm meditation app: cross-sectional survey. JMIR Mhealth Uhealth 2019 Nov 03;7(11):e15648 [FREE Full text] [doi: 10.2196/15648] [Medline: 31682582]
60. Chow PI, Showalter SL, Gerber M, Kennedy EM, Brenin D, Mohr DC, et al. Use of mental health apps by patients with breast cancer in the United States: pilot pre-post study. JMIR Cancer 2020 Apr 15;6(1):e16476 [FREE Full text] [doi: 10.2196/16476] [Medline: 32293570]
61. 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]
62. Tran C, Dicker A, Leiby B, Gressen E, Williams N, Jim H. Utilizing digital health to collect electronic patient-reported outcomes in prostate cancer: single-arm pilot trial. J Med Internet Res 2020 Mar 25;22(3):e12689 [FREE Full text] [doi: 10.2196/12689] [Medline: 32209536]

63. Ifejika NL, Bhadane M, Cai CC, Noser EA, Grotta JC, Savitz SI. Use of a smartphone-based mobile app for weight management in obese minority stroke survivors: pilot randomized controlled trial with open blinded end point. JMIR Mhealth Uhealth 2020 Apr 22;8(4):e17816 [FREE Full text] [doi: 10.2196/17816] [Medline: 32319963]
64. Mao AY, Chen C, Magana C, Caballero Barajas K, Olayiwola JN. A mobile phone-based health coaching intervention for weight loss and blood pressure reduction in a national payer population: a retrospective study. JMIR Mhealth Uhealth 2017 Jun 08;5(6):e80 [FREE Full text] [doi: 10.2196/mhealth.7591] [Medline: 28596147]
65. Rudin RS, Fanta CH, Qureshi N, Duffy E, Edelen MO, Dalal AK, et al. A clinically integrated mHealth app and practice model for collecting patient-reported outcomes between visits for asthma patients: implementation and feasibility. Appl Clin Inform 2019 Oct;10(5):783-793 [FREE Full text] [doi: 10.1055/s-0039-1697597] [Medline: 31618782]
66. Johnston DC, Mathews WD, Maus A, Gustafson DH. Using smartphones to improve treatment retention among impoverished substance-using Appalachian women: a naturalistic study. Subst Abuse 2019 Jul 8;13:1178221819861377 [FREE Full text] [doi: 10.1177/1178221819861377] [Medline: 31312084]
67. Pagoto S, Tulu B, Agu E, Waring ME, Oleski JL, Jake-Schoffman DE. Using the habit app for weight loss problem solving: development and feasibility study. JMIR Mhealth Uhealth 2018 Jun 20;6(6):e145 [FREE Full text] [doi: 10.2196/mhealth.9801] [Medline: 29925496]
68. Everett E, Kane B, Yoo A, Dobs A, Mathioudakis N. A novel approach for fully automated, personalized health coaching for adults with prediabetes: pilot clinical trial. J Med Internet Res 2018 Feb 27;20(2):e72 [FREE Full text] [doi: 10.2196/jmir.9723] [Medline: 29487046]
69. 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]
70. Burke LE, Zheng Y, Ma Q, Mancino J, Loar I, Music E, et al. The SMARTER pilot study: testing feasibility of real-time feedback for dietary self-monitoring. Prev Med Rep 2017 Mar 31;6:278-285 [FREE Full text] [doi: 10.1016/j.pmedr.2017.03.017] [Medline: 28409090]
71. Bricker JB, Copeland W, Mull KE, Zeng EY, Watson NL, Akioka KJ, et al. Drug Alcohol Depend 2017 Jan 01;170:37-42 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.10.029] [Medline: 27870987]
72. Hoeppner BB, Hoeppner SS, Carlon HA, Perez GK, Helmuth E, Kahler CW, et al. Leveraging positive psychology to support smoking cessation in nondaily smokers using a smartphone app: feasibility and acceptability study. JMIR Mhealth Uhealth 2019 Jul 03;7(7):e13436 [FREE Full text] [doi: 10.2196/13436] [Medline: 31271147]
73. Dillingham R, Ingersoll K, Flickinger TE, Waldman AL, Grabowski M, Laurence C, et al. PositiveLinks: a mobile health intervention for retention in HIV care and clinical outcomes with 12-month follow-up. AIDS Patient Care STDS 2018 Jun;32(6):241-250 [FREE Full text] [doi: 10.1089/apc.2017.0303] [Medline: 29851504]
74. Affordance. Merriam-Webster. URL: https://www.merriam-webster.com/dictionary/affordance [accessed 2022-04-07]
75. Husain I. ResearchKit app shows high levels of enrollment and engagement on iPhone. iMedical Apps. 2015 Jun 17. URL: https://www.imedicalapps.com/2015/06/researchkit-engagement-iphone/ [accessed 2022-04-07]
76. Brooke J. SUS: a 'quick and dirty' usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability evaluation in industry. London, UK: CRC Press; 1996:189-204.

Abbreviations

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

Edited by A Mavragani; submitted 22.11.21; peer-reviewed by A Mao, R Marshall; comments to author 20.01.22; revised version received 16.03.22; accepted 17.03.22; published 26.04.22.

Please cite as:
Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC
Challenges in Participant Engagement and Retention Using Mobile Health Apps: Literature Review
J Med Internet Res 2022;24(4):e35120
URL: https://www.jmir.org/2022/4/e35120
doi:10.2196/35120
PMID:35471414

Review

6G and Artificial Intelligence Technologies for Dementia Care: Literature Review and Practical Analysis

Zhaohui Su[1], PhD; Barry L Bentley[2], PhD; Dean McDonnell[3], PhD; Junaid Ahmad[4], PhD; Jiguang He[5], PhD; Feng Shi[6], PhD; Kazuaki Takeuchi[7,8], MSc; Ali Cheshmehzangi[9,10], PhD; Claudimar Pereira da Veiga[11], PhD

[1]School of Public Health, Southeast University, Nanjing, China
[2]Cardiff School of Technologies, Cardiff Metropolitan University, Cardiff, United Kingdom
[3]Department of Humanities, Institute of Technology Carlow, Carlow, Ireland
[4]Prime Institute of Public Health, Peshawar Medical College, Peshawar, Pakistan
[5]Centre for Wireless Communications, University of Oulu, Oulu, Finland
[6]Department of Research and Development, Shanghai United Imaging Intelligence, Shanghai, China
[7]Ory Laboratory Inc, Tokyo, Japan
[8]Kanagawa Institute of Technology, Kanagawa, Japan
[9]Department of Architecture and Built Environment, University of Nottingham Ningbo China, Ningbo, China
[10]Network for Education and Research on Peace and Sustainability, Hiroshima University, Hiroshima, Japan
[11]Fundação Dom Cabral, Nova Lima, Brazil

Corresponding Author:
Claudimar Pereira da Veiga, PhD
Fundação Dom Cabral
632 Prefeito Lothário Meissner Ave
Av Princesa Diana, 760 Alphaville, Lagoa dos Ingleses
Nova Lima, 34018-006
Brazil
Phone: 55 41 3360 4366
Email: claudimar.veiga@gmail.com

Abstract

Background: The dementia epidemic is progressing fast. As the world's older population keeps skyrocketing, the traditional incompetent, time-consuming, and laborious interventions are becoming increasingly insufficient to address dementia patients' health care needs. This is particularly true amid COVID-19. Instead, efficient, cost-effective, and technology-based strategies, such as sixth-generation communication solutions (6G) and artificial intelligence (AI)-empowered health solutions, might be the key to successfully managing the dementia epidemic until a cure becomes available. However, while 6G and AI technologies hold great promise, no research has examined how 6G and AI applications can effectively and efficiently address dementia patients' health care needs and improve their quality of life.

Objective: This study aims to investigate ways in which 6G and AI technologies could elevate dementia care to address this study gap.

Methods: A literature review was conducted in databases such as PubMed, Scopus, and PsycINFO. The search focused on three themes: dementia, 6G, and AI technologies. The initial search was conducted on April 25, 2021, complemented by relevant articles identified via a follow-up search on November 11, 2021, and Google Scholar alerts.

Results: The findings of the study were analyzed in terms of the interplay between people with dementia's unique health challenges and the promising capabilities of health technologies, with in-depth and comprehensive analyses of advanced technology-based solutions that could address key dementia care needs, ranging from impairments in memory (eg, Egocentric Live 4D Perception), speech (eg, Project Relate), motor (eg, Avatar Robot Café), cognitive (eg, Affectiva), to social interactions (eg, social robots).

Conclusions: To live is to grow old. Yet dementia is neither a proper way to live nor a natural aging process. By identifying advanced health solutions powered by 6G and AI opportunities, our study sheds light on the imperative of leveraging the potential of advanced technologies to elevate dementia patients' will to live, enrich their daily activities, and help them engage in societies across shapes and forms.

(J Med Internet Res 2022;24(4):e30503) doi:10.2196/30503

KEYWORDS

COVID-19; 6G; digital health; artificial intelligence; dementia; first-perspective health solutions

Introduction

The dementia epidemic is prevalent and pernicious [1], and it is engulfing the world at an alarming speed [2]. Approximately every 3 seconds, a new dementia case occurs, which translates into 10 million new annual cases added to the ever-increasing pool of dementia patients worldwide [3]. Globally, it is estimated that the number of people living with dementia will rise from 50 million in 2018 to 152 million in 2050 [2]. Dementia is not one disease but rather a family of health conditions that affect "memory and other cognitive abilities and behavior that interfere significantly with a person's ability to maintain their activities of daily living" [4]. While Alzheimer's disease is the most common type of dementia—accounts for 60% to 80% of all dementia cases, the term dementia could refer to a wide range of unique cognitive impairments rooted in diverse risk factors, ranging from Alzheimer's disease, frontotemporal dementia, Lewy body dementia, vascular dementia, to mixed dementia [4]. While impacts such as adverse drug effects could also cause transitory or reversal dementia symptoms [5], unfortunately, due to limitations to developments in science and technology, the majority of dementia cases are difficult or impossible to reverse.

To further compound the matter, available pharmaceutical interventions that could effectively cure or curb dementia range from scarce to nonexistent, a situation which has been further plagued by controversies, rather than promoted by consensus, attracted by high-profile treatments such as Biogen's aducanumab [6]. Overall, the increasing prevalence of dementia and the sobering lack of effective care could further aggravate the toll of the disease on lives, livelihoods, and economics. Take the more tangible economic consequences, for instance. It is estimated that the global economic cost of dementia has recently reached approximately US $1 trillion per annum [2] and is expected to inch towards US $9.12 trillion by 2050 [7]. What is too complex to materialize might be the most daunting task society has to shoulder—addressing the bevy of dementia needs and wants that are rooted in the kaleidoscopic range of symptoms that could manifest across the disease continuum [4].

The debilitating nature of the disease, particularly its corrosive impacts on patients' cognitive and functional capabilities, dictates that dementia patients often have to depend on others for care and support [8]. Two main sources of care dementia patients rely on are professional health care providers (eg, doctors, nurses, and other formal caregivers) and informal caregivers (eg, family, friends, and other close social ties) [9]. Evidence suggests that, depending on contextual factors such as family values and cultural norms, family care could constitute from 65% to 96% of all health care services dementia patients receive [9-11]. While the existing laborious and caregiver-reliant care paradigm serves a purpose and holds well-deserved and hard-earned merits, it also hinders dementia care development—patients' health and wellbeing, let alone access to care, are often contingent upon their informal caregivers' could-be problematic abilities and ever-changing willingness to provide care [12-15].

Take caregivers' dementia awareness and knowledge, for instance. In South Africa, for example, it is common for dementia patients, especially those of the female gender, to be considered as witches, even by health care professionals and patients' family caregivers—as a result, rather than receiving quality care and pre-eminent support, these patients are often bullied, beaten, or burned many thanks to their caregivers' exceedingly invalid but deeply ingrained dementia beliefs [13]. Furthermore, for some dementia caregivers, the seemingly never-ending, ever-worsening, and extremely-demanding duties and responsibilities associated with dementia care, particularly when compounded by (1) their lack of awareness, knowledge, and training, (2) competing interests and engagements, as well as (3) potential learned helplessness due to the continued absence of a tangible dementia cure, might be enough to trick or trigger them into committing inexcusable, yet possibly inescapable, neglect and abuse of dementia patients [16-18].

These insights combined underscore the need for innovative approaches to address challenges patients, caregivers, health care professionals, as well as society at large face regarding dementia care. Overall, while an effective pharmaceutical solution to dementia or a magic bullet for fixing the patient-caregiver relationship might be difficult to secure, what can be relatively easy to change in order to improve the effectiveness and efficiency of dementia care and management is how current dementia services are designed and delivered [19-23]. In other words, the existing dementia caregiving model could possibly benefit from a paradigm shift. Rather than solely relying on laborious, time-consuming, and human-dependent solutions that suffer from great variability in terms of care quality [16-18]), efficient, cost-effective, and technology-based interventions could be the much-needed solution to tame the dementia epidemic, interventions such as those that are powered by advanced technological solutions such as the sixth-generation wireless technologies (6G) and artificial intelligence (AI) [24-26]. However, while 6G and AI technologies hold great promise, no research has examined how 6G and AI applications can effectively and efficiently address dementia patients' health care needs and improve their quality of life. Thus, to bridge the research gap, this study aims to investigate ways in which 6G and AI technologies could elevate dementia care. A detailed research flow can be found in Figure 1.

Figure 1. A schematic representation of the study framework.

Methods

Overview

To address the research objective, a literature review was conducted in databases such as PubMed, Scopus, and PsycINFO. The search was focused on three themes: dementia, 6G, and AI technologies. An example search string applied in PubMed could be found in Textbox 1. The initial search was conducted on April 25, 2021, focused on literature published in the past five years at the time. A subsequent search was conducted on November 21, 2021, to ensure all relevant insights were included in the review. Furthermore, Google Scholar Alerts focusing on dementia, 6G, and AI technologies were set up to make sure the review could incorporate the most recent evidence. In addition, we also included up-to-date, validated news reports in the review to ensure industry insights that are most relevant to the current study are included.

Textbox 1. Themes and search terms adopted.

Dementia: dementia*[MeSH] OR dementia*[TIAB] OR Alzheimer*[MeSH] OR Alzheimer*[TIAB] OR "cognitive decline*"[MeSH] OR "cognitive decline*"[TIAB]

AI: "artificial intelligence"[MeSH] OR "artificial intelligence"[TIAB] OR "machine learning"[MeSH] OR "machine learning"[TIAB] OR "deep learning"[MeSH] OR "deep learning"[TIAB]

6G: "sixth-generation communication*"[MeSH] OR "sixth-generation communication*"[TIAB] OR "sixth-generation network*"[MeSH] OR "sixth-generation network*"[TIAB] OR "sixth-generation technolog*"[MeSH] OR "sixth-generation technolog*"[TIAB] OR "sixth-generation cellular" [MeSH] OR "sixth-generation cellular" [TIAB] OR "6G communication*"[MeSH] OR "6G communication*"[TIAB] OR "6G network*"[MeSH] OR "6G network*"[TIAB] OR "6G technolog*"[MeSH] OR "6G technolog*"[TIAB] OR "6G wireless"[MeSH] OR "6G wireless"[TIAB] OR "6G cellular"[MeSH] OR "6G cellular"[TIAB]

Eligibility Criteria

The inclusion criteria adopted to screen relevant papers are listed in Textbox 2. Overall, we excluded articles if they (1) were not published in English, (2) did not focus on dementia patients (eg, dementia caregivers), (3) did not focus on either AI or 6G technologies, or (4) did not report empirical findings (eg, protocol studies), (5) did not focus on the noninvasive application of 6G or AI technologies (eg, collection of biomarkers such as blood), and (6) did not offer detailed information on the utilization of 6G or AI in the context of dementia care and management.

Textbox 2. Data type and study inclusion criteria.

Language: English.

Study context: The use and application of 6G or AI in the context of dementia care and management.

Technology type: AI and 6G technologies.

Use of technology: Nonintrusive application of AI and 6G techniques or technologies.

Study design: Research that reports empirical findings.

Review and Analysis Structure

Overall, as posited by the renowned social scientists Urie Bronfenbrenner [27], the quantity and quality of dementia care could be influenced by a wide array of factors, rooted in all levels of society—from the individual level, the interpersonal level, the community level, the social level, to the policy level. In light of the scope of the study, we only focused on dementia care accessed on the individual level and from the patients' perspective. Furthermore, in order to comprehensively answer our overarching research aim—to understand ways in which 6G and AI technologies could elevate dementia care, two foundational research questions were identified and investigated: (1) what are the unique characteristics of dementia care, and (2) what 6G and AI-powered technologies could address the needs and wants of dementia patients. For ease of understanding and consistency in presentation, we organized the review and the subsequent analysis in accordance with these two research questions. A schematic representation of the review process can be found in Figure 1.

Results

A total of 32 peer-reviewed articles were included in the final review. In addition, up-to-date and vetted news reports were examined to further enhance the rigor of the subsequent analysis. The overarching aim of this study is to investigate ways in which 6G and AI technologies could elevate dementia care. To effectively answer this research question, we first need to identify the unique characteristics of dementia care and then the specific 6G and AI-powered technologies that could address the logistics of dementia care. Subsequently, the results are divided into two sections, with the first one focusing on the unique characteristics of dementia care and the second section centering on specific ways in which 6G and AI technologies could be best leveraged to mitigate the health challenges dementia patients face. While complex in nature, drawing insights from the literature [28-34], dementia could be roughly categorized into three stages, the early, middle, and the late stage, each with unique disease manifestations, particularly in terms of patients' cognitive and physical abilities (Table 1).

Table 1. Dementia care by stage.

Stage	Key characteristic	Care needed
Early stage	The symptoms are often overlooked due to the gradual onset of the disease	• Regular forgetfulness • Often lose track of time • Frequently get lost in familiar places
Middle stage	The symptoms become clearer and more manifested over time.	• Have difficulties in remembering recent events and acquittances' names • Confusion about time and space • Incapable of communicating coherently • Need assistance in self-care • Behavioral irregularities like wandering and repeated questioning
Late stage	The symptoms become increasingly evident and debilitating.	• Incognizance of time and space • Unable to recognize relatives and friends • Dependent on others for self-care • Incapable to walk • Behavioral irregularities like aggression and violence

Subsequently, the list of 6G and AI-powered technologies that could address the unique needs and preferences of people with dementia can be found in Table 2. In the following section, we will contextualize the findings of the study by delineating and dissecting the specific ways in which 6G and AI-powered technologies could be utilized to address people with dementia's unique care needs and preferences.

Table 2. Key dementia care needs and advanced technology-based solutions

Dementia care needs and solution		Detail
Memory impairment		
	RFusion [35]	• A robotic arm that could help dementia patients find even deeply hidden items based on camera and antenna data analyzed by advanced AI algorithms.
	Egocentric Live 4D Perception [36]	• An AI-powered project that could enable machines like virtual reality headsets to better help people like dementia patients to better navigate daily activities, ranging from finding lost items, limiting accidental over-medications, to enabling social interactions.
Speech impairment	Project Relate [37]	• An AI-powered communication tool built by Google that aims to help people with speech impairments communicate smoothly via Google Assistant.
Motor impairment		
	Project Activate [38]	• An AI-powered algorithm that allows people with speech and motor impairments to use facial expressions as smartphone commands.
	Avatar Robot Café [39]	• A robot and AI-powered system that allows people who are house-bound or bed-ridden, such as dementia patients, to engage in society as robot pilots—work remotely in the form of physical robot servers via virtually controlling these robots using an AI-powered system at home or even in bed.
Cognitive and motor impairments	Affectiva [40]	• An AI system that could recognize and analyze car drivers' emotional and cognitive states, such as distraction, fatigue, and heatstroke, information which can then be used to send alerts to the drivers to prevent potential accidents.

Social connections	PARO [41-46] and other social robots [47-49]	• A sensor-based therapeutic robot that could improve people with dementia's mood, social interaction, and wellbeing, which could be further improved via 6G and AI technologies:
		• Connect with more advanced AI health surveillance systems.
		  • Computer vision for assistive medical diagnosis based on facial images [50], which could further facilitate personalized care design and development.
		  • Video-based vital signs monitoring (eg, Oxehealth [51]).
		  • Brain-machine interface devices, like Neurable [52], to gain insights into patients' focus and productivity using headphones outfitted with EEG sensors.
		• Enable assistive robots with more competent health monitoring and managing functions.
		• Connect patients with their loved ones remotely "through" robots (eg, telepresence robots, which can transit voices, mimics, and head motions [53]).
		• Transform assistive robots into multi-functional care assistants (eg, perform memory evaluation test [54]; provide dementia patients with a wider range of services, from fetch and carry, fall detection, to transfer patients from floor to chair/bed [55,56]).

Discussion

Principal Findings

Our study aimed to investigate ways in which 6G and AI technologies could elevate dementia care. This is the first study that examined how society at large could better meet people with dementia's care needs and wants with the aid of advanced technologies such as 6G and AI. From a technological perspective, by pointing out areas where 6G and AI could help further leverage dementia care, the study shed light on the importance of ingenuity and interoperability across the older (eg, 4G and 5G) and the newer networking platforms (eg, 6G) and analytical capabilities (eg, AI algorithms). From a care perspective, by detailing the promise and prowess of advanced technologies like 6G and AI in leveraging dementia care, the current study highlighted the importance of developing people-centric health solutions that could benefit the betterment of society and humanity, above and beyond the advancement of niche technologies. Overall, two fundamental research questions were answered to sufficiently address the research objective: (1) what are the unique characteristics of dementia care, and (2) what 6G and AI-powered technologies could address the unique needs and wants of dementia patients.

Unique Characteristics of Dementia Care

While much remains to be known about dementia, meaningful understandings about people with dementia are available in the literature. Overall, a bevy of definitions of dementia was present in the literature, ranging from the ones delineated by national or international organizations (eg, World Health Organization, the United Kingdom's National Health Service, the US Centers for Diseases Control and Prevention, the Government of Ireland, etc) [4,5,28-30], to conceptualizations developed to echo the unique characteristics of the research contexts [8,31-34]. Noticeably, almost all reviewed conceptualizations of dementia share the following similarities: (1) dementia is an abnormal part of aging that represents a diverse range of health conditions

that are hallmarked by the gradual decline of (2) cognitive and (3) physical functionalities. In other words, though dementia is common among older people, it is not a condition of aging—young people can develop dementia, and some older people do not have dementia [5].

Furthermore, rather than one disease, dementia represents a family of cognitive impairments, which includes conditions such as Alzheimer's Disease, vascular dementia, Lewy body dementia, and mixed dementia, often caused by varied disease manifestations [28-30]. To further complicate the situation, people can have multiple dementias at the same time or across the dementia continuum, conditions which are often difficult to detect or diagnose due to the similarities between types of dementia and limitations in medical sciences [28-30]. For instance, it is not uncommon for people with Alzheimer's Disease to simultaneously have vascular dementia (linked to issues with blood flow to the brain), especially among older people.

Subsequently, from the perspective of dementia care, the unique characteristics of dementia symptoms could be further classified as disease manifestations that are related to cognitive impairments and the ones rooted in the decline of physical functions. Overall regardless of disease stage, from the patient's perspective, across the dementia continuum, changes they experience regarding their dementia condition are often shown as fluctuations in cognitive and/or physical capabilities. In other words, people with dementia who experience optimal cognitive and physical abilities may have little to no need to rely on other people or advanced technologies to assist with their activities, whereas for individuals who experience poor cognitive and physical functionalities (eg, have difficulties in remembering, thinking, reasoning, walking, or moving), exterior help like advanced technologies are have-to-have enablers of their self-care needs and wants. A detailed illustration of the interplay between people with dementia's abilities and their need for health technologies can be found in Table 3.

Table 3. The interplay between people with dementia's abilities and their need for health technologies.

	High cognitive abilities	Low cognitive abilities
High physical abilities	• Little to no dependence on assistive technologies. • Health technologies are nice-to-have add-ons to daily activities.	• Extremely high dependence on assistive technologies. • Health technologies are have-to-have directors of daily activities.
Low physical abilities	• High dependence on assistive technologies. • Health technologies are have-to-have assistants of daily activities.	• Extremely high dependence on assistive technologies. • Health technologies are have-to-have enablers of daily activities.

6G and AI-Powered Technologies for Dementia Care

Compared to previous networking technologies such as 5G, 6G is considerably advanced, ranging from high-capacity connectivity to powerful real-time data analytics [57-64]. Overall, the relationship between 6G and AI will be that of symbiotic [65]—6G will lay the groundwork needed for the ecosystem to exist, whereas AI will make the 6G-enabled infrastructure produce meaningful health solutions for patients in a timely fashion, which in turn, facilitate the intelligentization of the hyper-connected community of solutions for dementia care. It is worth noting that though not all technical details of 6G have been fully explored, available research on its advantages in the following aspects renders investigations into 6G and AI-empowered dementia care justifiably relevant and urgent [65-68]: (1) 6G will address the technical limitations of previous networks (eg, 5G) with markedly superior communication solutions, (2) 6G can help elevate and enrich AI applications' full potentials, (3) 6G technologies coupled with AI techniques can truly materialize smart dementia care—networking infrastructure with the ability to seamlessly transform data points into evidence-based and intelligent decision-making processes [69], and in turn, (4) offer much-needed insights into dementia care and management from a big data analysis perspective, effectively bridging some of the most ingrained research gaps in the literature (eg, lack of large population-based empirical research) [47].

For instance, Project Relate is an AI-powered communication tool developed by Google that aims to help people with speech impairments communicate smoothly [37]. Different from the general public, people with speech impairments, such as dementia patients, often exhibit more obtuse or convoluted linguistic characteristics [70]. By first asking people to record their personal speech and then analyzing and identifying the unique linguistic identifiers using AI-powered algorithms, Project Relate can subsequently help users "translate" what could be considered difficult-to-understand speeches into comprehensible conversations [37]. In other words, by building their own virtual villages with personal data, advanced technological solutions such as Project Relate could help dementia patients with speech impairments communicate with others more effectively and smoothly. The algorithms and the logics behind Project Relate could have revolutionary impacts on existing dementia care solutions that patients have full control over, such as social robots.

Social robots are interactive robots that could offer companionship and serve as an emotional or mental health booster for dementia patients [71-73]. In a study on the effects of PARO, a robotic seal, for instance, researchers found that the social robot helped dementia patients uphold their sense of life, increased their perceptions of social connection, and improved their positive emotions [74]. However, the use and application of social robots in dementia care face a wide range of issues [75,76]. For instance, the way PARO works is that it has a tactile stimulation-enabled "value" system, which dictates

its rule of engagement: the robotic seal enjoys being stroked and dislikes being hit [72]. In its relationship with its owner, PARO is programmed to remember if it's been stroked and act in ways that make it more likely to be stroked—subsequently, it gradually learns to develop a "personality" that its owner might appreciate [72]. In other words, when it comes to interacting with dementia patients, PARO's technical capabilities allow it to be reactive and receptive but fall short of equipping it with capabilities to be proactive and responsive. Not to mention that one of the most noticeable flaws of social robots is that they cannot initiate verbal communication with dementia patients.

One way to address this issue is via 6G and AI technologies. For instance, when pair advanced AI algorithms such as those applied in health solutions like Project Relate with real-time data analytics enabled by 6G technologies, social robots will no longer serve as reactive companions to dementia patients but function as proactive caregivers that can provide intelligent health analytics (eg, noninvasive video-based health monitoring systems like Oxehealth) [51] and decision assistance to patients. In other words, as indicated in Table 2, social robots enhanced with 6G and AI technologies could have the potential to: (1) detect and predict patients' agitation and/or violence episodes prior to disease onset or progression, (2) initiate communications, rather than waiting to be approached, (3) engage the patients with a wide range of interactions (eg, speech, physical interactions, etc) prior to the onset of aggression or violence, rather than solely based on reactive motion-activation technologies, and (4) process and analyze data gained from the interactions to build a patient-by-patient database system to facilitate personalized dementia care.

Advanced Technology-Enabled Work

Recent advancements, such as the "avatar work" model, could further expand the promises of advanced technology-enabled dementia care [39,77]. With the help of advanced AI technologies, researchers are able to build systems that could enable people with physical impairments to "work" as robots in cafés that are appropriately named as Diverse Avatar Working Network (DWAN) [78]. Essentially, what is innovative about the "avatar work" or the "Work-Remotely-Café" model is that it allows people who are house-bound or bed-ridden, such as dementia patients, to work remotely in the form of physical robots via virtually controlling these robots using an AI-powered system at home or even in bed (Figure 2) [39]. In other words, what the DAWN model shows is that, with the help of advanced AI-powered algorithms, people with dementia could work as pilots of the robots (Figure 3).

Figure 2. User interface of the "Avatar robot café" AI system.

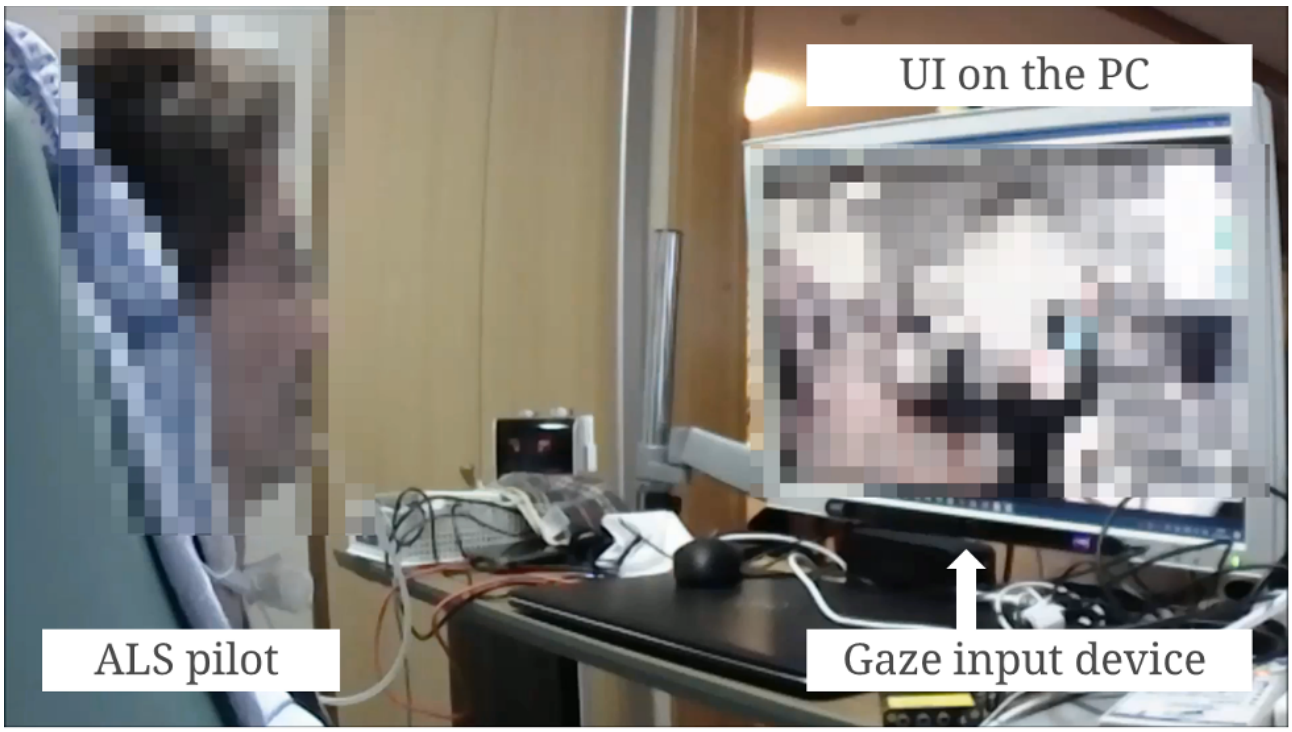


Figure 3. An example of piloted robot in action.

It is important to underscore that even people with severe debilitating physical impairment can benefit from the AI-powered robots—similar to the abovementioned Project Relate, only minimum motor capabilities are required for people to work as pilots. Unfortunately, at the moment, the DAWN avatar robot café is only available in Japan, and its first permanent location opened in June 2021 in Tokyo [39]. However, in light of the fast development of AI technologies and the ever-expanding dementia population [79-81], it is safe to assume there will be similar or more advanced 6G and AI-powered frameworks that would enable "limited-mobility-yet-unlimited-opportunities" for people with dementia in the future.

Limitations

While this study bridges vital gaps in the literature, it is not without limitations. First, the current study only focused on care solutions for people with dementia rather than informal caregivers or formal caregivers like health care professionals. To address this limitation, future research could investigate dementia health solutions that are designed for stakeholders other than the patients. Furthermore, in light of the parameters of our research aim, we only gauged advanced technology-based dementia care solutions in the context of 6G and AI opportunities. It is important to note that nontechnology-based dementia care solutions as well as technologies that are beyond the scale and scope of 6G and AI also play a critical role in dementia care [82,83]. In addition, the current study is not a systematic review, a limitation that could be sufficiently addressed as evidence continues to accumulate.

Conclusions

To live is to grow old. Yet dementia is neither a proper way to live nor a natural way to grow old. While scientists continue to work on finding effective pharmaceutical interventions that could help cure or cub dementia, speedy, supportive, and successful non-pharmaceutical interventions are needed to address and alleviate the everyday health challenges people with dementia face. By identifying advanced health solutions powered by 6G and AI opportunities, our study shed light on how patients with dementia can live and prosper.

Acknowledgments

The authors wish to express their gratitude to the editor and reviewers for their kind feedback and constructive input.

Data Availability

Data are available upon reasonable request.

Authors' Contributions

ZS conceived the work, reviewed the literature and drafted and edited the manuscript. BLB, DMD, JA, JH, FS, KT, AC, and CPdV reviewed the literature and edited the manuscript. All authors approved the manuscript for submission.

Conflicts of Interest

None declared.

References

1. Abbott A. Dementia: A problem for our age. Nature 2011 Jul 13;475(7355):S2-S4. [doi: 10.1038/475s2a]
2. Alzheimer's Disease International. World Alzheimer report 2018. Alzheimer's Disease International. 2018. URL: https://www.alzint.org/u/WorldAlzheimerReport2018.pdf [accessed 2022-02-02]
3. World HO. Global Dementia Observatory. 2021. URL: https://apps.who.int/gho/data/node.dementia [accessed 2021-11-20]
4. World HO. Dementia. 2021. URL: https://www.who.int/health-topics/dementia#tab=tab_1 [accessed 2021-11-20]
5. What Is Dementia? 2021. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/aging/dementia/index.html [accessed 2021-11-17]
6. Mullard A. Landmark Alzheimer's drug approval confounds research community. Nature 2021 Jun 08;594(7863):309-310. [doi: 10.1038/d41586-021-01546-2]
7. Katzman R. Dementia. Alzheimer Disease & Associated Disorders 1996;10(1):54. [doi: 10.1097/00002093-199603000-00011]
8. Biessels GJ, Despa F. Cognitive decline and dementia in diabetes mellitus: mechanisms and clinical implications. Nat Rev Endocrinol 2018 Jul 18;14(10):591-604. [doi: 10.1038/s41574-018-0048-7]
9. Cohen G, Russo MJ, Campos JA, Allegri RF. Living with dementia: increased level of caregiver stress in times of COVID-19. Int. Psychogeriatr 2020 Jul 30;32(11):1377-1381. [doi: 10.1017/s1041610220001593]
10. Reckrey JM, Morrison RS, Boerner K, Szanton SL, Bollens‐Lund E, Leff B, et al. Living in the Community With Dementia: Who Receives Paid Care? J Am Geriatr Soc 2019 Nov 06;68(1):186-191. [doi: 10.1111/jgs.16215]
11. Zhang Z, Chen X, Liu X, Tang M, Zhao H, Jue Q, et al. [A caregiver survey in Beijing, Xi'an, Shanghai and Chengdu: health services status for the elderly with dementia]. Zhongguo Yi Xue Ke Xue Yuan Xue Bao 2004 Apr;26(2):116-121. [Medline: 15171545]
12. Black BS, Johnston D, Rabins PV, Morrison A, Lyketsos C, Samus QM. Unmet Needs of Community‐Residing Persons with Dementia and Their Informal Caregivers: Findings from the Maximizing Independence at Home Study. J Am Geriatr Soc 2013 Nov 25;61(12):2087-2095. [doi: 10.1111/jgs.12549]
13. Mkhonto F, Hanssen I. When people with dementia are perceived as witches. Consequences for patients and nurse education in South Africa. J Clin Nurs 2017 Nov 20;27(1-2):e169-e176. [doi: 10.1111/jocn.13909]
14. Jutkowitz E, Scerpella D, Pizzi LT, Marx K, Samus Q, Piersol CV, et al. Dementia Family Caregivers' Willingness to Pay for an In-home Program to Reduce Behavioral Symptoms and Caregiver Stress. Pharmacoeconomics 2019 Apr 15;37(4):563-572 [FREE Full text] [doi: 10.1007/s40273-019-00785-6] [Medline: 30877638]
15. Huang M, Tsai CF, Lee WJ, Wang WF, Jhang KM, Wang SJ, et al. Caregivers' willingness to pay for Alzheimer's disease medications in Taiwan. J Chin Med Assoc 2021 Apr 01;84(4):446-452. [doi: 10.1097/JCMA.0000000000000504] [Medline: 33595992]
16. Lin M. Communication neglect, caregiver anger and hostility, and perceptions of older care receivers' cognitive status and problem behaviors in explaining elder abuse. J Elder Abuse Negl 2020 Jun 17;32(3):235-258. [doi: 10.1080/08946566.2020.1741054] [Medline: 32183610]
17. Fang B, Yan E, Yang X, Pei Y. Association between Caregiver Neurotic Personality Trait and Elder Abuse: Investigating the Moderating Role of Change in the Level of Caregiver Perceived Burden. Gerontology 2021 Jan 15;67(2):243-254. [doi: 10.1159/000512238] [Medline: 33454703]
18. Wang M, Sun H, Zhang J, Ruan J. Prevalence and associated factors of elder abuse in family caregivers of older people with dementia in central China cross-sectional study. Int J Geriatr Psychiatry 2019 Feb 27;34(2):299-307. [doi: 10.1002/gps.5020] [Medline: 30370657]
19. Black BS, Johnston D, Leoutsakos J, Reuland M, Kelly J, Amjad H, et al. Unmet needs in community-living persons with dementia are common, often non-medical and related to patient and caregiver characteristics. Int. Psychogeriatr 2019 Feb 04;31(11):1643-1654. [doi: 10.1017/s1041610218002296]

20. Handels RL, Sköldunger A, Bieber A, Edwards RT, Gonçalves-Pereira M, Hopper L, et al. Quality of Life, Care Resource Use, and Costs of Dementia in 8 European Countries in a Cross-Sectional Cohort of the Actifcare Study. JAD 2018 Nov 23;66(3):1027-1040. [doi: 10.3233/jad-180275]
21. Monin JK, Jorgensen TD, MacNeil Vroomen JL. Self-Reports and Caregivers' Proxy Reports of Unmet Needs of Persons With Dementia: Implications for Both Partners' Health-Related Quality of Life. Am J Geriatr Psychiatry 2020 Mar;28(3):363-367 [FREE Full text] [doi: 10.1016/j.jagp.2019.10.006] [Medline: 31708379]
22. Janssen N, Handels RL, Sköldunger A, Woods B, Jelley H, Edwards RT, et al. Impact of Untimely Access to Formal Care on Costs and Quality of Life in Community Dwelling People with Dementia. JAD 2018 Nov 23;66(3):1165-1174. [doi: 10.3233/jad-180531]
23. Herron R, Rosenberg M. "Not there yet": Examining community support from the perspective of people with dementia and their partners in care. Soc Sci Med 2017 Jan;173:81-87. [doi: 10.1016/j.socscimed.2016.11.041] [Medline: 27930919]
24. Wong-Lin K, McClean PL, McCombe N, Kaur D, Sanchez-Bornot JM, Gillespie P, et al. Shaping a data-driven era in dementia care pathway through computational neurology approaches. BMC Med 2020 Dec 16;18(1):398 [FREE Full text] [doi: 10.1186/s12916-020-01841-1] [Medline: 33323116]
25. Enshaeifar S, Barnaghi P, Skillman S, Markides A, Elsaleh T, Acton ST, et al. The Internet of Things for Dementia Care. IEEE Internet Comput 2018 Jan;22(1):8-17. [doi: 10.1109/mic.2018.112102418]
26. Koo BM, Vizer LM. Examining Mobile Technologies to Support Older Adults With Dementia Through the Lens of Personhood and Human Needs: Scoping Review. JMIR Mhealth Uhealth 2019 Nov 11;7(11):e15122 [FREE Full text] [doi: 10.2196/15122] [Medline: 31710305]
27. Bronfenbrenner U. Toward an experimental ecology of human development. American Psychologist 1977;32(7):513-531. [doi: 10.1037/0003-066x.32.7.513]
28. National HS. About dementia. 2021. URL: https://www.nhs.uk/conditions/dementia/about/ [accessed 2021-11-21]
29. National IOA. What Is dementia? Symptoms, types, and diagnosis. 2021. URL: https://www.nia.nih.gov/health/what-is-dementia [accessed 2021-11-21]
30. Government OI. The Irish National Dementia Strategy. 2020. URL: https://www.gov.ie/en/publication/62d6a5-national-dementia-strategy/ [accessed 2021-11-21]
31. Faruq RN, D'Silva P, Lau FD, Zhao C, Majumdar S. Early-Onset Vascular Dementia in a 43-Year-Old Man with Accelerated Atherosclerotic Disease, Elevated Lipoprotein (a), and a Missense DNAJC5 Variant with Potential Association to Adult-Onset Ceroid Lipofuscinosis. Case Rep Neurol 2021 Aug 24;13(2):565-571 [FREE Full text] [doi: 10.1159/000518194] [Medline: 34720963]
32. Hanson E, Hellström A, Sandvide, Jackson GA, MacRae R, Waugh A, et al. The extended palliative phase of dementia - An integrative literature review. Dementia (London) 2019 Jan 26;18(1):108-134. [doi: 10.1177/1471301216659797] [Medline: 27460046]
33. Gogia B. Differentiating Delirium Versus Dementia in the Elderly. In: StatPearls. Treasure Island, Florida, USA: StatPearls; 2021:1.
34. Banovic S, Zunic L, Sinanovic O. Communication Difficulties as a Result of Dementia. Mater Sociomed 2018 Oct;30(3):221-224 [FREE Full text] [doi: 10.5455/msm.2018.30.221-224] [Medline: 30515063]
35. Zewe A. A robot that finds lost items. 2021. URL: https://news.mit.edu/2021/robot-finds-items-camera-antenna-1005 [accessed 2021-11-21]
36. Facebook. Egocentric Live 4D Perception (Ego4D). 2021. URL: https://ai.facebook.com/tools/ego4d [accessed 2021-11-21]
37. Cattiau J. A communication tool for people with speech impairments. 2021. URL: https://blog.google/outreach-initiatives/accessibility/project-relate/ [accessed 2021-11-17]
38. Lillianfeld L, Nikka A. Two new tools that make your phone even more accessible. Google. URL: https://blog.google/outreach-initiatives/accessibility/making-android-more-accessible/ [accessed 2021-11-17]
39. Takeuchi K, Yamazaki Y, Yoshifuji K. Avatar Work: Telework for Disabled People Unable to Go Outside by Using Avatar Robots. In: ACM. 2020 Presented at: 2020 ACM/IEEE International Conference on Human-Robot Interaction; 2020; Cambridge, United Kingdom p. 53-60. [doi: 10.1145/3371382.3380737]
40. Affectiva. Interior Sensing AI. 2021. URL: https://go.affectiva.com/auto [accessed 2021-11-21]
41. Jøranson N, Pedersen I, Rokstad AMM, Ihlebæk C. Effects on Symptoms of Agitation and Depression in Persons With Dementia Participating in Robot-Assisted Activity: A Cluster-Randomized Controlled Trial. J Am Med Dir Assoc 2015 Oct 01;16(10):867-873. [doi: 10.1016/j.jamda.2015.05.002] [Medline: 26096582]
42. Petersen S, Houston S, Qin H, Tague C, Studley J. The Utilization of Robotic Pets in Dementia Care. JAD 2016 Nov 19;55(2):569-574. [doi: 10.3233/jad-160703]
43. Jøranson N, Pedersen I, Rokstad AMM, Ihlebaek C. Change in quality of life in older people with dementia participating in Paro-activity: a cluster-randomized controlled trial. J Adv Nurs 2016 Dec 16;72(12):3020-3033. [doi: 10.1111/jan.13076] [Medline: 27434512]
44. Liang A, Piroth I, Robinson H, MacDonald B, Fisher M, Nater UM, et al. A Pilot Randomized Trial of a Companion Robot for People With Dementia Living in the Community. J Am Med Dir Assoc 2017 Oct 01;18(10):871-878. [doi: 10.1016/j.jamda.2017.05.019] [Medline: 28668664]

45. Moyle W, Jones CJ, Murfield JE, Thalib L, Beattie ER, Shum DK, et al. Use of a Robotic Seal as a Therapeutic Tool to Improve Dementia Symptoms: A Cluster-Randomized Controlled Trial. J Am Med Dir Assoc 2017 Sep 01;18(9):766-773 [FREE Full text] [doi: 10.1016/j.jamda.2017.03.018] [Medline: 28780395]
46. Demange M, Lenoir H, Pino M, Cantegreil-Kallen I, Rigaud AS, Cristancho-Lacroix V. Improving well-being in patients with major neurodegenerative disorders: differential efficacy of brief social robot-based intervention for 3 neuropsychiatric profiles. CIA 2018 Jul;Volume 13:1303-1311. [doi: 10.2147/cia.s152561]
47. Obayashi K, Masuyama S. Pilot and Feasibility Study on Elderly Support Services Using Communicative Robots and Monitoring Sensors Integrated With Cloud Robotics. Clin Ther 2020 Feb;42(2):364-371.e4. [doi: 10.1016/j.clinthera.2020.01.001] [Medline: 31987582]
48. Obayashi K, Kodate N, Masuyama S. Measuring the impact of age, gender and dementia on communication-robot interventions in residential care homes. Geriatr Gerontol Int 2020 Apr 19;20(4):373-378. [doi: 10.1111/ggi.13890] [Medline: 32077237]
49. Rouaix N, Retru-Chavastel L, Rigaud A, Monnet C, Lenoir H, Pino M. Affective and Engagement Issues in the Conception and Assessment of a Robot-Assisted Psychomotor Therapy for Persons with Dementia. Front Psychol 2017 Jun 30;8:950 [FREE Full text] [doi: 10.3389/fpsyg.2017.00950] [Medline: 28713296]
50. Thevenot J, Lopez MB, Hadid A. A Survey on Computer Vision for Assistive Medical Diagnosis From Faces. IEEE J. Biomed. Health Inform 2018 Sep;22(5):1497-1511. [doi: 10.1109/jbhi.2017.2754861]
51. US Food and Drug Administration. Device classification under section 513(f)(2)(de novo). 2021. URL: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/denovo.cfm?id=DEN200019 [accessed 2021-04-29]
52. Neurable. Smart headphones for smarter focus. 2021. URL: https://neurable.com/headphones [accessed 2021-05-02]
53. Kuwamura K, Nishio S, Sato S. Can We Talk through a Robot As if Face-to-Face? Long-Term Fieldwork Using Teleoperated Robot for Seniors with Alzheimer's Disease. Front Psychol 2016 Jul 19;7:1066 [FREE Full text] [doi: 10.3389/fpsyg.2016.01066] [Medline: 27486416]
54. Cormons L. Testing social robot acceptance: What if you could be assessed for dementia by a robot? A pilot study. In: IEEE. 2020 Presented at: 2020 6th International Conference on Mechatronics and Robotics Engineering (ICMRE). 2020; 2020; Barcelona, Spain p. 92-98. [doi: 10.1109/ICMRE49073.2020.9065021]
55. The strong robot with the gentle touch. RIKEN. 2015. URL: https://www.riken.jp/en/news_pubs/research_news/pr/2015/20150223_2/ [accessed 2021-05-02]
56. Chen K, Lou VW, Tan KC, Wai M, Chan L. Effects of a Humanoid Companion Robot on Dementia Symptoms and Caregiver Distress for Residents in Long-Term Care. J Am Med Dir Assoc 2020 Nov;21(11):1724-1728.e3. [doi: 10.1016/j.jamda.2020.05.036] [Medline: 32713772]
57. Saad W, Bennis M, Chen M. A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems. IEEE Network 2020 May;34(3):134-142. [doi: 10.1109/mnet.001.1900287]
58. Samsung Research. 6G: The next hyper connected experience for all. Samsung.: Samsung; 2020. URL: https://cdn.codeground.org/nsr/downloads/researchareas/20201201_6G_Vision_web.pdf [accessed 2021-06-10]
59. Zhou Y, Liu L, Wang L, Hui N, Cui X, Wu J, et al. Service-aware 6G: An intelligent and open network based on the convergence of communication, computing and caching. Digital Communications and Networks 2020 Aug;6(3):253-260. [doi: 10.1016/j.dcan.2020.05.003]
60. Nawaz F, Ibrahim J, Awais M, Junaid M, Kousar S, Parveen T. A Review of Vision and Challenges of 6G Technology. IJACSA 2020;11(2):643-649. [doi: 10.14569/ijacsa.2020.0110281]
61. Alsharif MH, Kelechi AH, Albreem MA, Chaudhry SA, Zia MS, Kim S. Sixth Generation (6G) Wireless Networks: Vision, Research Activities, Challenges and Potential Solutions. Symmetry 2020 Apr 24;12(4):676. [doi: 10.3390/sym12040676]
62. Dang S, Amin O, Shihada B, Alouini M. What should 6G be? Nat Electron 2020 Jan 24;3(1):20-29. [doi: 10.1038/s41928-019-0355-6]
63. Chowdhury MZ, Shahjalal M, Ahmed S, Jang YM. 6G Wireless Communication Systems: Applications, Requirements, Technologies, Challenges, and Research Directions. IEEE Open J. Commun. Soc 2020;1:957-975. [doi: 10.1109/ojcoms.2020.3010270]
64. Letaief KB, Chen W, Shi Y, Zhang J, Zhang YA. The Roadmap to 6G: AI Empowered Wireless Networks. IEEE Commun. Mag 2019 Aug;57(8):84-90. [doi: 10.1109/mcom.2019.1900271]
65. Khan LU, Yaqoob I, Imran M, Han Z, Hong CS. 6G Wireless Systems: A Vision, Architectural Elements, and Future Directions. IEEE Access 2020;8:147029-147044. [doi: 10.1109/access.2020.3015289]
66. Zhang S, Liu J, Guo H, Qi M, Kato N. Envisioning Device-to-Device Communications in 6G. IEEE Network 2020 May;34(3):86-91. [doi: 10.1109/mnet.001.1900652]
67. Sheth K, Patel K, Shah H, Tanwar S, Gupta R, Kumar N. A taxonomy of AI techniques for 6G communication networks. Computer Communications 2020 Sep;161:279-303. [doi: 10.1016/j.comcom.2020.07.035]
68. Calvanese Strinati E, Barbarossa S, Gonzalez-Jimenez JL, Ktenas D, Cassiau N, Maret L, et al. 6G: The Next Frontier: From Holographic Messaging to Artificial Intelligence Using Subterahertz and Visible Light Communication. IEEE Veh. Technol. Mag 2019 Sep;14(3):42-50. [doi: 10.1109/mvt.2019.2921162]

69. Janbi N, Katib I, Albeshri A, Mehmood R. Distributed Artificial Intelligence-as-a-Service (DAIaaS) for Smarter IoE and 6G Environments. Sensors (Basel) 2020 Oct 13;20(20):5796 [FREE Full text] [doi: 10.3390/s20205796] [Medline: 33066295]
70. Long Z, Irish M, Piguet O, Kiernan MC, Hodges JR, Burrell JR. Clinical and neuroimaging investigations of language disturbance in frontotemporal dementia-motor neuron disease patients. J Neurol 2019 Apr 1;266(4):921-933. [doi: 10.1007/s00415-019-09216-0] [Medline: 30707358]
71. Choi SSW, Budhathoki C, Gitlin LN. Co-Occurrence and Predictors of Three Commonly Occurring Behavioral Symptoms in Dementia: Agitation, Aggression, and Rejection of Care. Am J Geriatr Psychiatry 2017 May;25(5):459-468. [doi: 10.1016/j.jagp.2016.10.013] [Medline: 27914870]
72. Shibata T. Therapeutic Seal Robot as Biofeedback Medical Device: Qualitative and Quantitative Evaluations of Robot Therapy in Dementia Care. Proc. IEEE 2012 Aug;100(8):2527-2538. [doi: 10.1109/jproc.2012.2200559]
73. Koh WQ, Ang FXH, Casey D. Impacts of Low-cost Robotic Pets for Older Adults and People With Dementia: Scoping Review. JMIR Rehabil Assist Technol 2021 Feb 12;8(1):e25340 [FREE Full text] [doi: 10.2196/25340] [Medline: 33497349]
74. Hung L, Gregorio M, Mann J, Wallsworth C, Horne N, Berndt A, et al. Exploring the perceptions of people with dementia about the social robot PARO in a hospital setting. Dementia (London) 2021 Feb 10;20(2):485-504 [FREE Full text] [doi: 10.1177/1471301219894141] [Medline: 31822130]
75. Robinson NL, Cottier TV, Kavanagh DJ. Psychosocial Health Interventions by Social Robots: Systematic Review of Randomized Controlled Trials. J Med Internet Res 2019 May 10;21(5):e13203 [FREE Full text] [doi: 10.2196/13203] [Medline: 31094357]
76. Pu L, Moyle W, Jones C, Todorovic M. The Effectiveness of Social Robots for Older Adults: A Systematic Review and Meta-Analysis of Randomized Controlled Studies. Gerontologist 2019 Jan 09;59(1):e37-e51. [doi: 10.1093/geront/gny046] [Medline: 29897445]
77. Lee S, O'Neill D, Moss H. Dementia-inclusive group-singing online during COVID-19: A qualitative exploration. Nordic Journal of Music Therapy 2021 Oct 22:1-19. [doi: 10.1080/08098131.2021.1963315]
78. Diverse AWN. Giving all the possibility to stay connected with society. 2021. URL: https://dawn2021.orylab.com/en/ [accessed 2021-11-21]
79. Nichols E, Szoeke C, Vollset S, Abbasi N, Abd-Allah F, Abdela J, et al. Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet Neurology 2019 Jan;18(1):88-106. [doi: 10.1016/S1474-4422(18)30403-4]
80. Wolters FJ, Chibnik LB, Waziry R, Anderson R, Berr C, Beiser A, et al. Twenty-seven-year time trends in dementia incidence in Europe and the United States. Neurology 2020 Jul 01;95(5):e519-e531. [doi: 10.1212/wnl.0000000000010022]
81. Cao Q, Tan C, Xu W, Hu H, Cao X, Dong Q, et al. The Prevalence of Dementia: A Systematic Review and Meta-Analysis. JAD 2020 Feb 04;73(3):1157-1166. [doi: 10.3233/jad-191092]
82. Nielsen TR, Nielsen DS, Waldemar G. Barriers in access to dementia care in minority ethnic groups in Denmark: a qualitative study. Aging Ment Health 2021 Aug 03;25(8):1424-1432. [doi: 10.1080/13607863.2020.1787336] [Medline: 32619352]
83. Ibsen TL, Kirkevold, Patil GG, Eriksen S. People with dementia attending farm-based day care in Norway - Individual and farm characteristics associated with participants' quality of life. Health Soc Care Community 2020 May;28(3):1038-1048. [doi: 10.1111/hsc.12937] [Medline: 31884707]

Abbreviations

6G: sixth-generation communication solutions
AI: artificial intelligence
DWAN: Diverse Avatar Working Network

Edited by A Mavragani; submitted 19.05.21; peer-reviewed by D Yang Meier, R Schmaelzle, F Cruz, P Zou, H Lee; comments to author 04.08.21; revised version received 23.11.21; accepted 31.01.22; published 27.04.22.

Please cite as:
Su Z, Bentley BL, McDonnell D, Ahmad J, He J, Shi F, Takeuchi K, Cheshmehzangi A, da Veiga CP
6G and Artificial Intelligence Technologies for Dementia Care: Literature Review and Practical Analysis
J Med Internet Res 2022;24(4):e30503
URL: https://www.jmir.org/2022/4/e30503
doi:10.2196/30503
PMID:35475733

Review

Consumers' Evaluation of Web-Based Health Information Quality: Meta-analysis

Yan Zhang[1,2*], PhD; Yeolib Kim[3*], PhD

[1]School of Information, The University of Texas at Austin, Austin, TX, United States

[2]Center for Health Communication, Moody College of Communication and Dell Medical School, The University of Texas at Austin, Austin, TX, United States

[3]Graduate School of Technology & Innovation Management, School of Business Administration, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea

[*]all authors contributed equally

Corresponding Author:
Yeolib Kim, PhD
Graduate School of Technology & Innovation Management
School of Business Administration
Ulsan National Institute of Science and Technology
50 UNIST-gil
Ulsan, 44919
Republic of Korea
Phone: 82 52 217 3189
Email: yeolib.kim@unist.ac.kr

Abstract

Background: The internet has become a major source of health information for general consumers. Web-based health information quality varies widely across websites and applications. It is critical to understand the factors that shape consumers' evaluation of web-based health information quality and the role that it plays in their appraisal and use of health information and information systems.

Objective: This paper aimed to identify the antecedents and consequences of consumers' evaluation of web-based health information quality as a means to consolidate the related research stream and to inform future studies on web-based health information quality.

Methods: We systematically searched 10 databases, examined reference lists, and conducted manual searches. Empirical studies that investigated consumers' evaluation of web-based health information quality, credibility, or trust and their respective relationships with antecedents or consequences were included.

Results: We included 147 studies reported in 136 papers in the analysis. Among the antecedents of web-based health information quality, system navigability (ρ=0.56), aesthetics (ρ=0.49), and ease of understanding (ρ=0.49) had the strongest relationships with web-based health information quality. The strongest consequences of web-based health information quality were consumers' intentions to use health information systems (ρ=0.58) and satisfaction with health information (ρ=0.46). Web-based health information quality relationships were moderated by numerous cultural dimensions, research designs, and publication moderators.

Conclusions: Consumers largely rely on peripheral cues and less on cues that require more information processing (eg, content comprehensiveness) to determine web-based health information quality. Surprisingly, the relationships between individual differences and web-based health information quality are trivial. Web-based health information quality has stronger effects on cognitive appraisals and behavioral intentions than on behavior. Despite efforts to include various moderators, a substantial amount of variance is still unexplained, indicating a need to study additional moderators. This meta-analysis provides broad and consistent evidence for web-based health information quality relationships that have been fractured and incongruent in empirical studies.

(J Med Internet Res 2022;24(4):e36463) doi:10.2196/36463

KEYWORDS

online health information; information quality; credibility; trust; consumer health information behavior; meta-analysis

Introduction

The internet has become a major source of health information for general consumers. However, health information quality (IQ) varies widely across websites and web applications, and the overall quality is concerning [1,2]. Low-quality information conveys incomplete, inaccurate, or outdated knowledge, which may lead users to form erroneous health beliefs and cause negative, or even detrimental, health outcomes. Owing to the immense ramifications, web-based health IQ has attracted continued attention from researchers, health care professionals, and consumers alike.

The IQ construct has been defined in a disparate fashion. Some researchers have taken an objective view, defining IQ in relation to currently accepted medical guidelines [3]. Others recognized that the evaluation of IQ is contingent on users' tasks, goals, and value judgments [4-6] and defined IQ, from a subjective view, as users' perceptions of IQ [7] or "fitness for use" [8]. For the purpose of this review, we adopted the view of IQ in the study by McKinney et al [7] and defined web-based health IQ as users' perceptions of the quality of health information on the internet. In the internet context, two other concepts share this notion: credibility and trust. Credibility is often defined as perceived IQ, whereas trust denotes users' willingness to trust web-based information [9].

Some researchers have differentiated these 3 concepts. For instance, some view IQ as a dimension of credibility or a factor that influences credibility judgment [10], whereas others view credibility as a major dimension of IQ [11]. Some view IQ [12-14] or credibility [15] as antecedents of trust, whereas others view trustworthiness as a major dimension of credibility [16]. Despite these differences, the 3 concepts are intertwined. In the literature on consumers' web-based health information seeking, they all, to some degree, refer to consumers' perceived quality of web-based health information [17,18]. To achieve comprehensive coverage of the literature, we included studies that used any of the 3 terms to refer to health consumers' perceptions of web-based health IQ.

Systematic reviews concerning IQ, trust, and credibility of web-based health information have recently been published. Sun et al [19] identified the criteria and indicators that consumers use to evaluate web-based health IQ. Sbaffi and Rowley [18] identified factors that affect consumers' trust in and the perceived credibility of web-based health information. Kim [20] identified antecedents of trust in web-based health information. On a related note, Diviani et al [17] examined the relationship between health literacy and consumer evaluation of web-based health information. These reviews provide a comprehensive view of how consumers evaluate web-based health IQ and outline categories of antecedents of web-based health IQ, such as individual factors (eg, sociodemographic and health status), source factors (eg, reputation), content factors (eg, relevance and usefulness), and design factors (eg, layout and ease of use).

However, these reviews have several limitations. First, a plethora of antecedents of web-based health IQ was identified; however, few syntheses and comparisons were performed, resulting in a rather murky view of the most influential antecedents and how they affect web-based health IQ evaluation. Second, little effort was made to amalgamate the consequences of web-based health IQ. Third, little effort was made to explain inconsistent results across studies. For example, health literacy (and education levels and other skill-based proxies for health literacy) had a significant positive effect on perceived web-based health IQ and trust in some studies [21-24] but a negative [25-27] or insignificant [28-30] effect in others. These inconsistent results indicate that web-based health IQ relationships may be moderated by contextual factors [31].

To further enhance our knowledge of the existence, nature, and magnitude of web-based health IQ relationships and elucidate the conceptual and practical significance of the concept [32], we performed a meta-analysis to address the following research questions: (1) what antecedents and consequences are relevant to consumers' evaluations of web-based health IQ, and (2) what moderators intervene in web-based health IQ relationships?

Methods

Search Strategy

A systematic search of the literature published since 2000 was performed in July 2020 on 10 databases (eg, PubMed, CINAHL, and PsycINFO), using the search query *health information* AND (*credibility* OR *quality* OR *reliability* OR *trust*) AND (*online* OR *Internet* OR *web*) within the title, abstract, and keyword fields of these databases. In addition, we tracked the references of the included papers using Google Scholar. To reduce publication bias, we also searched the ProQuest dissertation and thesis database and reviewed the proceedings of several related conferences.

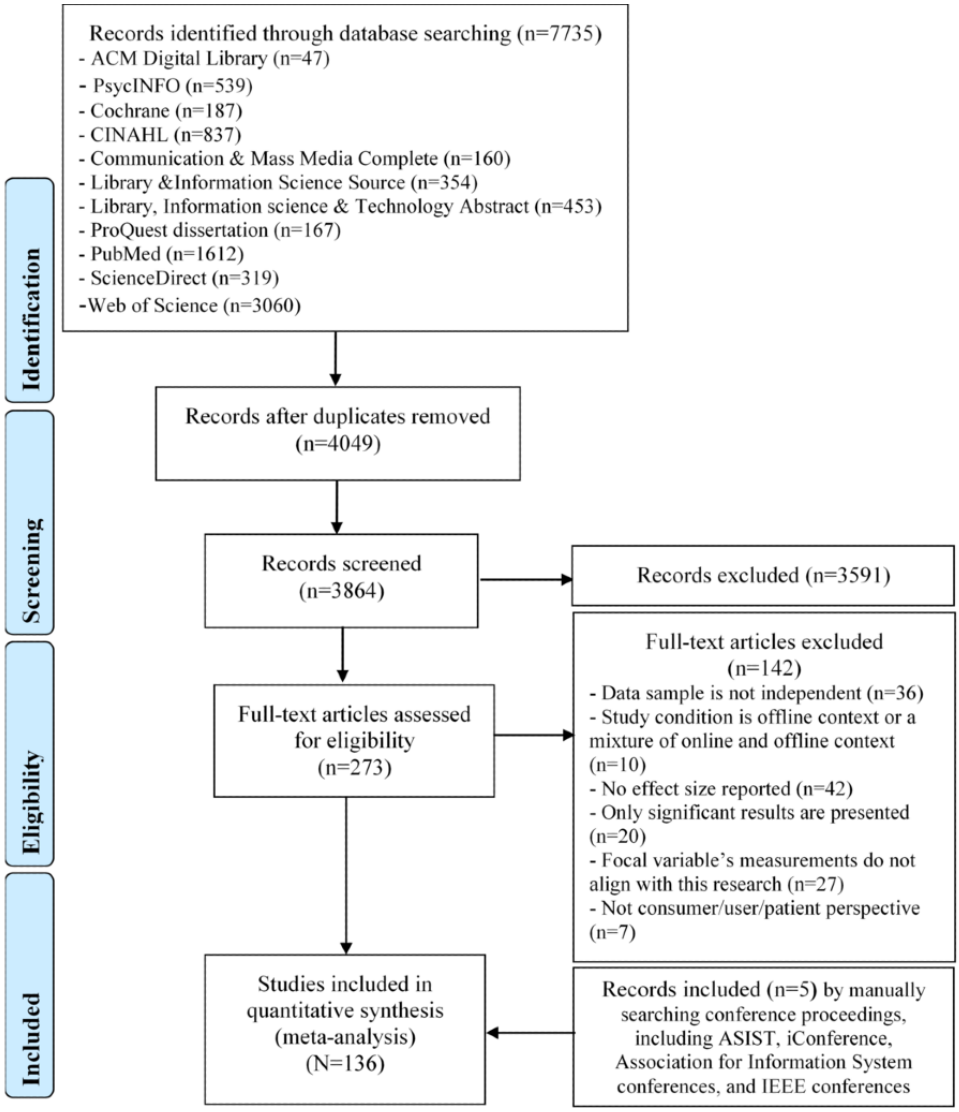
Inclusion Criteria, Exclusion Criteria, and Screening

The studies included in this review were empirical studies that reported effect sizes for web-based health IQ relationships. Studies were excluded if they met the following criteria: (1) focused on health care providers, (2) used qualitative research methods, (3) studied patients' or consumers' perceptions of the quality of information from noninternet sources (eg, health care providers and newspapers), (4) were not independent samples, (5) did not report effect sizes, (6) only reported significant results, and (7) were not in English.

Unique records resulting from the search were screened against the inclusion and exclusion criteria. First, two reviewers (YZ and SS) independently reviewed the titles and abstracts of the 100 randomly selected records. The full text was retrieved and perused when a decision could not be reached based on the title and abstract. The intercoder agreement was moderate (Cohen κ=0.51). Discrepancies were discussed, and we clarified the inclusion and exclusion criteria. Then, the 2 coders independently coded another 50 randomly selected records. The intercoder agreement reached 0.71. Discrepancies were discussed and resolved again. SS then screened the rest of the records. The screening was purposely kept broad to avoid missing relevant studies. The overall process resulted in 273 papers. YK reviewed the full text of these papers and further excluded 142. Relevant papers from related conference

proceedings were added, resulting in a final sample of 136 papers, which reported 147 studies. The paper screening and identification procedures are illustrated in Figure 1. The list of studies is reported in Multimedia Appendix 1.

Figure 1. Paper screening and identification procedure.



Data Extraction and Meta-analytic Approach

One of the reviewers (YK) extracted and coded the following data from the included studies using Microsoft Excel:

1. Basic paper features: title, publication outlet, author, publication year, and publication type (journal and nonjournal)
2. Research design: stimulus type (specific vs general), technology context (social media vs nonsocial media), sample size, sample clinical status (patient vs nonpatient), sample type (student vs nonstudent), operationalization (quality vs credibility vs trust), sample year, number of instrument items, measurement reliability, sample country, sample culture dimensions, and study methods (survey vs experiment); the values for cultural dimensions were obtained by inputting the sample country into the website of Hofstede [33]
3. Antecedents and consequences: antecedent and consequence variables (when authors of the included papers used different terms to describe the same or similar concept, the terms were grouped under a preferred name; eg, the construct of direct experience with cancer in the study by Feng and Yang [34] and the construct of perceived severity of mental health in the study by McKinley and Ruppel [35] were coded as health experience and beliefs; constructs were categorized as antecedents or consequences as in the original studies), reliability scores when available, and effect sizes for specific antecedents and consequences (ie, correlations, odds ratio, β, chi-square, *F* statistic, and *t* statistic; the latter 4 were subsequently converted into correlations using formulas [36-38]).

YZ reviewed the coded data against the original full-text papers to ensure accuracy and consistency. The interrater agreement (Cohen κ) reached 0.93 for basic paper features and research design, 0.87 for grouping concepts, and 0.91 for effect sizes. Disagreements were discussed and resolved. Interested readers can contact the authors to obtain the meta-analysis database.

Following the methods by Hunter and Schmidt [36], this meta-analysis used a random-effects model to analyze

correlations (*r*s). Weighted mean correlations (ρ or main effects) were computed by correcting for measurement and sampling errors. Reliabilities from each study were used to correct measurement errors. In studies that did not report a reliability value, the mean reliability (Multimedia Appendix 2) was used as the substitute. Reliability was set at 1.00 for variables assumed to have no measurement error (eg, gender, age, education, income, and race). Sample sizes were used to correct for sampling errors. Various supporting statistics such as the 95% CI, 90% credibility interval, *Q* statistic, I^2 statistic, and Begg test were computed in addition to ρs. Heterogeneity was detected if the *Q* statistic was significant ($P<.05$), the I^2 was >75%, or the 90% credibility interval was *wide*. The Begg test [39] exposes where publication bias exists in the meta-analysis via funnel plot asymmetry, whereby $P<.05$ implicates publication bias.

Informed by prior meta-analyses on relevant topics [40-42] and the characteristics of the included studies, we examined three categories of moderators—cultural, research design, and publication—and the operationalization of web-based health IQ (quality, credibility, and trust), resulting in a total of 13 factors (Table 1).

All moderators were categorical; thus, subgroup analyses using a random-effects model [43] were conducted to calculate the mean ρs. Q_M, an omnibus test, was calculated to statistically compare subgroup means. Antecedents and consequences with a sufficient number of observations (≥20) were analyzed against moderators. Those without sufficient observations were combined into composite variables for the analysis based on conceptual similarities. The metafor package [44] in R was used to analyze the main and moderating effects.

Table 1. Moderators for web-based health information quality relationships.

Moderator	Definition and operationalization
Sample culture	This refers to the culture that sample participants belong to. It is operationalized by 5 cultural dimensions outlined in the cultural dimension theory by Hofstede [45].
Individualism versus collectivism	Individualism is "a preference for a loosely-knit social framework" where people are supposed to take care of only themselves or their close family members [46]. Collectivism represents a preference for in-group loyalties. People in a collectivistic society must unconditionally be in service to other in-group members to show their loyalty [47].
Power distance	This refers to the degree to which "the less powerful members of a society accept and expect that power is distributed unequally" [46]. In a society with high power distance, people "accept a hierarchical order in which everybody has a place and which needs no further justification." In a society with low power distance, people strive to "equalize the distribution of power and demand justification for inequalities of power" [46].
Uncertainty avoidance	This expresses "the degree to which members of a society feel uncomfortable with uncertainty and ambiguity." Societies with strong uncertainty avoidance are "intolerant of unorthodox behavior and ideas" [46], whereas societies with weak uncertainty avoidance exhibit a more relaxed attitude [48].
Long-term versus short-term orientation	A society with a long-term orientation fosters virtues oriented toward future rewards, in particular, perseverance and thrift [48]. A society with a short-term orientation "prefers to maintain time-honored traditions and norms while viewing societal change with suspicion" [46].
Indulgence versus restraint	Indulgence stands for a society's tendency to allow "relatively free gratification of basic and natural human desires related to enjoying life and having fun." Restraint stands for "a society that suppresses gratification of needs and reregulate it by strict social norms." [46,48].
Research design	This refers to the study's research methods to address research problems, including research settings, data collection, measurement, and the analysis of data.
Technology context	This refers to the internet technology platforms where a study situates their examination of web-based health information quality. The technology context was categorized into social media (eg, web-based health communities, Twitter, and Facebook) and non–social media (ie, general health websites).
Sample type	This refers to whether a sample comprises students or nonstudents.
Sample clinical status	This refers to people who assume to have no specific conditions or patients who have been diagnosed with particular conditions.
Study methods	This refers to the research methods that a study used to collect data. Two specific research methods were frequently used and thus coded for this meta-analysis: survey and experiment.
Stimulus type	This refers to the stimuli used in the included studies. Two types of stimuli were identified: general and specific. General stimuli are web-based health information in general (without specifications of information source and content). Specific stimuli are specific health information or health information systems (eg, a specific health website or a specific health message).
Publication	
Publication outlet	This refers to the venue where a study was published. Two types of publication outlets were defined: journal and nonjournal (including conference proceedings and theses and dissertations).
Time	This refers to when a study was published. Two periods were defined—before 2014 and in or after 2014—by applying the median split on the publication year.
Operationalization of web-based health information quality	This refers to the three focal concepts included in the analysis: web-based health information quality, credibility, and trust in web-based health information.

Results

Basic Characteristics of the Included Papers

The 136 papers included 109 (80.1%) journal articles, 20 (14.7%) conference papers, and 7 (5.2%) theses and dissertations. The publication years ranged from 2000 to 2020, with 75% (102/136) of the papers published after 2010. The health domains covered included both general and specific health topics (eg, schizophrenia, cancer, HIV, and prescription medications).

The included papers reported 147 independent studies. Sample sizes ranged from 34 to 8586 (median 252); 67.3% (99/147) of samples involved nonstudents, 32.7% (48/147) involved students, 8.8% (13/147) of samples were patients, and 91.2% (134/147) were nonpatients. Among the 133 samples that reported countries (15 countries), 76 (57.1%) were from the United States, followed by 16 (12%) from China, 10 (7.5%) from Korea, 8 (6%) from Germany, and 5 (3.8%) from Australia.

Antecedents and Consequences of Web-Based Health IQ

Table 2 presents 18 antecedents and 8 consequences of web-based health IQ with at least 6 observations. Those with the number of samples <6 were not included in the analysis as the results tend to be less generalizable [41]. The antecedents fell into four categories: individual difference, source, content,

and design. The consequences fell into three categories: cognitive appraisals, behavioral intentions, and behaviors.

Table 3 presents the main effects of the antecedents and consequences. Using the Cohen criteria [49] for judging the magnitude of correlation effect sizes, the design factor—navigability—was most strongly related to web-based health IQ (ρ=0.56), followed by the other design factor—aesthetics (ρ=0.49)—and a content factor—ease of understanding (ρ=0.49). Four other factors—source trustworthiness (ρ=0.28), health knowledge (ρ=0.15), internet experience (ρ=0.13), and social endorsement (ρ=0.10)—showed significant but weak relationships with web-based health IQ.

On the basis of the Begg test, which takes into account publication bias (Begg *P*=.02), and using the trim-and-fill method [50] with 10 imputed studies on the right side of the funnel plot, age had a significant association with web-based health IQ (ρ=0.27; 95% CI 0.06-0.48; Q=3753.86). Thus, the age and web-based health IQ relationship changed from nonsignificant to significant, with individuals who were older rating the web-based health IQ higher than those who were younger. The remaining factors were not significantly related to web-based health IQ.

Regarding consequences, the web-based health IQ exerted the strongest effect on intentions to use health information systems (ρ=0.58). Its relationship with intentions to use health information was also significant but not as strong (ρ=0.37). Web-based health IQ's relationships with cognitive appraisal factors were mostly moderate, with the effect size for satisfaction being the largest (ρ=0.46). Web-based health IQ was moderately related to health information seeking (ρ=0.30) and did not have a significant relationship with health information use.

Across the results of the main effects, Q statistics were *substantial*, indicating that the effect size distribution was heterogeneous and that some variables other than subject-level sampling and measurement errors contributed to the effect size variances [51]. Confirming the Q statistics, the I^2 statistics indicated wide dispersion. The credibility interval for all relationships was wide, further implying that the effect size distribution was heterogeneous.

Table 2. Antecedents and consequences of web-based health information quality with at least 6 observations.

Variable	Definition
Antecedents	
Individual differences	
Gender	The gender of the participants included in study samples (female=1 and male=0)
Age	The age of the participants included in study samples
Education	The education levels of the participants included in study samples
Income	The income of the participants included in study samples
Race	The race of the participants included in study samples (White=1 and non-White=0)
Internet experience	An individual's experience with using the internet, as manifested in aspects such as the length or frequency of use and the use of the range of web-based services [52]
Personal involvement	An individual's perceived personal relevance of the web-based health information [53]
Perceived health status	Individuals' self-assessment of the status of their overall personal physical and mental health [54,55]
Condition experience and beliefs	An individual's experience with a health condition, perceived risk for developing the condition, and perceived severity of the condition
Health literacy	Individuals' ability to obtain health information from both electronic and nonelectronic sources and their ability to process, understand, and apply the obtained health information to solve health problems and make appropriate health decisions [56,57]
Health knowledge	Individuals' knowledge about their health problems and the care for the problems [58]
Source-related factors	
Source trustworthiness	The extent to which an individual believes that a specific web-based health information provider has attributes (eg, reputation) that are beneficial to the consumer [14]
Source expertise	The extent to which the source or the author of a message, webpage, or website is perceived to be capable of making correct assertions [59]
Content factors	
Ease of understanding	Whether the provided information is easy to understand (eg, in everyday language) and informative to users [60,61]
Social endorsement	Endorsements from other users of a website and could be manifested in forms ranging from sharing, commenting, and rating to liking [62]
Content comprehensiveness	Whether information provided is comprehensive, providing users with comparatively complete information (eg, necessary information to establish a medical claim, statistics, references, testimonials, source and author information, and user support information) [63,64]
Design factors	
Navigability	Whether a website has clear navigation menus and effective hyperlinks and whether the information is easy to access by searching or browsing [65,66]
Aesthetics	The visual design of a website, including the structural features such as typography, images, color, and aesthetics (eg, whether the website is professional and appealing) [67]
Consequences	
Cognitive appraisals	
Attitudes	Individuals' evaluations of and feelings about health websites or web-based health information [64,68]
Perceived usefulness	The degree to which consumers believe that using health information on the internet would enhance their health-related activities [47]
Perceived health benefits	The perceived level of rewards or risks that people have about the consequences of using or acting on web-based health information [14,65]
Satisfaction with health information	Individuals' satisfaction with health websites or web-based health information
Behavioral intentions	
Intentions to use health information	Individuals' intentions or willingness to use web-based health information to make health decisions, manage health problems, or inform health behaviors

Variable	Definition
Intentions to use health information systems	Individuals' intentions or willingness to use web-based health information systems to seek health information
Behavior	
Health information seeking	Individuals' use of web-based or offline sources to find health-related information, which is manifested in aspects such as the types of information sought and the frequency and intensity of health information seeking
Use health information	Individuals' use or application of health information (from web-based or offline sources) to make health decisions, manage health problems, or inform health behaviors

Table 3. Antecedents and consequences of quality of web-based health information.

Factors	Samples, n	Sample size, N	r[a]	ρ[b], mean (SD)	95% CI	90% CV[c]	Q[d]	I^2 (%)[e]	Begg *P* value[f]
Antecedents									
Individual differences									
Gender (female)	25	20,101	0.04	0.04 (0.13)	–0.04 to 0.13	–0.17 to 0.25	216.29[g]	86.45	.22
Age	20	23,463	0.04	0.04 (0.13)	–0.11 to 0.20	–0.18 to 0.26	834.84[g]	97.15	.02
Education	15	16,874	0.05	0.05 (0.17)	–0.09 to 0.20	–0.23 to 0.33	332.32[g]	94.29	.70
Internet experience	14	6235	0.12	0.13 (0.15)	0.01 to 0.24	–0.11 to 0.37	186.18[g]	91.99	.75
Personal involvement	13	4171	0.11	0.13 (0.22)	–0.05 to 0.31	–0.23 to 0.49	465.37[g]	97.08	.31
Perceived health status	9	26,207	0.04	0.04 (0.21)	–0.12 to 0.19	–0.31 to 0.39	715.44[g]	98.55	.92
Income	9	18,177	0.05	0.05 (0.10)	–0.08 to 0.19	–0.11 to 0.21	270.04[g]	95.65	.61
Race (White)	9	14,162	–0.07	–0.08 (0.26)	–0.52 to 0.37	–0.51 to 0.35	2609.14[g]	99.56	.08
Condition experience and beliefs	7	7772	0.05	0.05 (0.10)	–0.01 to 0.11	–0.12 to 0.22	18.74[g]	48.88	.99
Health literacy	6	3661	0.18	0.22 (0.28)	–0.01 to 0.45	–0.25 to 0.69	298.22[g]	97.83	.27
Health knowledge	6	2797	0.13	0.15 (0.11)	0.07 to 0.22	–0.03 to 0.33	18.26[g]	63.64	.72
Source-related factors									
Source trustworthiness	17	4154	0.25	0.28 (0.25)	0.10 to 0.45	–0.14 to 0.70	950.38[g]	97.78	.66
Source expertise	13	5988	0.17	0.20 (0.27)	–0.08 to 0.49	–0.25 to 0.65	649.35[g]	97.66	.44
Content-related factors									
Ease of understanding	14	3981	0.41	0.49 (0.28)	0.35 to 0.63	0.03 to 0.95	698.28[g]	97.90	.47
Social endorsement	7	2267	0.09	0.10 (0.17)	0.00 to 0.19	–0.18 to 0.38	32.03[g]	77.57	.14
Content comprehensiveness	7	1373	0.18	0.21 (0.30)	–0.11 to 0.54	–0.29 to 0.71	549.04[g]	97.57	.56
Design-related factors									
Navigability	12	3099	0.47	0.56 (0.33)	0.44 to 0.67	0.02 to 1.00	436.08[g]	96.85	.21
Aesthetics	11	4307	0.40	0.49 (0.26)	0.30 to 0.68	0.06 to 0.92	434.41[g]	95.99	.06
Consequences									
Cognitive appraisals									
Attitudes	14	3934	0.38	0.43 (0.24)	0.36 to 0.50	0.04 to 0.82	1871.08[g]	88.04	.75
Perceived usefulness	10	11,110	0.25	0.29 (0.23)	0.08 to 0.50	–0.09 to 0.67	557.55[g]	97.78	.60
Perceived health benefits	9	6292	0.32	0.37 (0.16)	0.26 to 0.49	0.10 to 0.64	189.17[g]	94.86	.61

Factors	Samples, n	Sample size, N	r[a]	ρ[b], mean (SD)	95% CI	90% CV[c]	Q[d]	I^2 (%)[e]	Begg *P* value[f]
Satisfaction with health information	9	3334	0.41	0.46 (0.29)	0.41 to 0.51	–0.01 to 0.93	1699.85[g]	84.31	.36
Behavioral intentions									
Intentions to use health information	17	7663	0.32	0.37 (0.33)	0.32 to 0.43	–0.18 to 0.92	4162.85[g]	99.23	.66
Intentions to use health information systems	8	1614	0.49	0.58 (0.24)	0.43 to 0.72	0.19 to 0.97	290.34[g]	95.73	.55
Behavior									
Health information seeking	18	26,259	0.25	0.30 (0.28)	0.15 to 0.46	–0.16 to 0.76	12,308.96[g]	99.59	.08
Health information use	15	15,021	0.21	0.25 (0.28)	–0.00 to 0.50	–0.21 to 0.71	1083.33[g]	98.36	.77

[a]Weighted mean correlation.

[b]Corrected weighted mean correlation and SD of ρ.

[c]90% credibility interval.

[d]Heterogeneity statistic.

[e]Percentage of variation across studies that is because of heterogeneity.

[f]The Begg test for funnel plot asymmetry.

[g]*P*<.01.

Moderators of Web-Based Health IQ Relationships

Substantial heterogeneity calls for moderator analyses to explain the variance. The examined moderators included culture, research design, publication factors, and one operationalization-related moderator—the focal variable. The analysis was performed on web-based health IQ's relationships with eight factors: two individual factors—gender and age—enabled by adequate sample numbers and six composite factors—source, content, design, cognitive appraisals, behavioral intentions, and behavior—formed by combining lower-level antecedents and consequences to offer adequate observations for the analysis. For the moderator analysis involving age and web-based health IQ, we did not include the 10 imputed studies, given that incorporating simulated data can distort the subgroup comparison. Table 4 presents the subgroup mean values and Q_M statistics. Other relevant statistics (95% CI, 90% credibility interval, Q_E, and R^2) can be found in Multimedia Appendices 3-10. All moderators were significantly related to the effect size of at least one web-based health IQ relationship examined; 6 moderators significantly affected ≥3 relationships. The following interpretations focused on subgroups with significant differences.

Culture moderated the three antecedents of web-based health IQ: gender, age, and source. Females in individualistic (ρ=0.06 vs ρ=–0.11), low power distance (ρ=0.06 vs ρ=–0.02), and high uncertainty avoidance (ρ=0.08 vs ρ=–0.03) cultures rated web-based health IQ higher than males. Older individuals in low uncertainty avoidance (ρ=0.21 vs ρ=–0.07) and indulgence cultures (ρ=0.19 vs ρ=–0.05) rated web-based health IQ higher. Individuals with high uncertainty avoidance (ρ=0.37 vs ρ=0.20), long-term orientation (ρ=0.32 vs ρ=0.19), and restraint (ρ=0.37 vs ρ=0.17) cultures exhibited a stronger source and web-based health IQ relationship.

Culture moderated two consequences of web-based health IQ: cognitive appraisals and behavioral intentions. Individuals in long-term cultures had higher cognitive appraisals of web-based health IQ (ρ=0.40 vs ρ=0.27). Individuals with low uncertainty avoidance (ρ=0.59 vs ρ=0.41), short-term orientation (ρ=0.80 vs ρ=0.43), and indulgence cultures (ρ=0.60 vs ρ=0.43) had higher behavioral intentions as a result of the web-based health IQ than individuals in their respective counterpart cultures.

Research design moderated two antecedents of web-based health IQ: gender and content. Women rated the web-based health IQ higher in studies using the survey method (ρ=0.06 vs ρ=–0.06), non–social media technology context (ρ=0.06 vs ρ=–0.10), and nonpatient samples (ρ=0.09 vs 0.00). The content and web-based health IQ relationships were stronger in studies using the survey method (ρ=0.51 vs ρ=0.21), general stimuli (ρ=0.73 vs ρ=0.30), and nonstudent samples (ρ=0.43 vs ρ=0.24).

Research design moderated three consequences of web-based health IQ: cognitive appraisals, behavioral intentions, and behavior. Studies using specific stimuli (ρ=0.48 vs ρ=0.27) and nonstudent samples (ρ=0.35 vs ρ=0.26) produced larger effect sizes for the web-based health IQ and cognitive appraisals relationship. Studies using specific stimuli (ρ=0.54 vs ρ=0.32), social media context (ρ=0.45 vs ρ=0.39), and student samples (ρ=0.53 vs ρ=0.39) reported higher behavioral intentions. Student samples also produced a larger effect size for the web-based health IQ and behavior relationship (ρ=0.66 vs ρ=0.20).

Publication factors moderated the gender and web-based health IQ relationship. Journal articles (ρ=0.06 vs ρ=–0.08) and papers published before 2014 (ρ=0.08 vs ρ=–0.01) reported larger

effect sizes than their respective counterparts. Publication year moderated web-based health IQ and cognitive appraisals and web-based health IQ and behavioral intentions, with recent publications (2014 and after) reporting lower cognitive appraisals (ρ=0.32 vs ρ=0.37) but higher behavioral intentions (ρ=0.55 vs ρ=0.25).

The three focal variables—quality, credibility, and trust—produced significant differences in 2 web-based health IQ relationships. The quality subgroup reported a stronger design and web-based health IQ relationship than the credibility subgroup (ρ=0.58 vs ρ=0.33). The omnibus test for comparing the focal variables in the web-based health IQ and behavioral intentions was significant (Q_M=30.50; *P*<.01). Post hoc tests revealed that the significant difference was because of the trust group being higher than the quality group (Q_M=13.63; *P*<.01) and the trust group being higher than the credibility group (Q_M=26.85; *P*<.01).

Table 4. Influence of moderators on quality of web-based health information relationships.

Moderators	Gender (female=1)	Age	Source-related factors	Content-related factors	Design-related factors	Cognitive appraisals	Behavioral intentions	Behavior
Culture								
Individualism versus collectivism, Q_M[a]	10.50[b]	0.32	3.20	0.03	0.24	1.88	2.71	1.67
Individualism, mean (k; N)[c]	0.06 (18; 17,745)	0.05 (12; 19,834)	0.22 (21; 7645)	0.37 (20; 5491)	0.39 (9; 2127)	0.31 (23; 15,533)	0.43 (13; 4684)	0.27 (20; 33,688)
Collectivism, mean (k; N)	–0.11 (4; 1242)	0.03 (5; 2542)	0.44 (6; 1460)	0.46 (4; 468)	0.57 (6; 1917)	0.39 (11; 4008)	0.58 (7; 1977)	0.37 (9; 5515)
Power distance, Q_M	3.85[d]	0.07	3.20	0.03	0.24	1.88	2.71	1.67
High, mean (k; N)	–0.02 (5; 1901)	0.01 (6; 3201)	0.44 (6; 1460)	0.46 (4; 468)	0.57 (6; 1917)	0.39 (11; 4008)	0.58 (7; 1977)	0.37 (9; 5515)
Low, mean (k; N)	0.06 (17; 17,086)	0.05 (11; 19,175)	0.22 (21; 7645)	0.37 (20; 5491)	0.39 (9; 2127)	0.31 (23; 15,533)	0.43 (13; 4684)	0.27 (20; 33,688)
Uncertainty avoidance, Q_M	6.78[b]	7.37[b]	6.25[d]	0.13	0.25	1.23	109.01[b]	2.02
High, mean (k; N)	0.08 (10; 13,180)	–0.07 (8; 12,795)	0.37 (14; 3022)	0.45 (11; 2177)	0.61 (4; 969)	0.31 (17; 7284)	0.41 (9; 4344)	0.42 (12; 14,139)
Low, mean (k; N)	–0.03 (12; 5807)	0.21 (9; 9581)	0.20 (13; 6083)	0.33 (13; 3782)	0.43 (11; 3075)	0.33 (17; 12,257)	0.59 (11; 2317)	0.21 (17; 25,064)
Orientation, Q_M	0.57	2.93	7.21[b]	0.52	0.02	3.88[d]	457.96[b]	.08
Long-term, mean (k; N)	0.07 (15; 14,013)	–0.05 (13; 14,411)	0.32 (19; 4457)	0.48 (15; 3551)	0.52 (9; 2556)	0.40 (23; 7935)	0.43 (15; 5761)	0.29 (17; 13,738)
Short-term, mean (k; N)	–0.01 (7; 4974)	0.22 (4; 7965)	0.19 (8; 4648)	0.22 (9; 2408)	0.39 (6; 1488)	0.27 (11; 11,606)	0.80 (5; 900)	0.28 (12; 25,465)
Indulgence versus restraint, Q_M	0.00	3.86[d]	11.47[b]	0.13	0.04	0.37	165.28[b]	0.92
Indulgence, mean (k; N)	0.00 (10; 6213)	0.19 (6; 8974)	0.17 (9; 5223)	0.34 (11; 3582)	0.41 (7; 1862)	0.31 (13; 11,190)	0.60 (8; 1704)	0.20 (14; 22,776)
Restraint, mean (k; N)	0.07 (12; 12,774)	–0.05 (11; 13,402)	0.37 (18; 3882)	0.43 (12; 2307)	0.53 (8; 2182)	0.38 (20; 6945)	0.43 (12; 4957)	0.29 (12; 11,502)
Methods								
Study method, Q_M	5.39[d]	0.01	0.57	9.55[b]	2.46	0.49	3.39	N/A[e]
Survey, mean (k; N)	0.06 (17; 17,929)	0.04 (13; 21,681)	0.25 (7; 4884)	0.51 (10; 3271)	0.56 (14; 5795)	0.34 (27; 21,314)	0.40 (22; 8870)	—[f]
Experiment, mean (k; N)	–0.06 (8; 2172)	–0.01 (7; 1782)	0.22 (23; 5258)	0.21 (17; 3772)	0.26 (6; 897)	0.41 (10; 1702)	0.46 (3; 407)	N/A
Stimulus type, Q_M	2.13	0.05	1.31	9.53[b]	0.49	4.74[a]	100.36[b]	1.04
General, mean (k; N)	0.06 (13; 16,088)	0.05 (11; 20,648)	0.35 (4; 1321)	0.73 (5; 822)	0.60 (9; 3963)	0.27 (18; 15,477)	0.32 (11; 5688)	0.29 (25; 35,528)
Specific, mean (k; N)	–0.01 (12; 4013)	–0.02 (9; 2815)	0.22 (26; 8821)	0.30 (22; 6221)	0.39 (11; 2729)	0.48 (19; 7539)	0.54 (14; 3589)	0.27 (6; 3983)
Technology context, Q_M	4.55[d]	0.06	3.33	0.72	N/A	N/A	5.18[d]	N/A
Social media, mean (k; N)	–0.10 (4; 2092)	0.02 (3; 1397)	0.32 (10; 2011)	0.24 (10; 2493)	—	—	0.45 (6; 1107)	—
Non–social media, mean (k; N)	0.06 (21; 18,009)	0.04 (17; 22,066)	0.21 (20; 8131)	0.41 (17; 4550)	N/A	N/A	0.39 (19; 8170)	N/A

Moderators	Gender (female=1)	Age	Source-related factors	Content-related factors	Design-related factors	Cognitive appraisals	Behavioral intentions	Behavior
Sample clinical status, Q_M	7.18[b]	N/A	N/A	N/A	N/A	N/A	0.57	0.04
Nonpatients, mean (k; N)	0.09 (4; 9478)	—	—	—	—	—	0.24 (3; 1646)	0.11 (6; 4192)
Patients, mean (k; N)	0.00 (21; 10,623)	N/A	N/A	N/A	N/A	N/A	0.44 (22; 7631)	0.31 (25; 35,319)
Sample type, Q_M	0.00	0.03	0.00	5.63[d]	.04	9.20[b]	4.69[d]	3.71[d]
Students, mean (k; N)	0.01 (7; 1978)	–0.03 (3; 594)	0.22 (13; 2959)	0.24 (13; 2947)	0.50 (4; 841)	0.26 (8; 1693)	0.53 (6; 1133)	0.66 (9; 7008)
Nonstudents, mean (k; N)	0.05 (18; 18,123)	0.04 (17; 22,869)	0.24 (17; 7183)	0.43 (14; 4096)	0.52 (16; 5851)	0.35 (29; 21,323)	0.39 (19; 8144)	0.20 (22; 32,503)
Publication								
Outlet, Q_M	6.41[d]	0.02	0.18	0.73	2.85	0.41	N/A	0.01
Journal, mean (k; N)	0.06 (17; 18,204)	0.05 (12; 21,288)	0.24 (20; 7528)	0.38 (19; 5426)	0.50 (13; 4751)	0.33 (30; 21,454)	—	0.29 (26; 38,726)
Nonjournal, mean (k; N)	–0.08 (8; 1897)	–0.01 (8; 2175)	0.21 (10; 2614)	0.26 (8; 1617)	0.54 (7; 1941)	0.45 (7; 1562)	N/A	0.22 (5; 785)
Year, Q_M	5.26[d]	0.15	0.00	0.01	0.21	6.42[d]	146.34[b]	0.00
Before 2014, mean (k; N)	0.08 (9; 12,145)	0.07 (7; 16,892)	0.26 (13; 5463)	0.33 (8; 1684)	0.52 (12; 4380)	0.37 (21; 14,357)	0.25 (10; 4430)	0.27 (15; 24,416)
2014 and after, mean (k; N)	–0.01 (16; 7956)	–0.02 (13; 6571)	0.20 (17; 4679)	0.36 (19; 5359)	0.52 (8; 2312)	0.32 (16; 8659)	0.55 (15; 4847)	0.31 (16; 15,095)
Focal variable, Q_M	0.07	0.94	0.09	2.19	9.34[b]	3.59	30.50[b]	4.12
Quality, mean (k; N)	—	0.04 (3; 737)	0.30 (8; 1899)	0.31 (6; 1632)	0.58 (10; 4815)	0.45 (11; 4510)	0.40 (7; 3355)	0.40 (6; 4095)
Credibility, mean (k; N)	0.02 (9; 2867)	0.22 (8; 7801)	0.20 (18; 4859)	0.29 (16; 3943)	0.33 (9; 1704)	0.26 (17; 9301)	0.31 (8; 1577)	0.19 (7; 7341)
Trust, mean (k; N)	0.06 (9; 14,572)	–0.05 (8; 14,290)	0.25 (4; 3384)	0.57 (5; 1468)	—	0.36 (9; 9205)	0.45 (10; 4345)	0.30 (17; 27,889)

[a]Omnibus test comparing group means.

[b]$P<.01$.

[c]Cell entries show subgroup means (weighted mean correlation corrected for measurement unreliability); each parenthesis contains *k* (number of samples) and N (total sample size).

[d]$P<.05$.

[e]N/A: not applicable; insufficient effect sizes for subgroup comparison.

[f]Not available.

Discussion

Using a comprehensive meta-analytic approach, this study analyzed antecedents and consequences of consumer web-based health IQ evaluations and contextual factors that moderate the relationships based on 147 independent studies. The major findings are discussed in the following sections.

Web-Based Health IQ Antecedents

Consistent with systematic reviews of consumer web-based health information evaluation behavior [18,19], we identified four major categories of antecedents of web-based health IQ: individual, source, content, and design factors. Furthermore, we revealed the magnitude of the antecedents' effect. We found that among the 18 antecedents examined, navigability (design) was the strongest predictor of web-based health IQ, followed by ease of understanding (content) and aesthetics (design). Four factors had significant but weak relationships with web-based health IQ: source trustworthiness (source), health knowledge (individual), internet experience (individual), and social endorsement (content). Age (individual) was significantly related to web-based health IQ after correcting for publication bias. However, this result needs to be viewed with caution as imputed data were generated to obtain this result. The remaining 10 antecedents were not substantially related to web-based health IQ evaluation.

These results suggest that consumers rely prominently on peripheral cues (eg, navigability, aesthetics, and ease of understanding) and less on systematic cues (eg, content comprehensiveness) to evaluate web-based health IQ. This is consistent with the Fogg et al [69,70] findings from large-scale surveys that website design look and ease of use (including navigability) were the most prominent influencers of website credibility, exerting stronger impacts than source expertise and trustworthiness. According to dual processing models of information processing and assessment, such as the Elaboration Likelihood Model and Heuristic-Systematic Model [71,72], these results can be attributed to consumers' lack of motivation and/or ability to evaluate web-based health IQ [73,74]. However, the results were not conclusive. First, it is possible that theoretically significant motivational and ability factors, such as personal involvement and source expertise [71,72], did not show a significant direct impact on web-based health IQ in this research because their relationships were moderated by contextual factors, which were not analyzed because of insufficient observations. Second, other theoretically and/or empirically significant influencers of web-based health IQ that are closely related to systematic information processing, such as augment strength [14] and content consistency [75,76], were not analyzed because of insufficient observations; thus, their effects were not accounted for in this research. More research is needed to elucidate the antecedent and web-based health IQ relationships.

Consequences

Web-based health IQ was significantly related to all the consequences identified in the research, except for health information use. The effect of web-based health IQ on behavioral intentions (particularly intentions to use health information systems) was the strongest, followed by cognitive appraisal factors (particularly satisfaction with health information). The relationship of web-based health IQ with health information–seeking behaviors was moderate, consistent with the findings of another meta-analysis of credibility and health information seeking [77].

The information system success model posits that IQ predicts users' intention to use or use of and satisfaction with an information system [78,79]. The model of information adoption posits that IQ determines users' attitudes toward information (ie, usefulness) [80,81]. Empirical research in information systems has provided strong support for the IQ-satisfaction relationship [40], whereas support for the IQ-use relationship has been mixed [79]. Our meta-analyses of web-based health IQ consequences are largely consistent with these findings, suggesting that web-based health IQ is important for consumers' intentions to use and satisfaction with web-based health information systems and information and information-seeking behavior. The 2 aforementioned models, although primarily developed and tested in organizational or individual work settings, are applicable in the context of consumers' web-based health information seeking.

Moderators

Culture moderated three antecedent and web-based health IQ relationships (ie, age, gender, and source) and two web-based health IQ and consequence relationships (ie, cognitive appraisals and behavioral intentions), demonstrating itself as an important factor shaping both web-based health IQ evaluation and its consequences. However, few empirical studies have directly examined the culture and web-based health IQ relationships. Future studies should fill this gap, which is critical for informing the design of health information systems and policies that serve different cultural groups in and across nations.

Research design factors moderated two antecedent and web-based health IQ (ie, gender and content) and all 3 consequence and web-based health IQ relationships, reinforcing the importance of careful research design in studying web-based health IQ. It is worth noting that sample type and stimulus type affected the greatest number of relationships, with student samples and studies using general stimuli reporting stronger content and web-based health IQ relationships and having lower cognitive appraisals but stronger behavioral intentions (and stronger behavior for the student samples). The results caution the use of student samples and general stimuli when studying web-based health IQ relationships. The clinical status of the sample moderated the gender and web-based health IQ relationship. It may moderate more relationships for patients' personal involvement [72]; however, it remains inconclusive because of insufficient observations.

Limited publication venue bias was observed as the publication outlet moderated only the gender and web-based health IQ relationship. As a proxy to detect how web-based health IQ relationships have fluctuated over time, the publication year moderated three relationships—gender and web-based health IQ, web-based health IQ and cognitive appraisals, and web-based health IQ and behavioral intentions—revealing that individuals' cognitive appraisals of web-based health IQ lessened; however, intentions to act on the information increased over time. It is plausible that consumers are becoming more critical as arbiters of web-based health information; however, they are also becoming more receptive to web-based health information and information systems.

The focal variables (credibility, trust, and quality) moderated two relationships—design and web-based health IQ and web-based health IQ and behavioral intentions—out of the 8 relationships examined, indicating that some theoretical and/or methodological issues exist that promulgate this effect size disparity. Studies using quality identified a larger effect size than studies that used credibility in the design and web-based health IQ relationship. This can be attributed to the fact that studies that examined the relationship viewed quality as intrinsic merit of information (eg, accuracy, argument strength, consistency, and comprehensiveness) [14,82] and credibility as perceived reliability or trustworthiness of information [64]. In such a case, we speculate that consumers had more difficulty determining IQ than credibility [83]; thus, they need to rely more on design factors to form IQ perceptions. For the web-based health IQ and behavioral intentions relationship, studies using trust produced the largest effect size, followed by studies using quality and credibility, indicating that trust most strongly predicts behavioral intentions, followed by quality and credibility. This may be because studies on web-based health IQ and behavioral intentions were more likely to consider risk

and gain assessment as part of the trust formation process [84-86], such that trust showed a higher predictive power for behavioral intentions [12,14].

Limitations

As with all meta-analysis studies, the main effects of a small number of observations or small sample size (eg, race or health knowledge with web-based health IQ relationships) should be interpreted with caution. Insufficient observations also limit moderator analyses, whereby moderator analyses of some theoretically or practically important relationships (eg, race, personal involvement, and health literacy with web-based health IQ relationships) were not performed. Relatedly, some antecedents and consequences were combined to form high-level categories to enable moderator analyses, which inevitably masks how some important specific relationships (eg, web-based health IQ and use of health information) might be affected by moderators.

In terms of moderator analysis, consistent with prior meta-analysis findings, student-based results were biased [79], and survey-based results produced larger effect sizes than experience-based results [41]. The most noteworthy finding concerning moderator analyses was that the three conceptualizations of web-based health IQ (ie, quality, credibility, and trust) moderated two out of the eight relationships examined (ie, design and web-based health IQ and web-based health IQ and behavior intentions), suggesting that despite a significant conceptual overlap, theoretical and/or operationalization differences exist among the 3 constructs. This result should be interpreted in light of the fact that we took a phenomenological approach, adopting the authors' conceptualizations of web-based health IQ (quality, credibility, and trust). A detailed examination of the definitions and measures of these constructs is warranted to elucidate the differences among the concepts. A preliminary examination of the included papers revealed that not many studies provided explicit definitions of the constructs and that measures of the same construct varied, with many articles not including specific and complete measures. These observations call for future empirical studies to offer clearer definitions of the constructs and complete measures to enable a fair assessment of these concepts for future literature synthesis.

Despite attempts to apply various moderators to explain the variance across web-based health IQ relationships, substantial variance remained. Future research should prudently select additional moderators to explain this variance. For example, health topics merit investigation as an important contextual factor with theoretical significance for studying information-seeking behavior [87,88]. Website type also merits investigation in light of recent findings that it influences how consumers apply content, design, and source factors to evaluate web-based health IQ [89,90].

Conclusions

On the basis of a meta-analysis of 147 empirical studies, our study confirmed that consumers' evaluation of web-based health IQ significantly affects their cognitive appraisals of web-based health information, intentions to use web-based information systems and information, and information-seeking behavior, suggesting the important role that web-based health IQ plays in promoting health information seeking. The study also confirmed that consumers' evaluation of web-based health IQ is shaped by source, content, design, and individual factors, with the most influential factors being design, particularly navigability and aesthetics, and ease of understanding of content. Many individual factors, such as gender, race, education, personal involvement, and health literacy, did not show significant relationships with web-based health IQ. However, moderator analyses and the residual variance after the analyses suggest that these relationships may be moderated by numerous methodological and nonmethodological moderators. Patient empowerment and active participation in health care require individuals to have equal access to high-quality health information. More studies are needed to elucidate individual factors and web-based health IQ relationships to address potential information access disparities among different user groups.

Acknowledgments

The authors would like to thank Stephanie Tomasky and Shijie Song for their assistance in the article search and screening processes.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of studies included in sample.
[DOCX File , 59 KB - jmir_v24i4e36463_app1.docx]

Multimedia Appendix 2
Mean reliabilities.
[DOCX File , 13 KB - jmir_v24i4e36463_app2.docx]

Multimedia Appendix 3

Influence of moderators on the relationship between gender and web-based health information quality.
[DOCX File , 18 KB - jmir_v24i4e36463_app3.docx]

Multimedia Appendix 4
Influence of moderators on the relationship between age and web-based health information quality.
[DOCX File , 18 KB - jmir_v24i4e36463_app4.docx]

Multimedia Appendix 5
Influence of moderators on the relationship between source-related factors and web-based health information quality.
[DOCX File , 18 KB - jmir_v24i4e36463_app5.docx]

Multimedia Appendix 6
Influence of moderators on the relationship between content-related factors and web-based health information quality.
[DOCX File , 18 KB - jmir_v24i4e36463_app6.docx]

Multimedia Appendix 7
Influence of moderators on the relationship between design-related factors and web-based health information quality.
[DOCX File , 17 KB - jmir_v24i4e36463_app7.docx]

Multimedia Appendix 8
Influence of moderators on the relationship between web-based health information quality and cognitive appraisals.
[DOCX File , 18 KB - jmir_v24i4e36463_app8.docx]

Multimedia Appendix 9
Influence of moderators on the relationship between web-based health information quality and behavioral intentions.
[DOCX File , 18 KB - jmir_v24i4e36463_app9.docx]

Multimedia Appendix 10
Influence of moderators on the relationship between web-based health information quality and behavior.
[DOCX File , 18 KB - jmir_v24i4e36463_app10.docx]

References

1. Eysenbach G, Powell J, Kuss O, Sa ER. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002 May;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
2. 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]
3. Tsunehara ST, Gaster B. A critical assessment of diabetes information on the internet. Diabetes Educ 2002;28(2):186-195. [doi: 10.1177/014572170202800206]
4. Eppler MJ. Managing Information Quality: Increasing the Value of Information in Knowledge-intensive Products and Processes. 2nd edition. Berlin, Germany: Springer; 2006.
5. 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]
6. Taylor RS. Value-Added Processes in Information Systems. Norwood, NJ, USA: Ablex Publishing Corporation; 1986.
7. McKinney V, Yoon K, Zahedi FM. The measurement of web-customer satisfaction: an expectation and disconfirmation approach. Inf Syst Res 2002 Sep 1;13(3):296-315. [doi: 10.1287/isre.13.3.296.76]
8. Wang RY, Strong DM. Beyond accuracy: what data quality means to data consumers. J Manag Inf Syst 1996;12(4):5-33. [doi: 10.1080/07421222.1996.11518099]
9. Tseng S, Fogg BJ. Credibility and computing technology. Commun ACM 1999 May;42(5):39-44. [doi: 10.1145/301353.301402]
10. Sundar SS. Exploring receivers' criteria for perception of print and online news. Journal Mass Commun Q 1999;76(2):373-386. [doi: 10.1177/107769909907600213]
11. Rieh SY, Danielson DR. Credibility: a multidisciplinary framework. Ann Rev Info Sci Tech 2007;41(1):307-364. [doi: 10.1002/aris.2007.1440410114]
12. Harris PR, Sillence E, Briggs P. Perceived threat and corroboration: key factors that improve a predictive model of trust in internet-based health information and advice. J Med Internet Res 2011 Jul 27;13(3):e51 [FREE Full text] [doi: 10.2196/jmir.1821] [Medline: 21795237]

13. Song J, Zahedi FM. Trust in health infomediaries. Decis Support Syst 2007 Mar;43(2):390-407. [doi: 10.1016/j.dss.2006.11.011]
14. Yi MY, Yoon JJ, Davis JM, Lee T. Untangling the antecedents of initial trust in Web-based health information: the roles of argument quality, source expertise, and user perceptions of information quality and risk. Decis Support Syst 2013 Apr;55(1):284-295. [doi: 10.1016/j.dss.2013.01.029]
15. Corritore CL, Wiedenbeck S, Kracher B, Marble RP. Online trust and health information websites. Int J Technol Hum Interact 2012 Oct;8(4):92-115. [doi: 10.4018/jthi.2012100106]
16. Freeman KS, Spyridakis JH. An examination of factors that affect the credibility of online health information. Tech Commun 2004 Jun;51(2):239-263 [FREE Full text]
17. 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]
18. 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]
19. Sun Y, Zhang Y, Gwizdka J, Trace CB. Consumer evaluation of the quality of online health information: systematic literature review of relevant criteria and indicators. J Med Internet Res 2019 May 02;21(5):e12522 [FREE Full text] [doi: 10.2196/12522] [Medline: 31045507]
20. Kim Y. Trust in health information websites: a systematic literature review on the antecedents of trust. Health Informatics J 2016 Jun;22(2):355-369 [FREE Full text] [doi: 10.1177/1460458214559432] [Medline: 25518944]
21. Helft PR, Eckles RE, Johnson-Calley CS, Daugherty CK. Use of the internet to obtain cancer information among cancer patients at an urban county hospital. J Clin Oncol 2005 Aug 01;23(22):4954-4962. [doi: 10.1200/JCO.2005.09.621] [Medline: 16051948]
22. Hesse BW, Nelson DE, Kreps GL, Croyle RT, Arora NK, Rimer BK, et al. Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. Arch Intern Med 2005 Dec;165(22):2618-2624. [doi: 10.1001/archinte.165.22.2618] [Medline: 16344419]
23. Marrie RA, Salter AR, Tyry T, Fox RJ, Cutter GR. Preferred sources of health information in persons with multiple sclerosis: degree of trust and information sought. J Med Internet Res 2013 Mar 17;15(4):e67 [FREE Full text] [doi: 10.2196/jmir.2466] [Medline: 23635393]
24. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan 27;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
25. Bernhardt JM, Felter EM. Online pediatric information seeking among mothers of young children: results from a qualitative study using focus groups. J Med Internet Res 2004 Mar 01;6(1):e7 [FREE Full text] [doi: 10.2196/jmir.6.1.e7] [Medline: 15111273]
26. Lawson R, Forbes S, Williams J. Patterns of trust in sources of health information. N Z Med J 2011 Jan 21;124(1328):98-105. [Medline: 21475344]
27. Nguyen GT, Bellamy SL. Cancer information seeking preferences and experiences: disparities between Asian Americans and Whites in the Health Information National Trends Survey (HINTS). J Health Commun 2006;11 Suppl 1:173-180. [doi: 10.1080/10810730600639620] [Medline: 16641082]
28. Murray E, Lo B, Pollack L, Donelan K, Catania J, White M, et al. The impact of health information on the internet on the physician-patient relationship: patient perceptions. Arch Intern Med 2003 Jul 28;163(14):1727-1734. [doi: 10.1001/archinte.163.14.1727] [Medline: 12885689]
29. Yan YY. Online health information seeking behavior in Hong Kong: an exploratory study. J Med Syst 2010 Apr;34(2):147-153. [doi: 10.1007/s10916-008-9226-9] [Medline: 20433053]
30. Ye Y. Correlates of consumer trust in online health information: findings from the health information national trends survey. J Health Commun 2011 Jan;16(1):34-49. [doi: 10.1080/10810730.2010.529491] [Medline: 21086209]
31. Beldad A, de Jong M, Steehouder M. How shall I trust the faceless and the intangible? A literature review on the antecedents of online trust. Comput Human Behav 2010 Sep;26(5):857-869. [doi: 10.1016/j.chb.2010.03.013]
32. Hong W, Chan FK, Thong JY, Chasalow LC, Dhillon G. A framework and guidelines for context-specific theorizing in information systems research. Inf Syst Res 2014 Mar;25(1):111-136. [doi: 10.1287/isre.2013.0501]
33. Compare countries. Hofstede Insights. URL: https://www.hofstede-insights.com/fi/product/compare-countries/ [accessed 2022-02-27]
34. Feng H, Yang Y. A model of cancer-related health information seeking on the internet. China Media Res 2007 Jul;3(3):14-24.
35. McKinley CJ, Ruppel EK. Exploring how perceived threat and self-efficacy contribute to college students' use and perceptions of online mental health resources. Comput Human Behav 2014 May;34:101-109. [doi: 10.1016/j.chb.2014.01.038]
36. Hunter JE, Schmidt FL. Methods of Meta-Analysis: Correcting Error and Bias in Research Findings. 2nd edition. Thousand Oaks, CA, USA: SAGE Publications; Jun 3, 2004.
37. Peterson RA, Brown SP. On the use of beta coefficients in meta-analysis. J Appl Psychol 2005 Jan;90(1):175-181. [doi: 10.1037/0021-9010.90.1.175] [Medline: 15641898]

38. Bonett DG. Transforming odds ratios into correlations for meta-analytic research. Am Psychol 2007 Apr;62(3):254-255. [doi: 10.1037/0003-066X.62.3.254] [Medline: 17469909]
39. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics 1994 Dec;50(4):1088-1101. [Medline: 7786990]
40. Ghasemaghaei M, Hassanein K. Online information quality and consumer satisfaction: the moderating roles of contextual factors – a meta-analysis. Inf Manag 2015 Dec;52(8):965-981. [doi: 10.1016/j.im.2015.07.001]
41. Kim Y, Peterson RA. A meta-analysis of online trust relationships in e-commerce. J Interact Market 2017 May;38:44-54. [doi: 10.1016/j.intmar.2017.01.001]
42. Ma TJ, Atkin D. User generated content and credibility evaluation of online health information: a meta analytic study. Telematics Informatics 2017 Aug;34(5):472-486. [doi: 10.1016/j.tele.2016.09.009]
43. Treweek S, Oxman AD, Alderson P, Bossuyt PM, Brandt L, Brożek J, DECIDE Consortium. Developing and evaluating communication strategies to support informed decisions and practice based on evidence (DECIDE): protocol and preliminary results. Implement Sci 2013 Jan 09;8:6 [FREE Full text] [doi: 10.1186/1748-5908-8-6] [Medline: 23302501]
44. 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]
45. Hofstede G, Hofstede GJ, Minkov M. Cultures and Organizations: Software of the Mind. 3rd edition. New York, NY, USA: McGraw-Hill; May 3, 2010.
46. National culture. Hofstede Insights. URL: https://hi.hofstede-insights.com/national-culture [accessed 2021-05-03]
47. Yun EK, Park HA. Consumers' disease information-seeking behaviour on the Internet in Korea. J Clin Nurs 2010 Oct;19(19-20):2860-2868. [doi: 10.1111/j.1365-2702.2009.03187.x] [Medline: 20598000]
48. Hofstede G. Culture's consequences: comparing values, behaviors, institutions, and organizations across nations. 2nd edition. Thousand Oaks, CA, USA: Sage Publications; 2001.
49. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]
50. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 2000 Jun;56(2):455-463. [doi: 10.1111/j.0006-341x.2000.00455.x] [Medline: 10877304]
51. Lipsey MW, Wilson DB. Practical meta-analysis. Thousand Oaks, CA, USA: Sage Publications; 2001.
52. Rains SA. Seeking health information in the information age: the role of internet self-efficacy. Western J Commun 2008 Mar 11;72(1):1-18. [doi: 10.1080/10570310701827612]
53. Zaichkowsky JL. Conceptualizing involvement. J Advert 1986 Jun;15(2):4-34. [doi: 10.1080/00913367.1986.10672999]
54. Madans JH. Health surveys. In: Smelser NJ, Baltes PB, editors. International Encyclopedia of the Social & Behavioral Sciences. 2nd edition. Amsterdam, The Netherlands: Elsevier; 2015:725-730.
55. Gould SJ. Consumer attitudes toward health and health care: a differential perspective. J Consum Affairs 1988;22(1):96-118 [FREE Full text] [doi: 10.1111/j.1745-6606.1988.tb00215.x]
56. Institute of Medicine, Board on Neuroscience and Behavioral Health, Committee on Health Literacy. In: Nielsen-Bohlman L, Panzer AM, Kindig DA, editors. Health Literacy: A Prescription to End Confusion. Washington, DC, USA: National Academies Press; 2004.
57. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
58. Wilson EV, Lankton NK. Modeling patients' acceptance of provider-delivered e-health. J Am Med Inform Assoc 2004;11(4):241-248 [FREE Full text] [doi: 10.1197/jamia.M1475] [Medline: 15064290]
59. 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]
60. Thon FM, Jucks R. Believing in expertise: how authors' credentials and language use influence the credibility of online health information. Health Commun 2017 Jul;32(7):828-836. [doi: 10.1080/10410236.2016.1172296] [Medline: 27466693]
61. Yang F, Li C. The color of gender stereotyping: the congruity effect of topic, color, and gender on health messages' persuasiveness in cyberspace. Comput Human Behav 2016 Nov;64:299-307. [doi: 10.1016/j.chb.2016.07.001]
62. Borah P, Xiao X. The importance of 'Likes': the interplay of message framing, source, and social endorsement on credibility perceptions of health information on Facebook. J Health Commun 2018;23(4):399-411. [doi: 10.1080/10810730.2018.1455770] [Medline: 29601271]
63. Dutta-Bergman MJ. The impact of completeness and web use motivation on the credibility of e-health information. J Commun 2004 Jun;54(2):253-269. [doi: 10.1111/j.1460-2466.2004.tb02627.x]
64. Rains SA, Karmikel CD. Health information-seeking and perceptions of website credibility: examining Web-use orientation, message characteristics, and structural features of websites. Comput Human Behav 2009 Mar;25(2):544-553. [doi: 10.1016/j.chb.2008.11.005]
65. Cao W, Zhang X, Xu K, Wang Y. Modeling online health information-seeking behavior in China: the roles of source characteristics, reward assessment, and internet self-efficacy. Health Commun 2016 Sep;31(9):1105-1114. [doi: 10.1080/10410236.2015.1045236] [Medline: 26861963]

66. Shim M, Jo HS. What quality factors matter in enhancing the perceived benefits of online health information sites? Application of the updated DeLone and McLean Information Systems Success Model. Int J Med Inform 2020 May;137:104093. [doi: 10.1016/j.ijmedinf.2020.104093] [Medline: 32078918]
67. Robins D, Holmes J, Stansbury M. Consumer health information on the Web: the relationship of visual design and perceptions of credibility. J Am Soc Inf Sci 2010;61(1):13-29. [doi: 10.1002/asi.21224]
68. Dillard JP. Persuasion past and present: attitudes aren't what they used to be. Commun Monographs 2009 Jun 02;60(1):90-97. [doi: 10.1080/03637759309376299]
69. Fogg BJ, Soohoo C, Danielson DR, Marable L, Stanford J, Tauber ER. How do users evaluate the credibility of Web sites?: a study with over 2,500 participants. In: Proceedings of the 2003 conference on Designing for user experiences. 2003 Jun Presented at: DUX '03; Jun 6 - 7, 2003; San Francisco, CA, USA p. 1-15. [doi: 10.1145/997078.997097]
70. Fogg BJ, Marshall J, Laraki O, Osipovich A, Varma C, Fang N, et al. What makes Web sites credible?: a report on a large quantitative study. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2001 Mar Presented at: CHI '01; Mar 31- Apr 5, 2001; Seattle, WA, USA p. 61-68. [doi: 10.1145/365024.365037]
71. Chaiken S. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. J Pers Soc Psychol 1980 Nov;39(5):752-766. [doi: 10.1037/0022-3514.39.5.752]
72. Petty RE, Cacioppo JT. The effects of involvement on responses to argument quantity and quality: central and peripheral routes to persuasion. J Pers Soc Psychol 1984 Jan;46(1):69-81. [doi: 10.1037/0022-3514.46.1.69]
73. 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]
74. Wathen CN, Burkell J. Believe it or not: factors influencing credibility on the Web. J Am Soc Inf Sci 2002;53(2):134-144 [FREE Full text] [doi: 10.1002/asi.10016]
75. Lederman R, Fan H, Smith S, Chang S. Who can you trust? Credibility assessment in online health forums. Health Policy Technol 2014 Mar;3(1):13-25. [doi: 10.1016/j.hlpt.2013.11.003]
76. Sillence E, Blythe JM, Briggs P, Moss M. A revised model of trust in Internet-based health information and advice: cross-sectional questionnaire study. J Med Internet Res 2019 Nov 11;21(11):e11125 [FREE Full text] [doi: 10.2196/11125] [Medline: 31710297]
77. Chang CC, Huang MH. Antecedents predicting health information seeking: a systematic review and meta-analysis. Int J Inf Manag 2020 Oct;54:102115. [doi: 10.1016/j.ijinfomgt.2020.102115]
78. DeLone WH, McLean ER. Information systems success: the quest for the dependent variable. Inf Syst Res 1992 Mar;3(1):60-95. [doi: 10.1287/isre.3.1.60]
79. Petter S, DeLone W, McLean E. Measuring information systems success: models, dimensions, measures, and interrelationships. Eur J Inf Syst 2008;17(3):236-263. [doi: 10.1057/ejis.2008.15]
80. Erkan I, Evans C. The influence of eWOM in social media on consumers' purchase intentions: an extended approach to information adoption. Comput Human Behav 2016 Aug;61:47-55. [doi: 10.1016/j.chb.2016.03.003]
81. Sussman SW, Siegal WS. Informational influence in organizations: an integrated approach to knowledge adoption. Inf Syst Res 2003 Mar;14(1):47-65. [doi: 10.1287/isre.14.1.47.14767]
82. Purcarea VL, Gheorghe IR, Petrescu CM. Credibility elements of eWOM messages in the context of health care services. A Romanian perspective. J Med Life 2013 Sep 15;6(3):254-259 [FREE Full text] [Medline: 24146683]
83. Park JH, Chung HE, Yoo WS. Is the Internet a primary source for consumer information search?: group comparison for channel choices. J Retail Consum Serv 2009 Mar;16(2):92-99. [doi: 10.1016/j.jretconser.2008.11.002]
84. 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]
85. Lee SY, Choi J, Noh GY. Factors influencing health-related internet activities and their outcomes. J Health Commun 2016 Nov;21(11):1179-1186. [doi: 10.1080/10810730.2016.1236853] [Medline: 27802113]
86. Liu S, Zhang R, Lu X. The impact of individuals' attitudes toward health websites on their perceived quality of health information: an empirical study. Telemed J E Health 2019 Nov;25(11):1099-1107. [doi: 10.1089/tmj.2018.0217] [Medline: 30585763]
87. Marchionini G. Exploratory search: from finding to understanding. Commun ACM 2006 Apr 01;49(4):41-46. [doi: 10.1145/1121949.1121979]
88. Vakkari P. Task-based information searching. Ann Rev Info Sci Tech 2003;37(1):413-464. [doi: 10.1002/aris.1440370110]
89. Chang YS, Zhang Y, Gwizdka J. The effects of information source and eHealth literacy on consumer health information credibility evaluation behavior. Comput Human Behav 2021 Feb;115:106629. [doi: 10.1016/j.chb.2020.106629]
90. Liu J, Song S, Zhang Y. Linguistic features and consumer credibility judgment of online health information. University of Illinois. 2021. URL: https://www.ideals.illinois.edu/handle/2142/109674 [accessed 2021-05-07]

Abbreviations

IQ: information quality

Edited by A Mavragani; submitted 15.01.22; peer-reviewed by DJ Lim, E Sillence, S Song; comments to author 05.02.22; accepted 15.03.22; published 28.04.22.

Please cite as:
Zhang Y, Kim Y
Consumers' Evaluation of Web-Based Health Information Quality: Meta-analysis
J Med Internet Res 2022;24(4):e36463
URL: https://www.jmir.org/2022/4/e36463
doi:10.2196/36463
PMID:35482390

Original Paper

Training Resources Targeting Social Media Skills to Inform Rehabilitation for People Who Have an Acquired Brain Injury: Scoping Review

Melissa Brunner[1*], BAppSc, MHlthSc, PhD; Rachael Rietdijk[1*], BAppSc, PhD; Leanne Togher[1*], BAppSc, PhD

Faculty of Medicine and Health, University of Sydney, Eora Nation, Camperdown, Australia

[*]all authors contributed equally

Corresponding Author:
Melissa Brunner, BAppSc, MHlthSc, PhD
Faculty of Medicine and Health
University of Sydney
Level 6, Susan Wakil Health Building
Western Avenue
Eora Nation, Camperdown, 2006
Australia
Phone: 61 2 86279866
Email: melissa.brunner@sydney.edu.au

Abstract

Background: In 2020 and 2021, people increasingly used the internet to connect socially and professionally. However, people with an acquired brain injury (ABI) experience challenges in using social media, and rehabilitation professionals have reported feeling underprepared to support them in its use. To date, no review of social media skills training to inform ABI rehabilitation has been conducted.

Objective: This scoping review aimed to examine research on interventions addressing social media skills and safety, with a focus on people living with health conditions; free web-based resources for the general public on social media skills training; and currently available online support groups for people with ABI.

Methods: An integrative scoping review was conducted, with a systematic search strategy applied in March and November 2020 across OvidSP (MEDLINE, AMED, PsycINFO, and Embase), Scopus, Web of Science, CINAHL, Google Scholar, Google, and Facebook. The data collected were critically appraised and synthesized to describe the key content and features of social media training resources.

Results: This review identified 47 peer-reviewed academic articles, 48 social media training websites, and 120 online support groups for people with ABI. A key recommendation was interactive training with practical components addressing cybersafety, how to use platforms, and how to connect with others. However, no social media training resources that were relevant and accessible for people with ABI were identified.

Conclusions: Training resources to support people with ABI in safely using social media are limited. The key content to be addressed and the features to be incorporated into web-based social media training were determined, including the need for interactive training that is co-designed and safe and incorporates practical components that support people with ABI. These findings can be used to inform the development of web-based evidence-based support for people with ABI who may be vulnerable when participating in social media.

(J Med Internet Res 2022;24(4):e35595) doi:10.2196/35595

KEYWORDS

brain injury; social media; training; social communication; scoping review

Introduction

Background

Connecting with others and having conversations is an integral part of being a person in our society, where being social through developing and maintaining relationships is customary for many [1]. From birth, we learn through exploration [2]; as we learn, our brains discover how to adapt and accomplish more complex tasks [3]. Being social and having conversations are complex tasks that demonstrate the ability of our brain to learn and develop over time through our interactions with one another [4]. Throughout 2020 and 2021, because of restrictions being placed on real-life interactions physically, we observed immense changes in the way we interacted globally, with people increasingly using the internet to connect socially and professionally [5]. Using social media allows people living in this digital age to connect with others locally and internationally in various ways [6,7]. In addition to using social media to communicate with peers [6], communities have increasingly been strategically connecting on social media for climate change activism [8], disaster response and recovery [9], and global health issues relating to the COVID-19 pandemic [10]. These web-based communities include those who have communication difficulties resulting from brain injury.

An acquired brain injury (ABI) is defined as an injury to the brain that occurs after birth and is not related to congenital disorders, developmental disabilities, or degenerative processes [11] and can be the result of both traumatic and nontraumatic causes (eg, through a stroke, tumor, infection, or trauma). In 2016, there was a global incidence of 276 million people who were living with an ABI [12]. More recent epidemiological studies indicate that this number is increasing substantially, with reports of 12 million people who have a stroke [13] and 69 million people who sustain a traumatic brain injury (TBI) [14] annually. For people who experience an ABI, changes in their cognitive function alter their executive functioning and social communication skills [15]. As a result, many individuals have difficulties living independently, returning to work or study, and navigating their interpersonal relationships [16-19]. For those who experience aphasia (ie, “a communication disability due to an acquired impairment of language modalities caused by focal brain damage”) [20] or significant physical disability whereby speech clarity is difficult to achieve (ie, dysarthria or dyspraxia) [21], changes in communication after their injury can be marked, with social communication difficulties immediately apparent to their communication partners. However, for people who experience cognitive-communication disorders (commonly occurring after a right hemisphere stroke or TBI) [22], changes in a person’s social communication can present with more subtlety. For example, an individual with cognitive-communication difficulties may be able to have a conversation, yet they may either be verbose or alternatively have impoverished conversational skills (ie, they talk too much or too little in a conversation) [23]. Although subtle, these changes in their interactions and conversations resulting from cognitive-communication difficulties can have a dramatic impact on their life [24]. Within a year of their brain injury, many people lose their friends, and their remaining relationships can be strained [25]. This means that they are often socially isolated and can feel disconnected [26] and have difficulty engaging with their wider social networks as well as in their close relationships [27].

Using social media may present an opportunity to reduce social isolation after an ABI, offering an important way for people to connect with family, friends, and the broader community [28]. Social media platforms can provide a view of the lives of others, opportunities to explore topics of personal interest, and an avenue for people to interact with other users that may not be possible or accessible in their real-world environments. The use of social media may allow people to prepare for their communicative interactions without the cognitive overload that face-to-face interactions may present [29]. Asynchronous communications in social media may reduce pressure regarding the need for immediate responses and eliminate the need to use and interpret social nonverbal skills in real time [29]. In addition, there is the ready accommodation of spelling and grammatical errors observed in social media posts among the general public [30] as well as greater awareness of the need to reduce linguistic bigotry [31], as seen in fierce criticism of social media examples such as the Instagram account @celeb_spellcheck that mocks spelling and grammatical errors of other users [32]. Using social media can also give people the opportunity to interact with others on the web without disclosing their history of brain injury, or alternatively, it can provide a platform for individuals to advocate for themselves (and the ABI community) as people living with an ABI [29].

Although social media offers many opportunities for connection, it also presents safety risks to individuals [33], with >50% of the general public having experienced or observed negative social media interactions [34]. People who have physical impairments, intellectual disabilities, and specific chronic diseases are often specifically targeted on the web, which can result in depression, anxiety, and distress [35]. Cyberharassment incidents against people who have a disability occur with greater frequency than for the general public, with some cyberscammers pretending to have a disability themselves to insinuate themselves into the lives of their potential targets [35,36]. Although people with a TBI use social media for connection and communication, few report having received formal support during rehabilitation in social media skills or safety, and those who had been cyberscammed had not been offered training after these distressing experiences [37]. Limitations in our individual abilities to use social media, communication difficulties and patterns, and the social networks we interact with are sources of vulnerability in web-based interactions [38]. In seeking connection on the web, the vulnerability of people with an ABI can be exacerbated because of changes in their executive functioning and social cognition [39], which may make it challenging for them to recognize cyberscams or regulate their own interactions. The complexity of these issues is evidenced by people with a TBI who report having been on the receiving end of negative comments on the web as well as having been the perpetrator of cyberbullying [40]. Following the first 6 months of the COVID-19 pandemic when many countries underwent various levels of restrictions on movement and

gatherings, the eSafety Commissioner (Australia's national independent regulator for cybersafety) reported a noticeable increase in internet hate and harassment [41]. Recent work led by Gould et al [42] has also identified that injury-related cognitive impairments and social isolation increased the vulnerability of people with ABIs to cyberscams, particularly romance scams, and that the current lack of effective intervention may lead to scam revictimization. As such, it is imperative that resources and guidance are available to support the safe use of social media for connection, particularly for those who may be more vulnerable to web-based hate or scams through direct targeting from cyberscammers and trolls.

Even before the COVID-19 pandemic, it was reported that >60% of the general public used social media for everyday communication and socializing [43]. Thus, before acquiring a brain injury, it is highly likely that people will have developed a range of skills and competencies across various social media platforms and may want to return to using them in their daily interactions. Subsequently, during rehabilitation, clinicians should consider the person's use of social media when examining communication contexts before their injury [29]. Rehabilitation clinicians have expressed willingness to support people in using social media after their brain injury [44]. However, they report uncertainty and concern regarding the potential risks people with a brain injury may encounter on social media [44]. Standard speech pathology clinical practice and ABI rehabilitation are yet to include social media skills training, and evidence-based guidance on social media use for brain injury rehabilitation professionals is lacking [33]. In Australia, despite growing recognition that people with disabilities are overrepresented in reporting negative social media experiences, there are limited social media cybersafety resources available that are accessible for adults with cognitive impairments. For example, at the time of writing, the eSafety Commissioner has made only 4 of their extensive range of resources available on their website in an *easy English* format [45]. With little access to guidance or support, rehabilitation professionals use greater caution in therapy, resulting in restrictive or reactive approaches rather than proactive interventions [44].

Objective

To date, there has been one pioneering review of social media use by people with a TBI [28], indicating an urgent need for research in this area. Other reviews have demonstrated that peer support group interventions are a promising way to support individuals and promote adjustment following an ABI [46] and that digital health interventions can improve psychosocial and health outcomes for people with ABI [47]. Although peer support for people with ABI in online environments has been occurring with greater frequency, the efficacy of such support is yet to be explored. Recent studies have identified that, for people with TBI, similar barriers and facilitators affect both web-based and real-world social participation after their injury [48]. Although a previous review presented evidence for incorporating technology into cognitive-communication rehabilitation following TBI [49], no review has examined the evidence for training to support the use of social media generally, let alone for individuals who have cognitive-communication difficulties after a specific health condition such as an ABI. There is a pressing need to determine the outcomes and cost-effectiveness of social media training for people with an ABI [29]. As such, this review was developed to be intentionally broad, summarizing the literature across health conditions, population subgroups (ie, across the life span), and intervention strategies. It was also designed to incorporate user preference studies as well as freely accessible resources available to the general public, to elucidate what people with an ABI or their supporters may encounter when attempting to access support or guidance themselves through *Googling it*. The aims were well suited to a scoping review, which is ideal for reviewing a large and complex body of information across multiple sources not previously reviewed [50,51]. The specific research questions that guided this integrative scoping review are as follows:

1. What studies have investigated training for developing social media skills and safety?
2. What free web-based resources are used for social media skills training that are available to the general public?
3. What online support groups, aimed at providing peer connections and support, are available for people with an ABI?

Methods

Overview

An integrative scoping review was conducted to locate the following: (1) research relating to the use of training resources for developing social media skills and safety, with a focus on resources for people living with health conditions or their supporters (eg, rehabilitation professionals, family, friends, and direct support workers); (2) free web-based resources for social media skills training for the general public; and (3) online support groups for people with an ABI that are aimed at providing peer connections and support. The authors approached this scoping review with relativist ontological and constructivist epistemological positions [52], believing that meaning arises from interactions between individuals, their worlds, and their communities. As such, a pragmatic approach was applied [53], using mixed methods to understand the common and distinct components of social media skills training for people with an ABI. The review focused only on currently available web-based resources and support communities and used the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines [54]. A review protocol was written before commencing this study (Multimedia Appendix 1 [55-65]), with systematic searching initially conducted in March 2020 (by the first author [MB] and a student intern) and a secondary search conducted in November 2020 (by the first author, MB). The research team comprised 3 qualified speech pathologists who were also experienced researchers in the fields of ABI and digital health and a student intern. The first (MB) and second (RR) authors conducted data extraction in conjunction with the student. Multimedia Appendix 2 [54] provides the completed PRISMA-ScR checklist for this scoping review.

Each of the 3 research questions was explored using a specific approach for the search strategy, exclusion criteria, study selection, data extraction, and critical appraisal. The specific details of each approach are provided in Multimedia Appendix 1.

Search Strategy and Study Selection

Peer-reviewed Academic Literature Investigating Social Media Training

To identify relevant evidence, we included any publication in a peer-reviewed journal in the English language relating to the use of training resources for social media skills training. Resources for any population were included, although the search strategy was targeted toward people living with health conditions or their supporters. This included both descriptions of the development of such resources and reporting of the outcomes of the use of the resources, as well as user perspectives on such resources or training needs in the use of social media. Training resources were defined as any written content (including both webpages and downloadable PDFs), training packages (websites or apps), or audiovisual content. The systematic searching and selection based on title and abstract were conducted by the first author (MB), with selection based on full-text articles reviewed independently by the first (MB) and second (RR) authors and discrepancies resolved through discussion to achieve consensus.

Free Web-Based Social Media Training Resources

To identify currently available free web-based resources, we included any web-based resource targeted at the general public that was aimed at improving social media skills. The authors defined web-based resources as any written content (including both webpages and downloadable PDFs), training packages (websites or apps), or audiovisual content available via the internet, as identified through Google searching.

Online Support Groups for People With an ABI

To identify the currently available online support groups for people with a brain injury, we included any existing web-based community targeted at people with an ABI (and their supporters, such as family and friends) that was aimed at providing peer connections and support. The authors defined web-based communities as any networking groups or communities available via the internet, as identified through networks known to the researchers, Google searching, and Facebook searching.

Data Extraction, Critical Appraisal, and Synthesis

Data were extracted from all information sources and managed using Microsoft Excel spreadsheets [55] that were accessed by the research team to ensure consistency and transparency. Specific details regarding the data extraction, critical appraisal, and data synthesis processes used across the 3 information sources are presented in Multimedia Appendix 1. Data were synthesized descriptively to map different aspects of the literature and the resources outlined in our key questions. Descriptive statistics were calculated using Microsoft Excel [55]. The data collected from the charting process and critical appraisal were compared and synthesized qualitatively using open coding [66] across information sources to identify similarities in the format and content of web-based social media training and support resources. An inductive approach to content analysis was used, in which codes and categories were derived from the data, to enable the description and categorization of the data without potentially restricting the findings [67]. The authors discussed the findings in the context of ABI rehabilitation, and a constant comparison was used to compare and integrate the data [68]. This allowed the research team to identify key issues to address and features to use in social media training for people with an ABI [53] to inform future resource development and implementation research.

Results

Overview

A total of 47 peer-reviewed articles, 48 social media training websites, and 120 online support groups were included in this review. Figure 1 outlines the flow of sources through the inclusion process [54], and Table 1 provides a description of the target populations and behaviors identified across the 3 information sources). The completed PRISMA-ScR checklist can be found in Multimedia Appendix 2.

Figure 1. The flow of information sources through the inclusion process [54].

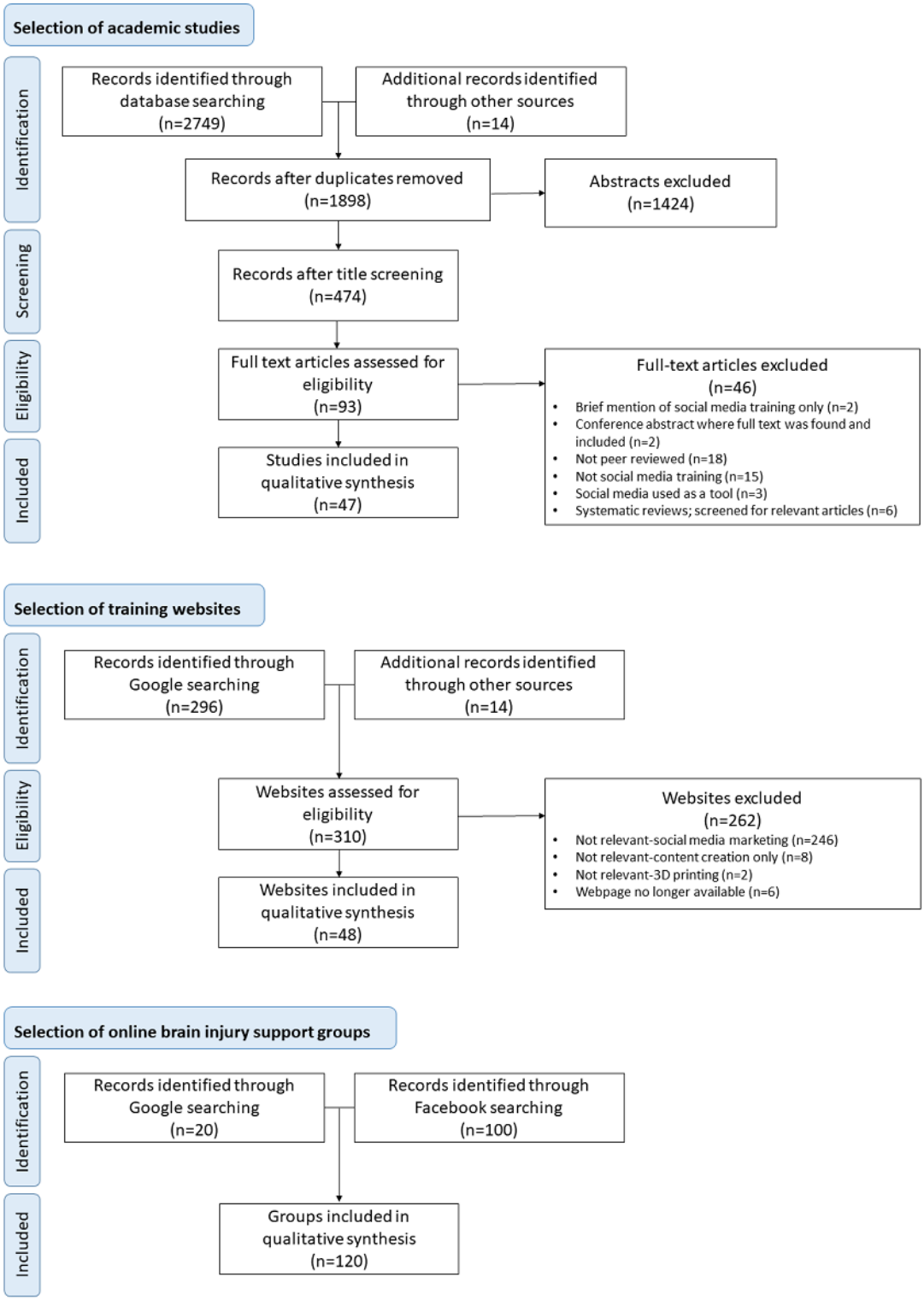


Table 1. Target populations and behaviors identified across the 3 information sources.

	Academic literature (N=47), n (%)	Social media training websites (N=48), n (%)	Online support groups (N=120), n (%)
Target population or audience			
Neurotypical	21 (44.7)	47 (97.9)	15 (12.5)
Cognitive disability	22 (46.8)	1 (2.1)	105 (87.5)
Physical disability	6 (12.8)	N/A[a]	N/A
Communication disability	11 (23.4)	N/A	N/A
Adults	35 (74.5)	46 (95.8)	120 (100)
Children (aged <13 years) and young people (aged 13-18 years)	14 (29.8)	2 (4.2)	N/A
Target behaviors			
Social participation	17 (36.2)	N/A	120 (100)
Professional use (eg, for employment or social marketing)	10 (21.3)	8 (16.7)	N/A
Cybersafety	7 (14.9)	11 (22.9)	N/A
Social media, information and communication technology knowledge, or use	8 (17)	15 (31.2)	N/A
Content creation	N/A	8 (16.7)	N/A
Well-being or other (mental health, sleep, eating, life skills, and support)	5 (10.6)	6 (12.5)	120 (100)

[a]N/A: not applicable.

Question 1: Peer-reviewed Academic Literature Investigating Social Media Training

Of the 47 included articles [37,69-114], most studies were conducted in either the United States (20/47, 43%) or Australia (20/47, 38%), with a median publication date of 2017 (range 2011-2020). Nearly two-thirds of the included studies were published as journal articles (30/47, 64%), with the remainder published as conference proceedings (13/47, 28%) or conference abstracts (4/47, 9%). These studies used a diverse range of qualitative (19/47, 40%), quantitative (15/47, 32%), and mixed methods (13/47, 28%) research designs. The target populations involved were also diverse but were identified as being vulnerable in some way when using social media. For example, there were studies with young people exploring cyberbullying [76,94,97,106] and studies with older adults working on building their internet and social media confidence [69,70,87]. Of the 47 included studies, 22 (47%) included people who had a cognitive disability, 21 (45%) included *neurotypical* individuals, and 11 (23%) specifically investigated social media training for people with communication disabilities. Behaviors of interest included social participation (17/47, 36%), professional use of social media (10/47, 21%), generic social media or internet knowledge and use (8/47, 17%), cybersafety (7/47, 15%), well-being (3/47, 6%), or life skills (2/47, 4%). Most intervention studies involved training delivered in person (25/33, 76%) and on the web (9/33, 27%).

The methodological reporting in the articles included in this review was variable in quality. A total of 13% (6/47) of the articles were conference abstracts [78,92,95,99,103,108] with limited study details where the methodological rigor could not be appraised. Of the 41 articles, 26 (63%) appraised were considered to have reported either strong methodology (a rating score >80% of the specified tool criterion) [37,69,70,75,77,80,83,86,88,91,93,94,97,100-102,106,112] or moderately strong methodology (a rating score >50% of the specified tool criterion) [72,74,76,79,87,89,96,104]. Multimedia Appendix 3 [37,69-114] provides a description and critical appraisal scores for the included studies.

Question 2: Free Web-Based Social Media Training Resources

Of the 48 included websites (Multimedia Appendix 4), 46 (96%) were directed toward *all ages* or an adult audience, with only 2 (4%) websites tailored for children or young people. Of the 48 websites, 9 (19%) were targeted toward teachers and parents, and 3 (6%) catered for more niche populations, specifically women, older people, and employees. Of these 48 websites, only 1 (2%) considered neurodiversity at all [115] and highlighted the risk that people with learning disabilities may encounter when using social media. Two-thirds of the training websites offered text-based resources (32/48, 67%), with the remainder offering courses and workshops (9/48, 19%) or videos and podcasts (7/48, 15%) and targeted generic *How to use social media* skills (15/48, 31%), cybersafety (11/48, 23%), social marketing or employment (8/48, 17%), content creation (8/48, 17%), or well-being (6/48, 13%).

The web accessibility and readability of the included social media training websites were evaluated to determine their suitability for someone who has difficulties with cognition or communication that can commonly occur after a brain injury.

Web accessibility is defined as the websites, tools, and technologies that are designed and developed so that people with disabilities can use and navigate the internet [116]. When websites are poorly designed, they can create barriers, often preventing people from using them as intended. The Web Content Accessibility Guidelines (WCAG) 2.1 [117] provide recommendations on how to make website content more accessible to a wider range of people with disabilities, such as visual, auditory, physical, speech, cognitive, language, learning, and neurological disabilities, "but will not address every user need for people with these disabilities" [117]. The WCAG 2.1 has 3 conformance levels, of which the highest is Level AAA, indicating that a webpage satisfies all the recommended criteria for making it as accessible as possible [117]. Of the 48 included websites, none conformed to Level AAA, with only 1 (2%) website [118] identified as having less than 10 Level AAA accessibility issues. Most of the websites (30/48, 63%) were found to have >20 Level AAA conformance issues identified, indicating that these websites are likely to be difficult for people with a brain injury to use and navigate.

Readability refers to how easily a person can read and understand written text [119]. Organizations such as the American Medical Association and the Agency for Healthcare Research and Quality recommend that the readability of health information materials should not be greater than a sixth-grade reading level [120,121]. In the Australian context, it is also generally recommended that health information be presented using plain language and should aim for a reading grade level of 6 (neurotypical children aged 11-12 years who have English as their first language; eg, [122-125]). More recently, it has been recommended that when preparing digital content, "you write your content for a reading level of age 9 or lower. If you're unable to achieve reading age 9, the more readable you can make your content, the better" [126]. Most included social media training websites had a reading grade level of 7 or above (39/48, 81%), a level considered high for the general public and indicating that these websites are highly likely to be difficult for people with a brain injury to read and understand. Table 2 provides an overview of the included websites, and Multimedia Appendix 3 provides a list of the included websites.

Table 2. Characteristics, accessibility, and readability of the web-based social media training websites (N=48).

	Value, n (%)
Text based	32 (67)
Courses or workshops	9 (19)
Videos or podcasts	7 (15)
Accessibility	
Conformance to international WCAG[a] 2.1 (Level AAA)	0 (100)
<10 issues[b] identified	1 (2)
<20 issues identified	18 (38)
≥20 issues identified	30 (63)
Readability	
Flesch-Kincaid reading grade level 6 or below[c]	9 (19)
Flesch-Kincaid reading grade level 7 or above	39 (81)

[a]WCAG: Web Content Accessibility Guidelines.

[b]Occurrences of an issue (error, warning, or review item) determined to be in contravention of the WCAG 2.1.

[c]Flesch-Kincaid reading grade level 6 or below is easily understood by neurotypical children aged 11 to 12 years who have English as their first language.

Question 3: Online Support Groups for People With an ABI

Of the 120 included online support groups, 102 (85%) were Facebook groups or pages, 16 (13.3%) were web-based discussion forums, and 2 (1.7%) offered support meetings via web-based videoconferencing. Of the 120 groups, 62 (51.7%) of the groups did not specify a location or were branded as being global. Of those that specified a location, the majority were based in the United States (38/120, 31.7%), with a smaller number of groups located in Australia (13/120, 10.8%), Canada (4/120, 3.3%), Aotearoa or New Zealand (2/120, 1.7%), the United Kingdom (2/120, 1.7%), and India (1/120, 0.8%). All groups aimed to provide social participation and support for members. Most of the groups catered to people with a brain injury (102/120, 85%), with 12.5% (15/120) of the groups targeted toward family and caregivers of people with a brain injury, and 2.5% (3/120) of the groups catering specifically to women with a brain injury. Most of the groups were closed (private) or required users to request permission to join the group (97/120, 80.8%), with a smaller number of groups using an open format (23/120, 19.2%), where information on membership and posts are available and easily accessible to the general public. Many of the groups were moderated by people with an ABI (20/120, 16.7%), family or caregivers (4/120, 3.3%), or brain injury support organizations (47/120, 39.2%). However, for 40.8% (49/120) of the groups, it was unclear from the public group description as to who managed the group, although several implied that it may be organized by someone with a brain injury, a family member, or a health professional. None of the support group *about* statements made any direct reference or inference to social media skills training. Table 3

provides an overview of the included online support groups, and Multimedia Appendix 5 provides a list of the included support groups.

Table 3. Characteristics of the online social media brain injury support groups (N=120).

	Value, n (%)
Type of platform	
Facebook group or page	102 (85)
Web-based discussion forum or space	16 (13.3)
Online support meetings (using web-based videoconferencing)	2 (1.7)
Type of space	
Open or public group	23 (19.2)
Closed or need to request to join the group	97 (80.8)
Target audience	
People with a brain injury (adults)	102 (85)
Women with a brain injury (adults)	3 (2.5)
Family or caregivers of people with a brain injury (adults and children)	15 (12.5)

Integration of Findings

Across the 3 information sources, key issues relating to the content and format of social media training were identified and synthesized. The topics of how to use social media and cybersafety were present across all 3 information sources. Other frequent topics identified were the following: developing relationships in social media, how to use technology to access social media, maintaining relationships, finding support for people who can guide social media use and troubleshooting, how to connect and interact with peers, navigating personal and professional use of social media, being able to access technology through assistive devices, and how to improve or maintain your well-being. Similarities in the training format were less apparent across the data sources. However, all the sources included training or guidance that incorporated interactive elements, allowing a two-way flow of information. In addition, the following features were identified in some of the data sources: led by people with an ABI, offer choice to cater to individual preferences, provide opportunities for real-life practice and where possible include practical examples, provide support for memory and recall, and, when able, tailor training to the individual's needs. Table 4 outlines in which of the 3 information sources these issues were identified.

Table 4. Key issues to address in social media training (or resources) identified across the 3 information sources.

Key issues	Academic literature	Social media training websites	Online support groups
Content (topics)			
Developing relationships	✓		✓
How to use technology	✓		
How to use social media	✓	✓	✓
Maintaining relationships	✓		
Cybersafety	✓	✓	✓
Support people	✓		✓
Peer connection	✓		✓
Personal and professional use		✓	
Technology access	✓		
Well-being		✓	✓
Format (features)			
Interactive	✓	✓	✓
Led by and designed with people with an acquired brain injury	✓		✓
Offer choice	✓		✓
Opportunity for real-life practice	✓		✓
Practical examples	✓	✓	
Support memory and recall	✓		
Individually tailored	✓		

Discussion

Principal Findings

In this review, we identified 47 articles that explored social media skills training across diverse populations, which provided information to support the future development of training for people with an ABI. Across the 48 websites that were freely accessible to the public, we found considerable variability in the content, readability, and accessibility of the information provided regarding social media skills training. For many of the 130 online support groups identified, it was unclear who managed the group, and none directly referenced social media skills training. Our synthesis of the data across the 3 information sources (peer-reviewed literature, websites, and online support groups) identified several key issues that are critical to addressing social media training for people with a brain injury. Evidence regarding the effectiveness of social media training was limited, and the authors were unable to draw robust conclusions about which active components of content and techniques for delivery were more successful than others. However, it was evident that training programs would best be engaging and interactive, developed using a co-design process by collaborating with people with an ABI and their supporters (eg, rehabilitation clinicians, family, friends, and industry stakeholders). This corroborates previous work that identified the need for specific research involving the user-centered co-design, development, and evaluation of web-based social media training for people with a brain injury [29]. It is also vital that training resources meet the needs of ABI rehabilitation professionals to facilitate the adoption of social media use as part of their rehabilitation intervention services [44].

Furthermore, the findings of this review indicated that the content of social media skills training should primarily provide general information about how to use social media and how to stay safe. In addition, training should demonstrate how to communicate and connect with others on key social media platforms (ie, Facebook, Twitter, and Instagram). Information should be presented using techniques that support recall and retention [85,87,110], along with support strategies for executive function (eg, strategies that promote learning or everyday problem solving) [90,101,111]. The importance of real-life practice that is personally meaningful [127-129] and being able to practice with everyday communication partners [128,130] has been well established in evidence-based cognitive rehabilitation programs for people with an ABI. Therefore, opportunities to practice key navigational, communication, and safety skills for using social media are likely to be critical components of successful training programs. Similarly, training programs that can be tailored to the individual, or at least provide users the ability to self-select which modules are completed and in which order they are completed, would be optimal. As few of the included websites met the recommended criteria for readability and accessibility of their information, we suggest that when developing web-based resources, it is imperative that the readability of the materials be appropriate for the general public (ie, below reading grade 6 level) and the information accessible (ie, comply with WCAG 2.1 recommendations). In combination with priorities for learning

about social media from interviews with people with a TBI [37,40] and brain injury rehabilitation professionals [44], the findings of this review can now be used to inform user-centered co-design discussions regarding the development of social media skills training programs for people with an ABI. In an interim step, clinicians may use the findings (specifically those outlined in Table 4) as guidance in clinical practice in providing social media training to support people with an ABI after their injury.

Regarding the online support groups available for people with a brain injury (and their supporters), there were an overwhelming number of groups, many moderated by people with an ABI or brain injury support organizations. However, for many others, it was unclear who ran the group. Most groups were closed or private, indicating a user preference for increased privacy. This may be due to the knowledge that open groups can leave people vulnerable to broader public commentary and being targeted by cyberscammers or trolls. Although closed groups provide users with some privacy, it should be noted that until joining the group, a user is unable to determine the membership of the group or the content of the group's posts. This may mean that users are faced with difficult decisions regarding which group to join. Findings from this review may facilitate clinicians' identification of relevant websites or online support groups to suggest to families of people with an ABI to support social media skills training and guidance.

Limitations and Directions for Future Research

This review had several limitations. First, in the interest of timeliness, the first author (MB) conducted the data extraction and critical appraisal with a student intern, and a second academic reviewer was not used. Second, although we used a systematic and comprehensive search strategy, some relevant websites and online support groups may have been omitted. Given the ephemeral nature of the internet, additional social media training website resources and online support groups for people with a brain injury may have been developed since the searches were conducted. Indeed, since this review was conducted, there have been websites that have been developed to address cybersafety and social media use for people with an ABI, such as the CyberABIlity [131] and social-ABI-lity [132] training programs. Similarly, the design and content of the included websites and online support groups are likely to change over time. In addition, as the evaluation of the included online support groups was beyond the scope of this review, the benefits and risks of these groups warrant further investigation. Therefore, clinicians using Multimedia Appendices 4 and 5 provided in this review (ie, the lists of websites and online support groups) should confirm the presence and credibility of any website or support group before referring individuals to it.

The findings of this research also suggest that, alongside research exploring the development and implementation of co-designed social media skills training resources for people with a brain injury and their supporters, greater insights into the experiences and perspectives of rehabilitation professionals are required, in addition to studies exploring cybersafety and web-based self-representation of people with an ABI on social media. Robust research is urgently needed to evaluate the effectiveness of the potential key ingredients for the content and delivery of social media training resources for people with an ABI that were identified in this scoping review and have been implemented in web-based training resources such as CyberABIlity [131] and social-ABI-lity [132]. Further research is needed to determine how using social media following an ABI may support social participation, cyber-resilience, and growth in personal agency. The results of such studies likely have the potential to inform resources adaptable for other people who experience difficulties in social communication, for example, young people with developmental language disabilities or adults with dementia.

Conclusions

There is limited research exploring social media skills training, with few web-based training resources available for people with an ABI. Social media offers an important means of connection for people with an ABI, providing continual opportunities for them to observe and interact with others, as well as a way to develop new relationships when their cognitive-communication skills may limit their in-person interactions. The findings from this review, along with priorities for learning about social media informed by people with a brain injury and rehabilitation professionals, can be used to inform the development of novel web-based training resources to support social media skill development in people with an ABI. The development of such resources may drive sustainable change through the provision of clinical practice guidance and the creation of web-based support networks to help people with a brain injury build their own social media mastery.

Acknowledgments

The authors would like to acknowledge the Gadigal lands of the Eora Nation upon which the University of Sydney now stands. The authors pay their respects to those who care and continue to care for Country and their continuing connection to land, waters, and community.

The authors would also like to acknowledge Abby McGuire, a University of Sydney intern from Washington University for their assistance with the Google search and data extraction phases for the websites and online support groups.

No funding was received for this study. The second author (RR) was supported by funding through Insurance and Care NSW (icare).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Scoping review protocol for the search strategy, exclusion criteria, study selection, data extraction, and critical appraisal applied across the 3 information sources.
[DOCX File , 17 KB - jmir_v24i4e35595_app1.docx]

Multimedia Appendix 2
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 506 KB - jmir_v24i4e35595_app2.pdf]

Multimedia Appendix 3
Included academic peer-reviewed literature (N=47).
[DOCX File , 30 KB - jmir_v24i4e35595_app3.docx]

Multimedia Appendix 4
List of included websites.
[DOCX File , 38 KB - jmir_v24i4e35595_app4.docx]

Multimedia Appendix 5
List of included online support groups.
[DOCX File , 45 KB - jmir_v24i4e35595_app5.docx]

References

1. McKeon R. Communication, truth, and society. Ethics 1957 Jan;67(2):89-99. [doi: 10.1086/291096]
2. Rochat P. Object manipulation and exploration in 2- to 5-month-old infants. Dev Psychol 1989;25(6):871-884. [doi: 10.1037/0012-1649.25.6.871]
3. Blakemore SJ, Choudhury S. Development of the adolescent brain: implications for executive function and social cognition. J Child Psychol Psychiatry 2006;47(3-4):296-312. [doi: 10.1111/j.1469-7610.2006.01611.x] [Medline: 16492261]
4. Frith CD, Frith U. Social cognition in humans. Curr Biol 2007 Aug 21;17(16):R724-R732 [FREE Full text] [doi: 10.1016/j.cub.2007.05.068] [Medline: 17714666]
5. Digital use around the world in July 2020. We are Social. 2020. URL: https://wearesocial.com/blog/2020/07/digital-use-around-the-world-in-july-2020 [accessed 2020-11-30]
6. boyd D. It's complicated: the social lives of networked teens. London, UK: Yale University Press; 2014. URL: http://www.danah.org/books/ItsComplicated.pdf
7. Media Consumer Survey 2019: Australian media and digital preferences. Deloitte. 2019. URL: https://www2.deloitte.com/au/en/pages/technology-media-and-telecommunications/articles/media-consumer-survey.html [accessed 2020-11-30]
8. Pearce W, Niederer S, Özkula SM, Sánchez Querubín N. The social media life of climate change: platforms, publics, and future imaginaries. WIREs Clim Change 2018 Dec 27;10(2):e569. [doi: 10.1002/wcc.569]
9. Fosso Wamba S, Edwards A, Akter S. Social media adoption and use for improved emergency services operations: the case of the NSW SES. Ann Oper Res 2017 May 29;283(1-2):225-245. [doi: 10.1007/s10479-017-2545-9]
10. Chen E, Lerman K, Ferrara E. Tracking social media discourse about the COVID-19 pandemic: development of a public coronavirus Twitter data set. JMIR Public Health Surveill 2020 May 29;6(2):e19273 [FREE Full text] [doi: 10.2196/19273] [Medline: 32427106]
11. Toronto Acquired Brain Injury Network. Definition of acquired brain injury. Toronto Acquired Brain Injury Network. 2020. URL: https://abinetwork.ca/individuals-families/about-brain-injury/definition/ [accessed 2020-11-30]
12. GBD 2016 Neurology Collaborators. Global, regional, and national burden of neurological disorders, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol 2019 May;18(5):459-480 [FREE Full text] [doi: 10.1016/S1474-4422(18)30499-X] [Medline: 30879893]
13. GBD 2019 Stroke Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Neurol 2021 Oct;20(10):795-820 [FREE Full text] [doi: 10.1016/S1474-4422(21)00252-0] [Medline: 34487721]
14. Dewan M, Rattani A, Gupta S, Baticulon RE, Hung YC, Punchak M, et al. Estimating the global incidence of traumatic brain injury. J Neurosurg 2018 Apr 01:1-18. [doi: 10.3171/2017.10.JNS17352] [Medline: 29701556]

15. MacDonald S. Introducing the model of cognitive-communication competence: a model to guide evidence-based communication interventions after brain injury. Brain Inj 2017;31(13-14):1760-1780. [doi: 10.1080/02699052.2017.1379613] [Medline: 29064304]
16. Corrigan JD, Cuthbert JP, Harrison-Felix C, Whiteneck GG, Bell JM, Miller AC, et al. US population estimates of health and social outcomes 5 years after rehabilitation for traumatic brain injury. J Head Trauma Rehabil 2014;29(6):E1-E9. [doi: 10.1097/HTR.0000000000000020] [Medline: 24495919]
17. Khan F, Baguley IJ, Cameron ID. 4: Rehabilitation after traumatic brain injury. Med J Aust 2003 Mar 17;178(6):290-295. [doi: 10.5694/j.1326-5377.2003.tb05199.x] [Medline: 12633489]
18. Ponsford JL, Olver JH, Curran C. A profile of outcome: 2 years after traumatic brain injury. Brain Inj 1995 Jan;9(1):1-10. [doi: 10.3109/02699059509004565] [Medline: 7874089]
19. Hewetson R, Cornwell P, Shum D. Social participation following right hemisphere stroke: influence of a cognitive-communication disorder. Aphasiology 2017 Apr 26;32(2):164-182. [doi: 10.1080/02687038.2017.1315045]
20. Berg K, Isaksen J, Wallace SJ, Cruice M, Simmons-Mackie N, Worrall L. Establishing consensus on a definition of aphasia: an e-Delphi study of international aphasia researchers. Aphasiology 2020 Dec 11;36(4):385-400. [doi: 10.1080/02687038.2020.1852003]
21. Strand EA. Neurologic substrates of motor speech disorders. Perspect Neurophysiol Neurogenic Speech Lang Disord 2013 Oct 1;23(3):98-104. [doi: 10.1044/nnsld23.3.98]
22. College of Audiologists and Speech-Language Pathologists of Ontario (CASLPO). Preferred Practice Guideline for Cognitive-Communication Disorders. Toronto, Canada: College of Audiologists and Speech-Language Pathologists of Ontario (CASLPO); 2015.
23. Tate RL. Executive dysfunction and characterological changes after traumatic brain injury: two sides of the same coin? Cortex 1999 Feb;35(1):39-55. [doi: 10.1016/s0010-9452(08)70784-6] [Medline: 10213533]
24. Togher L, Power E, Tate R, McDonald S, Rietdijk R. Measuring the social interactions of people with traumatic brain injury and their communication partners: the adapted Kagan scales. Aphasiology 2010 Feb 03;24(6-8):914-927. [doi: 10.1080/02687030903422478]
25. O'flaherty CA, Douglas JM. Living with cognitive-communicative difficulties following traumatic brain injury: using a model of interpersonal communication to characterize the subjective experience. Aphasiology 1997 Sep;11(9):889-911. [doi: 10.1080/02687039708250463]
26. Douglas JM. Conceptualizing self and maintaining social connection following severe traumatic brain injury. Brain Inj 2013;27(1):60-74. [doi: 10.3109/02699052.2012.722254] [Medline: 23252437]
27. Temkin NR, Corrigan JD, Dikmen SS, Machamer J. Social functioning after traumatic brain injury. J Head Trauma Rehabil 2009;24(6):460-467. [doi: 10.1097/HTR.0b013e3181c13413] [Medline: 19940679]
28. Brunner M, Hemsley B, Palmer S, Dann S, Togher L. Review of the literature on the use of social media by people with traumatic brain injury (TBI). Disabil Rehabil 2015;37(17):1511-1521. [doi: 10.3109/09638288.2015.1045992] [Medline: 25985717]
29. Brunner MJ. #TwitterMind - Social media use by people with communication difficulties after traumatic brain injury (TBI). University of Technology Sydney. 2020. URL: https://opus.lib.uts.edu.au/bitstream/10453/142209/1/01front.pdf [accessed 2020-11-30]
30. Hemsley B, Dann S, Palmer S, Allan M, Balandin S. "We definitely need an audience": experiences of Twitter, Twitter networks and tweet content in adults with severe communication disabilities who use augmentative and alternative communication (AAC). Disabil Rehabil 2015;37(17):1531-1542 [FREE Full text] [doi: 10.3109/09638288.2015.1045990] [Medline: 26030339]
31. Cameron D. Verbal hygiene. London, UK: Routledge; 1995.
32. Zhou M. The problem with celeb spellcheck. MTV. 2021. URL: https://www.mtv.com.au/news/fxzuxa/the-problem-with-celeb-spellcheck [accessed 2021-07-07]
33. Brunner M, Hemsley B, Togher L, Dann S, Palmer S. Social media and people with traumatic brain injury: a metasynthesis of research informing a framework for rehabilitation clinical practice, policy, and training. Am J Speech Lang Pathol 2021 Jan 27;30(1):19-33. [doi: 10.1044/2020_AJSLP-20-00211] [Medline: 33332986]
34. Duggan M. Online Harassment 2017. Pew Research Center. 2017. URL: https://www.pewresearch.org/internet/2017/07/11/online-harassment-2017/ [accessed 2018-01-14]
35. Alhaboby ZA, Barnes J, Evans H, Short E. Cyber-victimization of people with chronic conditions and disabilities: a systematic review of scope and impact. Trauma Violence Abuse 2019 Jul;20(3):398-415. [doi: 10.1177/1524838017717743] [Medline: 29333943]
36. Alhaboby ZA, al-Khateeb HM, Barnes J, Short E. 'The language is disgusting and they refer to my disability': the cyberharassment of disabled people. Disability Soc 2016 Oct 05;31(8):1138-1143. [doi: 10.1080/09687599.2016.1235313]
37. Brunner M, Palmer S, Togher L, Hemsley B. 'I kind of figured it out': the views and experiences of people with traumatic brain injury (TBI) in using social media-self-determination for participation and inclusion online. Int J Lang Commun Disord 2019 Mar;54(2):221-233 [FREE Full text] [doi: 10.1111/1460-6984.12405] [Medline: 29873159]

38. Orru K, Hansson S, Gabel F, Tammpuu P, Krüger M, Savadori L, et al. Approaches to 'vulnerability' in eight European disaster management systems. Disasters (forthcoming) 2021 Mar 24. [doi: 10.1111/disa.12481] [Medline: 33760259]
39. McDonald S, English T, Randall R, Longman T, Togher L, Tate RL. Assessing social cognition and pragmatic language in adolescents with traumatic brain injuries. J Int Neuropsychol Soc 2013 May;19(5):528-538. [doi: 10.1017/S1355617713000039] [Medline: 23442218]
40. Brunner M, Palmer S, Togher L, Dann S, Hemsley B. “If I knew what I was doing on Twitter then I would use it more”: Twitter experiences and networks of people with traumatic brain injury (TBI). Brain Impair 2020;21(1):1-18. [doi: 10.1017/BrImp.2019.12]
41. Covid-19 impact on Australian adults' online activities and attitudes. eSafety Commissioner. 2020. URL: https://www.esafety.gov.au/about-us/research/covid-19-impacts-australian-adults-online-activities-and-attitudes [accessed 2021-07-07]
42. Gould KR, Carminati JY, Ponsford JL. "They just say how stupid I was for being conned". Cyberscams and acquired brain injury: a qualitative exploration of the lived experience of survivors and close others. Neuropsychol Rehabil 2021 Dec 27:1-21. [doi: 10.1080/09602011.2021.2016447] [Medline: 34957919]
43. Yellow Social Media Report 2018: part one - consumers. yellow. 2018. URL: https://www.yellow.com.au/wp-content/uploads/2018/06/Yellow-Social-Media-Report-2018-Consumer.pdf [accessed 2019-11-16]
44. Brunner M, Togher L, Palmer S, Dann S, Hemsley B. Rehabilitation professionals' views on social media use in traumatic brain injury rehabilitation: gatekeepers to participation. Disabil Rehabil 2021 Jul;43(14):1955-1964. [doi: 10.1080/09638288.2019.1685604] [Medline: 31694430]
45. Help in Easy Read format. eSafety Commissioner. 2021. URL: https://www.esafety.gov.au/diverse-groups/living-with-disability/easy-read [accessed 2021-07-07]
46. Hughes R, Fleming P, Henshall L. Peer support groups after acquired brain injury: a systematic review. Brain Inj 2020 Jun 06;34(7):847-856. [doi: 10.1080/02699052.2020.1762002] [Medline: 32421382]
47. Shaffer KM, Tigershtrom A, Badr H, Benvengo S, Hernandez M, Ritterband LM. Dyadic psychosocial eHealth interventions: systematic scoping review. J Med Internet Res 2020 Mar 04;22(3):e15509 [FREE Full text] [doi: 10.2196/15509] [Medline: 32130143]
48. Ketchum JM, Sevigny M, Hart T, O'Neil-Pirozzi TM, Sander AM, Juengst SB, et al. The association between community participation and social internet use among adults with traumatic brain injury. J Head Trauma Rehabil 2020;35(4):254-261 [FREE Full text] [doi: 10.1097/HTR.0000000000000566] [Medline: 32108716]
49. Brunner M, Hemsley B, Togher L, Palmer S. Technology and its role in rehabilitation for people with cognitive-communication disability following a traumatic brain injury (TBI). Brain Inj 2017;31(8):1028-1043. [doi: 10.1080/02699052.2017.1292429] [Medline: 28471267]
50. 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]
51. Tricco AC, Lillie E, Zarin W, O'Brien K, Colquhoun H, Kastner M, et al. A scoping review on the conduct and reporting of scoping reviews. BMC Med Res Methodol 2016 Feb 09;16:15 [FREE Full text] [doi: 10.1186/s12874-016-0116-4] [Medline: 26857112]
52. Crotty M. The foundations of social research: meaning and perspective in the research process. St Leonards, Australia: Sage Publications; 1998.
53. Morgan DL. Paradigms lost and pragmatism regained: methodological implications of combining qualitative and quantitative methods. J Mix Methods Res 2007;1(1):48-76. [doi: 10.1177/2345678906292462]
54. 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]
55. Excel. Microsoft Corporation. URL: https://www.microsoft.com/en-us/microsoft-365/excel [accessed 2020-01-10]
56. EndNote. The EndNote Team. 2013. URL: https://endnote.com/product-details [accessed 2020-01-10]
57. Maher CG, Sherrington C, Herbert RD, Moseley AM, Elkins M. Reliability of the PEDro scale for rating quality of randomized controlled trials. Phys Ther 2003 Aug;83(8):713-721. [doi: 10.1093/ptj/83.8.713] [Medline: 12882612]
58. The Joanna Briggs Institute critical appraisal tools for use in JBI systematic reviews: checklist for quasi-experimental studies (non-randomized experimental studies). The Joanna Briggs Institute. 2017. URL: https://jbi.global/sites/default/files/2019-05/JBI_Quasi-Experimental_Appraisal_Tool2017_0.pdf [accessed 2020-11-30]
59. The Joanna Briggs Institute critical appraisal tools for use in JBI systematic reviews: checklist for cohort studies. The Joanna Briggs Institute. 2017. URL: https://jbi.global/sites/default/files/2019-05/JBI_Critical_Appraisal-Checklist_for_Cohort_Studies2017_0.pdf [accessed 2020-11-30]
60. Critical appraisal of a survey checklist. Center for Evidence-Based Management (CEBMa). 2019. URL: https://www.cebma.org/wp-content/uploads/Critical-Appraisal-Questions-for-a-Survey.pdf [accessed 2020-11-30]
61. The Joanna Briggs Institute critical appraisal tools for use in JBI systematic reviews: checklist for qualitative research. The Joanna Briggs Institute. 2017. URL: https://jbi.global/sites/default/files/2019-05/JBI_Critical_Appraisal-Checklist_for_Qualitative_Research2017_0.pdf [accessed 2020-11-30]

62. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Educ Inf 2018 Dec 18;34(4):285-291. [doi: 10.3233/EFI-180221]
63. Siteimprove Accessibility Checker (Google Chrome extension). Siteimprove. 2021. URL: https://chrome.google.com/webstore/detail/siteimprove-accessibility/djcglbmbegflehmbfleechkjhmedcopn?hl=en [accessed 2021-01-30]
64. Readability Test Tool. WebFX. 2020. URL: https://www.webfx.com/tools/read-able/ [accessed 2021-01-30]
65. Word. Microsoft Corporation. 2021. URL: https://www.microsoft.com/en-in/microsoft-365/word [accessed 2020-01-10]
66. 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]
67. Hsieh FH, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
68. Glaser BG, Strauss AL. The discovery of grounded theory: strategies for qualitative research. Boca Raton, FL, USA: Routledge; 1999.
69. Arthanat S, Vroman KG, Lysack C. A home-based individualized information communication technology training program for older adults: a demonstration of effectiveness and value. Disabil Rehabil Assist Technol 2016;11(4):316-324. [doi: 10.3109/17483107.2014.974219] [Medline: 25512061]
70. Arthanat S, Vroman KG, Lysack C, Grizzetti J. Multi-stakeholder perspectives on information communication technology training for older adults: implications for teaching and learning. Disabil Rehabil Assist Technol 2019 Jul;14(5):453-461 [FREE Full text] [doi: 10.1080/17483107.2018.1493752] [Medline: 30044659]
71. Barber LK, Cucalon MS. Modifying the Sleep Treatment Education Program for Students to include technology use (STEPS-TECH): intervention effects on objective and subjective sleep outcomes. Stress Health 2017 Dec 03;33(5):684-690. [doi: 10.1002/smi.2746] [Medline: 28156049]
72. Bayor A. Co-designing with young adults with intellectual disability to develop social life skills. In: Proceedings of the 2nd African Conference for Human Computer Interaction: Thriving Communities. 2018 Dec Presented at: AfriCHI '18; December 3-7, 2018; Windhoek, Namibia p. 1-5. [doi: 10.1145/3283458.3283517]
73. Bayor AA. HowToApp: supporting life skills development of young adults with intellectual disability. In: Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility. 2019 Oct Presented at: ASSETS '19; October 28-30, 2019; Pittsburgh, PA, USA p. 697-699. [doi: 10.1145/3308561.3356107]
74. Bayor AA, Sitbon L, Ploderer B, Bircanin F, Koplick S, Brereton M. Leveraging participation: supporting skills development of young adults with intellectual disability using social media. In: Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility. 2019 Oct Presented at: ASSETS '19; October 28-30, 2019; Pittsburgh, PA, USA p. 143-155. [doi: 10.1145/3308561.3353793]
75. Bramstedt KA, Ierna BN, Woodcroft-Brown VK. Using SurveyMonkey to teach safe social media strategies to medical students in their clinical years. Commun Med 2014;11(2):117-124. [doi: 10.1558/cam.v11i2.27556] [Medline: 26596120]
76. Cardoso P, Hawk DV, Cross D. What young people need to make better-informed decisions when communicating with digital images: implications for mental health and well-being. Health Educ Behav 2020 Feb;47(1):29-36. [doi: 10.1177/1090198119885433] [Medline: 31726855]
77. Caron JG, Light J. Social media experiences of adolescents and young adults with cerebral palsy who use augmentative and alternative communication. Int J Speech Lang Pathol 2017 Feb;19(1):30-42. [doi: 10.3109/17549507.2016.1143970] [Medline: 27063698]
78. Caton S, Landman R. Increasing awareness of online radicalisation for young people with intellectual disabilities: evaluating the get smart (Social Media Awareness and Resilience Training) project. J Intellect Disabil Res 2019 Jul;63(7):657. [doi: 10.1111/jir.12652]
79. Daems J, Dekelver J, De Vliegher A, Smets J, Martens L. Able To Include: focus groups with persons with IDD and their coaches communicating through social media. In: Conference on Raising Awareness for the Societal and Environmental Role of Engineering and (Re)Training Engineers for Participatory Design. 2015 Presented at: Engineering4Society '15; June 18-19, 2015; Leuven, Belgium p. 76-79. [doi: 10.1109/Engineering4Society.2015.7177903]
80. Flickinger TE, O'Hagan T, Chisolm MS. Developing a curriculum to promote professionalism for medical students using social media: pilot of a workshop and blog-based intervention. JMIR Med Educ 2015 Dec 01;1(2):e17 [FREE Full text] [doi: 10.2196/mededu.4886] [Medline: 27731846]
81. George DR. "Friending Facebook?" A minicourse on the use of social media by health professionals. J Contin Educ Health Prof 2011;31(3):215-219. [doi: 10.1002/chp.20129] [Medline: 21953663]
82. Gilster ME, Kleinschmit JL, Cummings SP, Ronnenberg MM. Teaching note—pick your platform: social media advocacy skill building. J Soc Work Educ 2020;56(1):170-178. [doi: 10.1080/10437797.2019.1593903]
83. Grace E, Raghavendra P, Newman L, Wood D, Connell T. Learning to use the internet and online social media: what is the effectiveness of home-based intervention for youth with complex communication needs? Child Lang Teach Ther 2014 Jan 14;30(2):141-157. [doi: 10.1177/0265659013518565]
84. Gruba P, Bettinson M, Mulder J, Grigg G. Developing social media training resources for AusAID scholarship students. In: Proceedings of the 30th ASCILITE-Australian Society for Computers in Learning in Tertiary Education Annual Conference. 2013 Presented at: ascilite '19; December 1-4, 2013; Sydney, Australia p. 330-339.

85. Hafez A, Wang Y. Designing training mechanism for the elderly to use social media mobile apps – a research proposal. In: Proceedings of the 10th International Conference on Social Computing and Social Media. 2018 Presented at: SCSM '18; July 15-20, 2018; Las Vegas, NV, USA. [doi: 10.1007/978-3-319-91521-0_4]
86. Hafez A, Wang Y. A training social media mobile app prototype for the elderly: a pilot study. In: Proceedings of the International Conferences on Usability & User Experience, and Human Factors and Assistive Technology. 2019 Presented at: AHFE '19; July 24-28, 2019; Washington, DC, USA. [doi: 10.1007/978-3-030-19135-1_19]
87. Harvey J, Beck A, Carr CT. A cognitive social media training program and intergenerational learning: a pilot study with older adults and speech-language pathology graduate students. Perspect ASHA Spec Interest Groups 2019 Aug 15;4(4):683-695. [doi: 10.1044/2019_pers-sig15-2018-0004]
88. Hemsley B, Palmer S, Dann S, Balandin S. Using Twitter to access the human right of communication for people who use Augmentative and Alternative Communication (AAC). Int J Speech Lang Pathol 2018 Feb;20(1):50-58. [doi: 10.1080/17549507.2017.1413137] [Medline: 29355398]
89. Henning MA, Hawken S, MacDonald J, McKimm J, Brown M, Moriarty H, et al. Exploring educational interventions to facilitate health professional students' professionally safe online presence. Med Teach 2017 Sep;39(9):959-966. [doi: 10.1080/0142159X.2017.1332363] [Medline: 28562145]
90. Holmes KM, O'Loughlin N. The experiences of people with learning disabilities on social networking sites. Br J Learn Disabil 2014;42(1):1-5. [doi: 10.1111/bld.12001]
91. Kydland F, Molka-Danielsen J, Balandin S. Examining the use of social media tool “Flickr” for impact on loneliness for people with intellectual disability. In: Proceedings of the 2012 Norsk konferanse for organisasjoners bruk av informasjonsteknologi. 2012 Presented at: NOKOBIT '12; November 19-21, 2012; Bodø, Norway p. 253-264.
92. Lellis-Santos C, Halpin PA. Workshop report: "Using social media and smartphone applications in practical lessons to enhance student learning" in Búzios, Brazil (August 6-8, 2017). Adv Physiol Educ 2018 Jun 01;42(2):340-342 [FREE Full text] [doi: 10.1152/advan.00011.2018] [Medline: 29676613]
93. Maben S, Colley K. #NotAllBaptismByFire: the state of social media training at student-run communication organizations. Southwestern Mass Commun J 2017 Mar 1;32(2).
94. McLean SA, Wertheim EH, Masters J, Paxton SJ. A pilot evaluation of a social media literacy intervention to reduce risk factors for eating disorders. Int J Eat Disord 2017 Jul;50(7):847-851. [doi: 10.1002/eat.22708] [Medline: 28370321]
95. Menger F, Morris J, Salis C. Interventions to support Internet use with aphasia: preliminary evidence for a tailored approach. Int J Stroke 2017 Nov 1;12(5 Supplement 2):40. [doi: 10.1177/1747493017732216]
96. Mentis H, Madjaroff G, Massey A. Upside and downside risk in online security for older adults with mild cognitive impairment. In: Proceedings of CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19; May 4-9, 2019; Glasgow, UK p. 1-13. [doi: 10.1145/3290605.3300573]
97. Prihastuty R, Swaraswati Y, Rahmawati DA, Khasanah SN. Social media literacy training for preventing cyberbullying intention among senior high school students. In: Proceedings of the International Conference on Science, Technology, and Humanity. 2019 Presented at: ISETH '19; December 3-4, 2019; Surakarta, Indonesia.
98. Probst D. Social media literacy as an IEP intervention for social and emotional learning. J Media Lit Educ 2017 Nov 04;9(2):45-57. [doi: 10.23860/jmle-2019-09-02-04]
99. Raghavendra P, Grace E, Hutchinson C, Newman L, Wood D. Online mentoring of young people with disabilities in rural South Australia: enhancing social networks via social media use. Dev Med Child Neurol 2016;58(Supplement 3):25-26. [doi: 10.1111/dmcn.13069]
100. Raghavendra P, Hutchinson C, Grace E, Wood D, Newman L. "I like talking to people on the computer": outcomes of a home-based intervention to develop social media skills in youth with disabilities living in rural communities. Res Dev Disabil 2018 May;76:110-123. [doi: 10.1016/j.ridd.2018.02.012] [Medline: 29571686]
101. Raghavendra P, Newman L, Grace E, Wood D. 'I could never do that before': effectiveness of a tailored Internet support intervention to increase the social participation of youth with disabilities. Child Care Health Dev 2013 Jul;39(4):552-561. [doi: 10.1111/cch.12048] [Medline: 23763256]
102. Raghavendra P, Newman L, Grace E, Wood D. Enhancing social participation in young people with communication disabilities living in rural Australia: outcomes of a home-based intervention for using social media. Disabil Rehabil 2015;37(17):1576-1590. [doi: 10.3109/09638288.2015.1052578] [Medline: 26043751]
103. Raghavendra P, Rillotta F, Thirumanickam A. Social media use training for students with intellectual and developmental disabilities: a step towards connecting students with and without disability. J Intellect Disabil Res 2019;63(7):759-760. [doi: 10.1111/jir.12658]
104. Raghavendra P, Wood D, Newman L, Lawry J. Why aren't you on Facebook?: patterns and experiences of using the internet among young people with physical disabilities. Technol Disabil 2012 Jun 01;24(2):149-162. [doi: 10.3233/tad-2012-0343]
105. Robertson M, Shoss MK, Broom MA. Social media: social intelligence training module. MedEdPORTAL 2016 Aug 26;12:10442 [FREE Full text] [doi: 10.15766/mep_2374-8265.10442] [Medline: 31008220]
106. Robinson J, Hill NT, Thorn P, Battersby R, Teh Z, Reavley NJ, et al. The #chatsafe project. Developing guidelines to help young people communicate safely about suicide on social media: a Delphi study. PLoS One 2018 Nov 15;13(11):e0206584 [FREE Full text] [doi: 10.1371/journal.pone.0206584] [Medline: 30439958]

107. Rogers AP, Barber LK. Addressing FoMO and telepressure among university students: could a technology intervention help with social media use and sleep disruption? Comput Human Behav 2019 Apr;93:192-199. [doi: 10.1016/j.chb.2018.12.016]
108. Schwertel U, Kammerer Y, Oloff C, Gerjets P, Schmidt M. Assessing and training social media skills in vocational education supported by TEL instruments. In: Proceedings of the 9th European Conference on Technology Enhanced Learning. 2014 Presented at: EC-TEL '14; September 16-19, 2019; Graz, Austria p. 588-589. [doi: 10.1007/978-3-319-11200-8_83]
109. Stott TC, MacEachron A, Gustavsson N. Social media and child welfare: policy, training, and the risks and benefits from the administrator's perspective. Adv Soc Work 2017 Jan 30;17(2):221-234. [doi: 10.18060/21263]
110. Tsaousides T, Matsuzawa Y, Lebowitz M. Familiarity and prevalence of Facebook use for social networking among individuals with traumatic brain injury. Brain Inj 2011;25(12):1155-1162. [doi: 10.3109/02699052.2011.613086] [Medline: 21961574]
111. Usoro I, Connolly T, Raman S, French T, Caulfield S. Using games based learning to support young people with learning disabilities stay safe online. In: Proceedings of the 10th European Conference on Games Based Learning. 2016 Presented at: ECGBL '16; October 6-7, 2016; Paisley, UK.
112. Wang XL, Eberhard D, Voron M, Bernas R. Helping students with cognitive disabilities improve social writing skills through email modeling and scaffolding. Educ Stud 2016 Mar 24;42(3):252-268. [doi: 10.1080/03055698.2016.1160825]
113. Weaver JL, Swank JM. Mindful connections: a mindfulness-based intervention for adolescent social media users. J Child Adolesc Couns 2019 Jun 14;5(2):103-112. [doi: 10.1080/23727810.2019.1586419]
114. Williams L, Johnson E, Patterson JE. The appropriate use and misuse of social media in MFT training programs: problems and prevention. Contemp Fam Ther 2013 Mar 6;35(4):698-712. [doi: 10.1007/s10591-013-9256-4]
115. Atkinson D. Social media risks for people who have learning disabilities. RCNi Plus. 2019 Nov 11. URL: https://rcni.com/learning-disability-practice/newsroom/research-news/social-media-risks-people-who-have-learning-disabilities-154876 [accessed 2020-11-30]
116. Introduction to Web Accessibility. W3C Web Accessibility Initiative. 2019. URL: https://www.w3.org/WAI/fundamentals/accessibility-intro/ [accessed 2021-02-25]
117. Web Content Accessibility Guidelines (WCAG) 2.1. W3C Web Accessibility Initiative. 2018 Jun 5. URL: https://www.w3.org/TR/WCAG21/ [accessed 2021-02-25]
118. Grady M. Intro to Facebook. Digitallearn.org. 2020. URL: https://www.digitallearn.org/courses/intro-to-facebook [accessed 2020-03-30]
119. DuBay WH. The Principles of Readability. Impact Information. 2004. URL: https://files.eric.ed.gov/fulltext/ED490073.pdf [accessed 2021-07-14]
120. Weiss BD. Health literacy: a manual for clinicians. Chicago, IL, USA: American Medical Association; 2003.
121. Brega AG, Barnard J, Mabachi NM, Weiss BD, DeWalt DA, Brach C, et al. AHRQ Health Literacy Universal Precautions Toolkit, Second Edition. Rockville, MD, USA: Agency for Healthcare Research and Quality; Jan 2015.
122. Checklist for designing consumer-friendly health information. Northern NSW Local Health District. 2016. URL: https://cbrhl.org.au/wp-content/uploads/2020/11/Northern-NSW-Health-Literacy-Checklist-for-writing-consumer-information.pdf [accessed 2021-07-14]
123. Assessing Readability. SA Health, Government of South Australia. 2013. URL: https://www.sahealth.sa.gov.au/wps/wcm/connect/fcb907004e455125ab8eaf8ba24f3db9/HLT-AssessingReability-T7-PHCS-SQ20130118.pdf?MOD=AJPERES&CACHEID=ROOTWORKSPACE-fcb907004e455125ab8eaf8ba24f3db9-niQAujZ [accessed 2021-07-14]
124. Assessing readability. Tasmanian Government Department of Health. 2019. URL: https://www.dhhs.tas.gov.au/publichealth/about_us/health_literacy/health_literacy_toolkit/assessing_readability [accessed 2021-07-14]
125. Health Literacy Guide: Health Communities Unit. Barwon Health. 2019. URL: https://www.barwonhealth.org.au/images/Healthy_Communities/HCU_Guidelines_FINAL-BH_External_MAR19.pdf [accessed 2021-07-14]
126. State of New South Wales (Customer Service NSW). Digital Service Toolkit, Design Standards: Write in plain English. New South Wales Government. 2021. URL: https://www.digital.nsw.gov.au/digital-service-toolkit/design-standards/design-with-users-for-users/accessibility-and-3 [accessed 2021-07-14]
127. Defense Health Agency (DHA), Defense and Veterans Brain Injury Center (DVBIC). Clinical recommendations: cognitive rehabilitation for service members and veterans following mild to moderate traumatic brain injury. Traumatic Brain Injury Center of Excellence. 2019 Feb. URL: https://health.mil/Reference-Center/Publications/2020/07/30/Cognitive-Rehabilitation-for-Following-Mild-to-Moderate-TBI-Clinical-Recommendation-Full [accessed 2012-12-29]
128. Togher L, McDonald S, Tate R, Power E, Rietdijk R. Communication skills of people with severe traumatic brain injury (TBI) can be improved by training everyday communication partners: findings from a single-blind multi-centre clinical trial. In: Armstrong B, editor. 40th clinical aphasiology conference: a special issue of aphasiology. West Sussex, UK: Psychology Press; 2010.
129. Ylvisaker M. Self-coaching: a context-sensitive, person-centred approach to social communication after traumatic brain injury. Brain Impair 2006;7(3):246-258. [doi: 10.1375/brim.7.3.246]

130. Togher L, McDonald S, Tate R, Power E, Rietdijk R. Training communication partners of people with severe traumatic brain injury improves everyday conversations: a multicenter single blind clinical trial. J Rehabil Med 2013 Jul;45(7):637-645 [FREE Full text] [doi: 10.2340/16501977-1173] [Medline: 23803687]
131. Cyberability. Monash University. 2022. URL: https://cyberability.org.au/ [accessed 2022-03-04]
132. social-ABI-lity. University of Sydney. 2022. URL: https://abi-communication-lab.sydney.edu.au/courses/social-abi-lity/ [accessed 2022-03-04]

Abbreviations

ABI: acquired brain injury
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
TBI: traumatic brain injury
WCAG: Web Content Accessibility Guidelines

Edited by T Leung; submitted 09.12.21; peer-reviewed by J Hoepner, N Behn; comments to author 19.02.22; revised version received 04.03.22; accepted 29.03.22; published 28.04.22.

Please cite as:
Brunner M, Rietdijk R, Togher L
Training Resources Targeting Social Media Skills to Inform Rehabilitation for People Who Have an Acquired Brain Injury: Scoping Review
J Med Internet Res 2022;24(4):e35595
URL: https://www.jmir.org/2022/4/e35595
doi:10.2196/35595
PMID:35482369

Viewpoint

Operationalizing and Evaluating Synchronous Virtual Group Health Interventions: Wide-Scale Implementation at a Tertiary Care Academic Hospital

Neesha Hussain-Shamsy[1,2,3], MHS; Ian McMillan[4], MSc; Sheridan Cook[5], MSW; Alyssa Furfaro-Argier[6], MHI; Andrea Sadler[7], BSc; Faith Delos-Reyes[8], BPHE; Lori Wasserman[9], MD; Sacha Bhatia[1,3,10], MD; Danielle Martin[1,3,11], MD; Emily Seto[1,2], PhD; Simone N Vigod[1,3,10,11], MD; Juveria Zaheer[11,12], MD; Payal Agarwal[3,13], MD; Geetha Mukerji[1,3,11,14], MD

[1]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada
[2]Centre for Global eHealth Innovation, Techna Institute, University Health Network, Toronto, ON, Canada
[3]Institute for Health System Solutions and Virtual Care, Women's College Hospital, Toronto, ON, Canada
[4]Women's Virtual, Women's College Hospital, Toronto, ON, Canada
[5]Gevity Consulting Inc, Toronto, ON, Canada
[6]Department of Information Management and Technology, Women's College Hospital, Toronto, ON, Canada
[7]Department of Psychiatry, Women's College Hospital, Toronto, ON, Canada
[8]Department of Cardiology, Women's College Hospital, Toronto, ON, Canada
[9]Reproductive Life Stages Program, Women's College Hospital, Toronto, ON, Canada
[10]Institute for Clinical Evaluative Sciences, Toronto, ON, Canada
[11]Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada
[12]Centre for Addiction and Mental Health, Toronto, ON, Canada
[13]Family Practice Health Centre, Women's College Hospital, Toronto, ON, Canada
[14]Department of Medicine, Women's College Hospital, Toronto, ON, Canada

Corresponding Author:
Geetha Mukerji, MD
Institute for Health System Solutions and Virtual Care
Women's College Hospital
76 Grenville Street
Toronto, ON, M5S 1B2
Canada
Phone: 1 416 323 6400
Email: geetha.mukerji@wchospital.ca

Abstract

Group-based health interventions are an important component of health promotion and management. To provide continuity of care throughout the COVID-19 pandemic, our institution undertook a rapid pivot to delivering group-based health interventions via a videoconferencing service which was securely embedded into both the electronic medical record and the patient portal to sustainably address immediate health service delivery needs during the pandemic and beyond. In this paper, we (1) describe the institutionally driven operationalization of a system to provide integrated synchronous video group visits across our hospital and (2) present a proposed strategy to comprehensively evaluate outcomes regarding their implementation, quality, and impact. Lessons for other institutions and the potential future role of synchronous video group visits to enhance how care can be scaled for delivery are discussed.

(J Med Internet Res 2022;24(4):e29841) doi:10.2196/29841

KEYWORDS

virtual care; group therapy; patient education; videoconferencing; sustainability; innovation; health systems; health promotion; patient portal; electronic medical records; health service delivery; video call

Introduction

Background

Group-based health interventions, where small groups of patients receive the same health intervention (eg, patient education or therapy) from one or more facilitators, are frequently used for health promotion and the self-management and treatment of many physical and mental illnesses [1-6]. Group-based health interventions are cost-effective as they allow multiple individuals to receive treatment simultaneously [1,4,7], particularly in settings with a limited supply of qualified providers [1,8].

With the conversion of many health care services to virtual care during the COVID-19 pandemic, the need for virtual approaches to group-based health interventions suddenly became pressing to ensure continuity of care. With little evidence to drive this shift [9], it was not uncommon for technologies and procedures that addressed immediate needs within the context of the pandemic, but which would not necessarily be sustainable in the long-term or optimized for high-quality care, to be used [10]. A systematic review (which included 15 studies of mixed quality) on the use of group videoconferencing for health interventions showed that it was satisfactory to patients with high attendance rates, but had mixed results in terms of its impact on health outcomes; high-quality studies in this review demonstrated positive results in terms of group process outcomes (eg, group cohesion and social support) [9]. The main limitations were that the studies in the review were typically pilot studies and the technical equipment used was supplied by the study, limiting the generalizability of the findings to institutionally driven and scalable approaches to deliver care in the long-term [9]. It is generally accepted that, after the pandemic, virtual care will take a more prominent and permanent place in the health care system. Therefore, it is important for health institutions to address the need for sustainable high-quality solutions with the ability to monitor impact.

This paper aims to reflect on and share the experience of Women's College Hospital (WCH) in developing and operationalizing a sustainable model of synchronous group video visits during the pandemic. We also present a proposed evaluation strategy that will be used to monitor and evaluate its impact, quality, and outcomes. This will aid other organizations when considering how to plan and move toward sustainable group video visit practices as postpandemic health system planning becomes a priority.

Methods

Institutional Context

WCH is an ambulatory tertiary care facility in Toronto, Canada, and is fully affiliated with the University of Toronto. Housed within the hospital is the WCH Institute for Health System Solutions and Virtual Care, a "living laboratory" for developing, testing, and implementing virtual solutions to improve health [11]. In December 2019, the hospital launched Women's Virtual (WV) [12], an institutional strategic initiative to systematically address virtual care barriers and facilitate a coordinated and widespread adoption of virtual care across the organization, with the introduction of individual video visits.

WCH offers several evidence-informed group health interventions, which were historically delivered solely in person. The groups range from 10 to 25 patient participants, span an array of health concerns (eg, mental health, chronic pain management, and cardiac rehabilitation), and offer a variety of health interventions (eg, therapy, education, and exercise). Shortly after the ramp-down of nonemergent in-person health care visits in March 2020 at the pandemic's onset, 5 divisions across WCH identified the urgent need to ensure continuity of care by delivering these group interventions virtually. The eventual implementation of virtual group health interventions was part of the WV strategic roadmap, but these events increased its priority as an initiative for the hospital to develop in a manner that both addressed the immediate needs triggered by the pandemic and the long-term need for a program of sustainable virtual care.

By April 2020, because of its ability to replicate elements of in-person groups (eg, immediate and reciprocal visual cues, facial expressions, and the ability to support conversational dialogue), hospital leadership, in consultation with relevant departments and providers, determined that synchronous videoconferencing would be the best option to use in pursuit of the continued delivery of group health interventions.

System Development and Operationalization

In line with the hospital's strategic mandate, the development and implementation of a system to deliver synchronous video group health interventions to patients would need to also serve the larger interest of building a sustainable virtual program of care. Therefore, the use of either an existing platform (ie, Zoom [Zoom Video Communications Inc], a videoconferencing service that can be licensed for secure use for health care purposes) or an alternative hospital vendor either external to or integrated within the hospital's existing electronic medical record (EMR; ie, Epic [Epic Systems Corp]) was considered. A process to select a technical platform was conducted to support implementation that would meet security, privacy, and quality considerations beyond a pandemic context. Multiple factors were considered (Textbox 1). Stakeholders decided to move forward with integrated Epic-Zoom group video visits as a long-term sustainable solution that could continue to be used post pandemic. While this approach provided many benefits, there were few previous examples of Epic-Zoom integration for synchronous video group visits and therefore the implementation team had to overcome the following 3 key challenges: (1) developing technical integration protocols from the ground up, (2) developing a process for patient identity verification prior to group session admission, and (3) reducing friction points in the workflows for patients and providers.

Over approximately 6 weeks, beginning in April 2020, the WCH information technology (IT) team integrated Zoom with Epic in a manner that addressed all requirements, including seamless recording in the patient's EMR (to facilitate documentation of patient care), automated creation and dissemination of meeting links (to reduce administrative and clinical burden), and secure

patient identity verification. These factors would not have been possible if a nonintegrated system was used. A smooth administrative, clinical, and patient workflow was simultaneously developed (Figure 1). This echoed the scheduling workflow for individual video visits to reduce administrative burden and training requirements.

The first group video visit occurred on May 1, 2020, followed by a rapid 3-phased rollout to quickly scale group video visits across the hospital, allowing for iterative feedback and system improvements (Figure 2). Group moderators trained in the early stages of the rollout (with full support directly from the technical development team) acted as clinical champions to train their peers (using augmented training materials) by Phase 3. At the end of each phase, stakeholders conducted a debrief to determine the hospital's readiness to move to the next phase of rollout and what, if any, technical or workflow changes needed to be made beforehand. Improvements over time included technical enhancements for improved audio and video quality and the development of training materials (Phase 1), clinical workflow enhancements for easier scheduling and patient identity verification (Phase 2), and minor format modifications to some groups to allow for logical and housekeeping considerations in the first group session (Phase 3). Across 5 hospital divisions, in the 6 months following the launch of synchronous video groups (ie, May to November 2020), 29 groups were run, involving 542 individually scheduled group sessions conducted with 767 unique participants across the hospital, representing diverse health concerns, ages, and care needs.

Textbox 1. The factors considered in selecting a sustainable technological platform to support the implementation of synchronous videoconferencing for group health interventions.

1. Meets all legal and privacy regulations and best practices.
2. Supports group videoconferencing for ~25 participants.
3. Generates a single access link for moderators and patients.
4. Allows for the secure sharing of access links with moderators and patients.
5. Has admittance functionality.
6. Has a smooth log-in workflow for patients.
7. Allows for the documentation of completed appointments in the electronic medical record.

Figure 1. Administrative, clinical, and patient workflow for synchronous group video visits for health interventions, including scheduling, entering the group, and verifying patient identity.

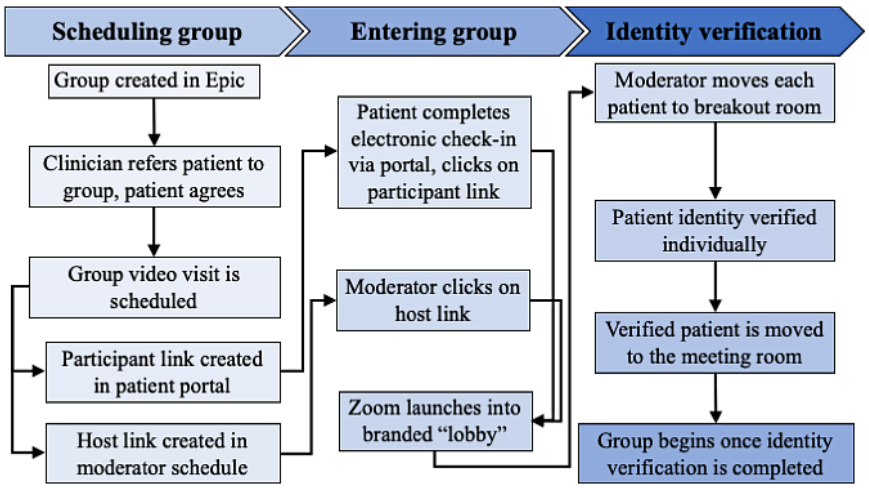
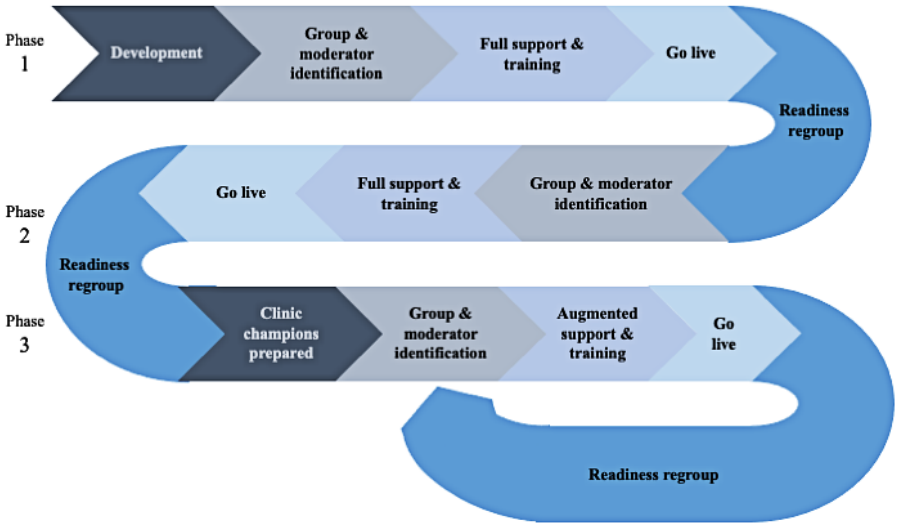


Figure 2. Overview of the 3-phased rollout of synchronous group video visits at Women's College Hospital.

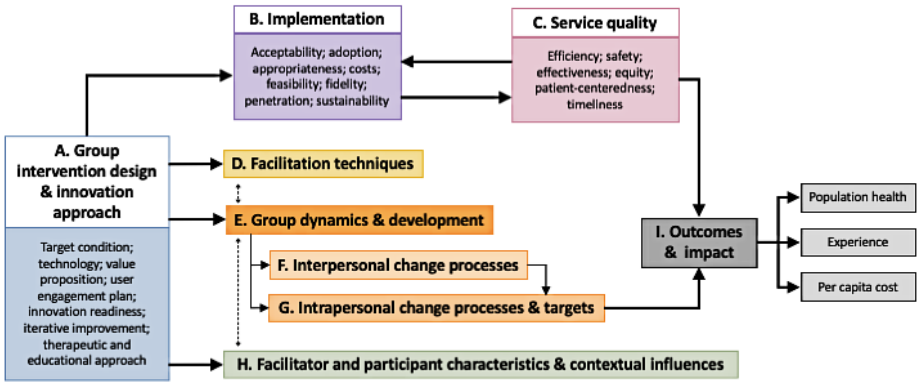


Evaluation Approach

Much has been documented related to the evaluation of individual video visits, but there is a significant gap in the literature related to the evaluation of the implementation and quality of synchronous video visits to deliver group-based health interventions, as well as group functioning in a virtual context. It is our view that appropriate evaluation is of critical importance to developing an evidence-based understanding of the role that these innovations can and should play to enhance the delivery of health care. Given the novelty of institutionally driven, Epic-integrated implementation of synchronous video group interventions, we developed the Virtual Group Evaluation Framework (Figure 3) to guide its short- and long-term evaluation at our institution; this framework could be considered for use by other institutions who similarly transition to system-integrated sustainable models of delivering group health interventions virtually. This framework integrates an adapted model from Proctor et al [13,14] that enables the evaluation of programs at multiple stages of their innovation lifecycle by considering measures of implementation (acceptability, adoption, appropriateness, feasibility, fidelity, cost, penetration, and sustainability), quality of care [15] (safe, effective, patient-centered, timely, efficient, and equitable) and impact (defined by the Quadruple Aim [16] as an improved experience of providing and receiving care, better population health, and reduced per capita cost of health care), and the mechanisms of action in group-based interventions (MAGI) framework [17,18], which outlines the theoretical mechanisms that lead to change in group-based health interventions. The MAGI framework was developed based on an extensive mixed-methods multidisciplinary literature review and primary research.

The Virtual Group Evaluation Framework can be operationalized using a mixed-methods approach to data collection and analysis, which is valuable and commonly used to evaluate both virtual and group interventions, which are inherently complex [1]. The following 3 layers of data collection can provide a comprehensive evaluation of each component of the framework: (1) hospital-based data (eg, patient records and IT services) to measure the innovation approach, as well as components of implementation, service quality, and outcomes (Multimedia Appendix 1); (2) quantitative data from patients (eg, based on self-report quality improvement surveys) and facilitators to measure individual characteristics, group dynamics, and components of implementation, service quality, and outcomes (Multimedia Appendix 2); and (3) qualitative interviews with key stakeholders (eg, patients, facilitators, referring clinicians, and hospital administrators) to fill in the remaining gaps in knowledge and provide a more holistic and contextual understanding of the experience. Each layer of data can be collected and analyzed alone or in combination with other layers of data depending on the specific research question being addressed by the institution or individual departments conducting their own evaluation projects. Where possible, specific validated tools were identified for quantitative data collection to ensure that the data collected address all components of the framework and allow for consistency and comparability across the variety of group interventions that take place throughout the hospital (Multimedia Appendix 1 and Multimedia Appendix 2). Simultaneously, we believe that this evaluation approach is flexible enough to remain receptive to the needs of different groups and patient populations that they serve.

Figure 3. The Virtual Group Evaluation Framework to guide evaluations of synchronous group video visits, adapted from Proctor et al [13,14] and Borek et al [17,18], and using definitions from the Institute of Medicine [15] and the Quadruple Aim [16].



Discussion

This paper describes the development and operationalization of group video visits to support the continued delivery of patient care in a virtual context, as well as a proposed strategy to guide the evaluation of their implementation, quality, and outcomes. To our knowledge, the experience of development and implementation of a clinically integrated and institutionally driven model of video-based group health interventions has not yet been described in the literature.

There are several key learnings that can be used to help inform similar undertakings by other institutions considering a similar approach. First, our focus in developing Epic-integrated synchronous video group visits was centered on the value proposition it could offer beyond the immediate context of COVID-19–related restrictions to care delivery. Therefore, we centered our efforts on addressing the needs of "tools, teams, and routines" to make meaningful improvements to care and quality of care [19]. This extended beyond the tool itself, to the design of administrative, clinical, and patient workflows, with rapid iterations at each stage of the phased rollout to make incremental improvements as feedback was received to ensure its long-term sustainable integration into the way group-based care is delivered at our institution.

Second, beyond ensuring patient care when in-person services are disrupted, it is vitally important to operationalize the virtual delivery of health care alongside approaches for evaluating its implementation, quality, and impact. Therefore, the development of an evaluation strategy that specifically considers the nuances of both virtual and group care will allow us to understand whether high-quality care that improves outcomes continues to be delivered to patients. The evaluation strategy developed as part of this work can be considered by other organizations pursuing similar innovations related to the delivery of virtual group health interventions and seeking to incorporate evaluation strategies. Our Virtual Group Evaluation Framework provides a theoretically driven yet practical way to support the short- and long-term evaluation of synchronous virtual group health interventions. It provides flexibility and adaptability to suit a variety of contexts (eg, group types, institutional capacities, and resources). Further, the evaluation strategy incorporates considerations of equity because although virtual health interventions may address some barriers to care, they simultaneously have the potential to exacerbate issues of inequity related to delivery and access. There are published examples of conceptually driven approaches for individual video visits [20], but none to our knowledge that consider the nuances of group health interventions when delivered in a virtual environment.

The limitations of this work should be noted. This paper describes the operationalization of group video visits in a tertiary care center with existing familiarity with video clinical visits and a high focus on innovation; therefore, this experience may not be broadly applicable to all centers, particularly those where the use of virtual care is more novel. However, we still feel that our experience in developing and operationalizing group video visits can be used to provide a guide for institutions looking to introduce such strategies. While the MAGI framework, which guides the evaluation of group process outcomes, has not yet been widely used, given its recent publication, nor has it been applied to group health interventions in a virtual setting, this work will add to the literature to address these gaps. Finally, the Virtual Group Evaluation Framework has not yet been used or validated in clinical virtual group interventions. Therefore, its utility and generalizability in this context need to be determined prior to more widespread use by other organizations. This will be explored as the framework is applied in practice; evaluation efforts at WCH that apply the use of this framework are underway.

Future work includes conducting evaluations using the Virtual Group Evaluation Framework to understand outcomes and gain a better understanding of group functioning in a virtual environment. Groups also need to be optimized to support the delivery of different types of group health interventions, such as movement-based groups (eg, exercise or yoga). Additionally, the experience of group health interventions through synchronous videoconferencing and results of evaluations should be used when considering how this type of care delivery fits into the postpandemic health system. If we can demonstrate its positive impact, it can be used to deliver care to those who might not otherwise be able to access the types of specialized care that our center provides due to geographic, social, or medical isolation. For example, some cardiac rehabilitation groups at WCH were able to operate over the summer months, a time when their target patient population may not be able to travel to the hospital to receive care due to high heat, which can exacerbate certain cardiac conditions and symptoms.

The rapid operationalization of virtual care has been a consistent feature in health care delivery during the COVID-19 pandemic, and one that is likely to have enduring effects on how care is delivered. The experience of our center in systematically operationalizing and delivering EMR-integrated synchronous video group health interventions is one such example that emerged based on need but was designed in a manner that has the potential to enhance how virtual care can be scaled and sustained. Its evaluation will provide the data needed to support its place in a postpandemic health system.

Acknowledgments

The authors gratefully acknowledge the contributions of Laura Pus (WV), Dhara Desai (IT), and members of the WCH Group Video Visits Subcommittee.

There was no direct funding provided for this study. The development of the intervention described in this manuscript was funded directly by in-kind operational resources from the Women's College Hospital (WCH). WCH had no role in the writing of the manuscript nor the decision to submit it for publication.

Conflicts of Interest

SNV receives royalties from UpToDate Inc for materials on perinatal depression.

Multimedia Appendix 1
The operational approach to evaluating synchronous video group health interventions using hospital-based data.
[DOCX File , 26 KB - jmir_v24i4e29841_app1.docx]

Multimedia Appendix 2
The operational approach to evaluating synchronous video group health interventions using self-reported quantitative data.
[DOCX File , 25 KB - jmir_v24i4e29841_app2.docx]

References

1. Biggs K, Hind D, Gossage-Worrall R, Sprange K, White D, Wright J, et al. Challenges in the design, planning and implementation of trials evaluating group interventions. Trials 2020 Jan 29;21(1):116 [FREE Full text] [doi: 10.1186/s13063-019-3807-4] [Medline: 31996259]
2. Okumura Y, Ichikura K. Efficacy and acceptability of group cognitive behavioral therapy for depression: a systematic review and meta-analysis. J Affect Disord 2014 Aug;164:155-164. [doi: 10.1016/j.jad.2014.04.023] [Medline: 24856569]
3. Barkowski S, Schwartze D, Strauss B, Burlingame GM, Rosendahl J. Efficacy of group psychotherapy for anxiety disorders: a systematic review and meta-analysis. Psychother Res 2020 Nov;30(8):965-982. [doi: 10.1080/10503307.2020.1729440] [Medline: 32093586]
4. Oei TP, Dingle G. The effectiveness of group cognitive behaviour therapy for unipolar depressive disorders. J Affect Disord 2008 Apr;107(1-3):5-21. [doi: 10.1016/j.jad.2007.07.018] [Medline: 17716745]
5. Singer J, Levy S, Shimon I. Group versus individual care in patients with long-standing type 1 and type 2 diabetes: a one-year prospective noninferiority study in a tertiary diabetes clinic. J Diabetes Res 2018;2018:1807246 [FREE Full text] [doi: 10.1155/2018/1807246] [Medline: 29998139]
6. Parry M, Bjørnnes AK, Victor JC, Ayala AP, Lenton E, Clarke H, et al. Self-management interventions for women with cardiac pain: a systematic review and meta-analysis. Can J Cardiol 2018 Apr;34(4):458-467. [doi: 10.1016/j.cjca.2017.12.011] [Medline: 29477931]
7. Green S, Haber E, Frey B, McCabe R. Cognitive-behavioral group treatment for perinatal anxiety: a pilot study. Arch Womens Ment Health 2015 Aug;18(4):631-638. [doi: 10.1007/s00737-015-0498-z] [Medline: 25652951]
8. Joyce AS, Tasca GA, Ogrodniczuk JS. Group psychotherapy in Canada. Int J Group Psychother 2015 Oct;65(4):583-593. [doi: 10.1521/ijgp.2015.65.4.583] [Medline: 26401801]
9. Banbury A, Nancarrow S, Dart J, Gray L, Parkinson L. Telehealth interventions delivering home-based support group videoconferencing: systematic review. J Med Internet Res 2018 Feb 02;20(2):e25 [FREE Full text] [doi: 10.2196/jmir.8090] [Medline: 29396387]
10. Birrell F, Lawson R, Sumego M, Lewis J, Harden A, Taveira T, et al. Virtual group consultations offer continuity of care globally during Covid‐19. Lifestyle Med 2020 Dec 20;1(2):e17. [doi: 10.1002/lim2.17]
11. Home. Women's College Hospital Institute for Health System Solutions and Virtual Care. URL: https://www.wchwihv.ca [accessed 2022-03-18]
12. Women's Virtual. Women's College Hospital. URL: https://www.womenscollegehospital.ca/care-programs/womens-virtual/ [accessed 2022-03-18]

13. Proctor EK, Landsverk J, Aarons G, Chambers D, Glisson C, Mittman B. Implementation research in mental health services: an emerging science with conceptual, methodological, and training challenges. Adm Policy Ment Health 2009 Jan;36(1):24-34 [FREE Full text] [doi: 10.1007/s10488-008-0197-4] [Medline: 19104929]
14. 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]
15. Institute of Medicine (US) Committee on Quality of Health Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academies Press; 2001.
16. Sikka R, Morath JM, Leape L. The Quadruple Aim: care, health, cost and meaning in work. BMJ Qual Saf 2015 Oct;24(10):608-610. [doi: 10.1136/bmjqs-2015-004160] [Medline: 26038586]
17. Borek A, Abraham C, Greaves C, Gillison F, Tarrant M, Morgan-Trimmer S, et al. Identifying change processes in group-based health behaviour-change interventions: development of the mechanisms of action in group-based interventions (MAGI) framework. Health Psychol Rev 2019 Jun 13:1-21. [doi: 10.1080/17437199.2019.1625282] [Medline: 31190619]
18. Borek AJ, Smith JR, Greaves CJ, Gillison F, Tarrant M, Morgan-Trimmer S, et al. Developing and applying a framework to understand mechanisms of action in group-based, behaviour change interventions: the MAGI mixed-methods study. Efficacy Mech Eval 2019;6(3):1-162. [doi: 10.3310/eme06030] [Medline: 31233299]
19. Shaw J, Agarwal P, Desveaux L, Palma DC, Stamenova V, Jamieson T, et al. Beyond "implementation": digital health innovation and service design. NPJ Digit Med 2018;1:48 [FREE Full text] [doi: 10.1038/s41746-018-0059-8] [Medline: 31304327]
20. Srinivasan M, Phadke AJ, Zulman D, Israni ST, Madill ES, Savage TR, et al. Enhancing patient engagement during virtual care: a conceptual model and rapid implementation at an academic medical center. NEJM Catalyst. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.20.0262 [accessed 2022-03-18]

Abbreviations

EMR: electronic medical record
IT: information technology
MAGI: mechanisms of action in group-based interventions
WCH: Women's College Hospital
WV: Women's Virtual

Edited by R Kukafka; submitted 22.04.21; peer-reviewed by R Krukowski, D West; comments to author 20.05.21; revised version received 24.06.21; accepted 17.12.21; published 07.04.22.

Please cite as:
Hussain-Shamsy N, McMillan I, Cook S, Furfaro-Argier A, Sadler A, Delos-Reyes F, Wasserman L, Bhatia S, Martin D, Seto E, Vigod SN, Zaheer J, Agarwal P, Mukerji G
Operationalizing and Evaluating Synchronous Virtual Group Health Interventions: Wide-Scale Implementation at a Tertiary Care Academic Hospital
J Med Internet Res 2022;24(4):e29841
URL: https://www.jmir.org/2022/4/e29841
doi:10.2196/29841
PMID:35389350

Viewpoint

Social Media, Public Health, and Community Mitigation of COVID-19: Challenges, Risks, and Benefits

Corey H Basch[1], MPH, EdD; Charles E Basch[2], PhD; Grace C Hillyer[3], MPH, EdD; Zoe C Meleo-Erwin[4], PhD

[1]Department of Public Health, William Paterson University, Wayne, NJ, United States

[2]Department of Health and Behavior Studies, Teachers College, Columbia University, New York, NY, United States

[3]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, United States

[4]see Acknowledgments

Corresponding Author:
Corey H Basch, MPH, EdD
Department of Public Health
William Paterson University
300 Pompton Rd
Wayne, NJ, 07470
United States
Phone: 1 9737202603
Email: baschc@wpunj.edu

Abstract

Shortly after the first case reports in 2019, COVID-19 was declared a pandemic. Early messages from trusted experts, which later proved to be inadequate or incorrect, highlight the need for continual adjustment of messages to the public as scientific knowledge evolves. During this time, social media exploded with greatly sought-after information, some of which was misinformation based on incomplete or incorrect facts or disinformation purposefully spread to advance a specific agenda. Because of the nature of social media, information, whether accurate or not at the time posted, lives on and remains accessible to the public even when its usefulness has been discredited. While the impact of mis/disinformation on COVID-19 risk-reducing behaviors is debatable, it is clear that social media has played a significant role in both extending the reach of COVID-19–related falsehoods and promoting evidence-based content. Over the last decade, social media has become a dominant source of information that consumers turn to for health information. A great deal of misinformation and disinformation has reached large numbers of social media users, which points to a need for the agencies of the US Public Health Service to create communications to convey accurate and current information and appeals that will actually be viewed. This viewpoint highlights the challenges, risks, and potential benefits that social media present in mitigating the COVID-19 pandemic.

(J Med Internet Res 2022;24(4):e36804) doi:10.2196/36804

KEYWORDS

COVID-19 pandemic; social media; misinformation; disinformation; COVID-19; pandemic; infodemiology; health literacy; health information; public health; COVID risk; information seeking

Background

At the close of 2019, SARS-CoV-2, the virus that causes COVID-19, emerged in Wuhan, China, and rapidly spread throughout the world, being labeled a pandemic in only a few months after its emergence [1,2]. Nearly 2 years after the initial outbreak, as of April 8, 2022, there have been 496,355,574 confirmed cases of COVID-19, 6,170,720 cumulative deaths, and 11,085,254,518 vaccine doses administered globally [3]. The Omicron variant is also circulating globally, quickly driving up cases in regions that had previously seen reprieve owing in part to high rates of vaccination [4,5].

The only way to prevent COVID-19 from occurring is through both reducing exposure to SARS-CoV-2 and reducing susceptibility. Since COVID-19 is primarily transmitted through air droplets and aerosols [6,7], the main strategies for reducing exposure are social distancing [8], mask use [9,10], and limiting time spent in poorly ventilated spaces where air is shared with other people [11,12]. The main strategy for reducing susceptibility is vaccination [13,14]. With respect to COVID-19 prevention, individuals' personal decision-making about behaviors to reduce their susceptibility and/or exposure influences risk at both the individual and community-wide levels [15,16].

Individuals' health and illness–related decision-making is affected by many different sources of influence such as family members and friends, health care providers, political leaders, and the media. A plethora of behavioral science theories demonstrate how individuals' levels of knowledge and beliefs, skills, availability and accessibility to resources, and social support shape health choices [17-21]. Over the past decade, examination of the role played by social media in personal decisions that affect individual, family, and population health, including those related to infectious disease, has expanded rapidly. This viewpoint highlights the challenges, risks, and potential benefits that social media present in mitigating the COVID-19 pandemic.

Challenges

Airborne transmission of droplets and aerosols is the main way SARS-CoV-2 is transmitted [6,7]. Secondarily, although far less frequent, the virus can also be transmitted through droplets and aerosols settling on fomites (inanimate objects) and having hand/nose or hand/mouth contact [22]. In the early stage of the pandemic, emphasis was placed on transmission through fomites, and on handwashing and surface disinfection, as well as social distancing. Yet, little to no emphasis was placed on masking. Messaging about fomite transmission was eventually dialed back [23,24], while messaging on mask recommendations was altered entirely.

In times of global and national crisis, the public clamors for the guidance of trusted experts to navigate rapidly evolving and potentially life-threatening situations. Early on, the experts at Centers for Disease Control and Prevention (CDC), the World Health Organization (WHO), and the US Surgeon General blasted messages that subsequently required modification. False messages of conspiracy rapidly surfaced, and charlatans promoted unsubstantiated claims to prevent disease. To some degree, future expert messages were discredited.

For example, on February 29, 2020, US Surgeon General Dr Jerome Adams used Twitter to send the following message to the public: "Seriously people- STOP BUYING MASKS! They are NOT effective in preventing general public from catching #Coronavirus, but if healthcare providers can't get them to care for sick patients, it puts them and our communities at risk!" [25]. This sentiment was in concert with the early mentality of the CDC, WHO, US Office of the Surgeon General, and other constituents at the time.

All of these entities would later reverse their views as further scientific evidence emerged regarding the ambient transmission of SARS CoV-2 [26-30]. With spread of the Delta variant during the summer of 2021, CDC messaging encouraged unvaccinated individuals to wear masks in crowded public places and for vaccinated individuals to wear masks indoors in areas of high transmission [31]. When the Omicron variant began to spread in late 2021, news reports began appearing with quotes from health experts advising the public to upgrade their masks [32-34]. However, the official CDC guidance on masking was not clarified until mid-January 2022 [31].

In a context of rapidly evolving scientific discovery, a key challenge for public health education is to adapt messaging in an information environment containing older content that continues to live and circulate online despite being obsolete or inaccurate from the outset.

A second major challenge is that public health experts may (and often do) disagree about recommendations. Communicating the nuances of scientific progress may be especially hampered by the fact that nearly 60% of the population has medium or low scientific literacy [35]. Given the uncertainty involved in an emerging global pandemic, the incremental and often contentious way in which scientific knowledge typically advances, and shifting and diverging recommendations, it is not surprising that public confusion is common and trust in official governmental agencies is hindered [35].

Risks

Large numbers of individuals search for health information online [36,37]. As with previous infectious diseases [38], the public turned to the internet for information on the novel coronavirus [39-41]. Unfortunately, false information about the COVID-19 pandemic on social media has been extensive [42-44] and viewed by millions of people [45]. False information involves both misinformation, or inaccurate conclusions about a phenomenon drawn from incomplete or incorrect facts, as well as disinformation, which pertains to the purposeful spread of false information in line with a specific agenda [46]. The early days of the pandemic saw such a proliferation of false information online, leading the WHO Director General to declare that the fight was not only against the pandemic but against an "infodemic" as well [47]. An "infodemic" can be described as "an overabundance of information – some accurate and some not – that makes it hard for people to find trustworthy sources and reliable guidance when they need it" [47].

Preventing COVID-19 involves both reducing exposure to SARS-CoV-2 through masking, social distancing, and the avoidance of crowded, poorly ventilated spaces, as well as through reducing susceptibility via vaccination. In the United States, there are two types of vaccines available: mRNA vaccines and an adenovirus-based vaccine [48,49]. It has become clear that one of the concerns that the public has about the vaccine pertains to the comparatively short time spent on development, testing, and approval (emergency use authorization). Officials sought to assure the public that neither safety nor scientific integrity was compromised in the process [50-52]. Nevertheless, public concern that mRNA would alter one's genes and fear surrounding an untested technology persisted [53,54]. Surveys in the United States about attitudes toward vaccination (whether for oneself or one's children) continue to reveal persistent levels of hesitancy and refusal [55], with notable demographic and political differences [56-60].

The trustworthiness of online information about COVID-19 vaccination is endangered by those whose aim is to spread disinformation. Disinformation can be spread by anyone. One estimate was that 12 individuals were responsible for 65% of the mis/disinformation spread on Twitter, Instagram, and Facebook [61]. These individuals were identified as

"anti-vaccine activists, alternative health entrepreneurs and physicians" [61] who spread a wealth of antivaccination content over the course of the pandemic [62]. The danger of mis/disinformation is epitomized by its use "to incite violence and crime targeted at ethnic minorities – which has resulted in deaths and displacement of children, led to lower child COVID vaccination rates, undermined trust in journalism and science, and drowned out marginalized voices" [63]. Mis/disinformation in times of health emergencies threatens national security and there have been calls for immediate intervention [64]; however, this is not easily accomplished in a societal context with freedom of speech and access to digital communication channels capable of reaching millions of people in minutes.

A small number of social media accounts spreading disinformation have had widespread reach and have gained many followers during the pandemic [62]. To the extent that this small number of antivaccination messengers attract many users who are skeptical or are "on the fence" about vaccination, they can undermine efforts to mitigate community transmission and increase the risk of more pathogenic and virulent variants.

The influence of misinformation and disinformation on acceptance of inaccurate COVID-19 information, behavioral intentions, and actual behavior remain equivocal [65-67]. Some evidence suggests that such communications are associated with acceptance of COVID-19 conspiracy theories [65,68] and other inaccurate information [69]. A belief in conspiracy theories may reduce the uptake of at least some preventative behaviors [70,71].

In sum, while the behavioral impact of misinformation and disinformation on social media is debated, what is clear is the role that social media has played in extending the reach of COVID-19–related falsehoods. Of course, even if low-quality or blatantly untrue COVID-19 online information has a minimal impact on COVID-19 exposure and susceptibility reduction, in the case of a highly contagious respiratory virus, the impact on community spread may nevertheless be of considerable concern.

Benefits of a Multifaceted Public Health Presence on Social Media

It has been well-established that social media can spread accurate, useful information. During the pandemic, social media companies have claimed to make efforts to remove harmful COVID-19 content and promote accurate information [72-74], although the degree to which they have been successful in doing so has been questioned [75]. Previous investigations have demonstrated examples of how social media have promoted positive behaviors such as the Global Handwashing Partnership [76]. Some research suggests that although misleading COVID-19 information has been widely shared on social media, evidence-based content may have been shared to a greater extent. Fortunately, public health agencies such as the WHO and CDC have increased their presence on key social media platforms [77]. Beyond creating their own content, public health agencies have begun to look to social media personalities to help disseminate prevention messages. For example, in 2020, the state of Texas moved to pay influencers to promote mask use and social distancing on TikTok [78].

We previously identified a shift in popular YouTube videos regarding the source of messages on COVID-19 from news media [79] to entertainers [80]. From January to March 2020, there was a large increase in the number and proportion of cumulative views garnered by widely viewed videos on hand hygiene, from 33,268,243 (26.6%) to 182,331,135 (51.3%) [80], and there was an increase coverage of face masks in the most popular videos [80]. Our work examining mechanisms of disease transmission across three successive samples indicated an increase from ~63 million views to more than 273 million views for videos specifically mentioning disease transmission. This increase coincided with a rise in the worldwide number of cases and the occurrence of COVID-19 transmission [81]. A Medline search performed at the time of writing this viewpoint indicated that these are among the first successive samples of YouTube videos published to date. Our studies suggest that tracking changes over time is useful and that social media may help to amplify scientific recommendations. Notably, successive sampling captures evolving situations rather than labeling something as misinformation simply because it was posted at a time when science about the topic was evolving.

Similar to YouTube, there is active conversation about COVID-19 on TikTok. Earlier research suggested that, despite 1 billion views with the #Coronavirus hashtag, there was minimal discussion about transmission [82]. However, with more directed searching, we identified messaging about prevention [83]. Notably, a search of the hashtag #WearAMask indicated that the number of views of TikTok videos mentioning masks uploaded by the WHO was far less than those in the "Wear a Mask" campaign, despite the fact that the messages from the WHO were more aligned with scientific recommendations [6].

Humor, music, and dance were large components of the #WearAMask posts, but were not common in those from the WHO [83]. Consistent with the study on YouTube, these findings suggest that popular entertainers and the use of nonconventional approaches to messaging can help draw attention to public health messages. While conversations on TikTok could be health-promoting (eg, using masks, social distancing, testing, handwashing) [83-85], they may also be counterproductive as was noted in our study of misinformation related to vaccination [86]. Thus, while there are challenges with and cause for concern regarding COVID-19 content on social media, social media clearly play a vital role in disseminating accurate information.

Moving Forward

Public health communication best practices suggest that in cases where science is not definitive, messages should be calibrated to make it clear that there is an evolving context [87]. Our work indicates that the content of official public health agency social media accounts does not receive as many views as communications posted by popular entertainers, influencers, or, in some cases, even consumers. Partnerships with carefully vetted content creators may help to extend the reach of accurate

health information on social media, particularly to young people who tend to use these platforms in higher numbers than older adults [88].

Because there is a propensity for research highlighting misinformation as a means to gain attention [89,90], the dialog should be clear about what constitutes misinformation. Researchers should consider the date that material was posted and how scientific understanding changed over time, rather than aim to produce "clickbait" titles. There are degrees of misinformation and calibrating messages with a disclaimer will also help in this circumstance. Studies claiming they have identified misinformation should be forthcoming about how this was coded and what factors were taken into consideration, with concrete examples of what did and did not constitute misinformation and disinformation.

Effective communication and education should be geared toward a very heterogenous public in terms of age, beliefs, culture, sociodemographics, literacy, and information-seeking sources. Over the last decade, social media have become a dominant source of information that consumers turn to for health and COVID-19 information. This situation prompted our early work to determine what is and is not being communicated. The findings to date suggest there is considerable room for improvement. A great deal of misinformation and disinformation has reached large numbers of social media users, pointing to a need for the agencies of the US Public Health Service to create communications to convey accurate and current information and appeals that will actually be viewed.

Acknowledgments

ZCME is an Independent Researcher in New York, United States.

Conflicts of Interest

CHB serves as an Editorial Board Member for JMIR; she did not have a role in the review or editorial process for this article.

References

1. Wu Y, Chen C, Chan Y. The outbreak of COVID-19: an overview. J Chin Med Assoc 2020 Mar;83(3):217-220 [FREE Full text] [doi: 10.1097/JCMA.0000000000000270] [Medline: 32134861]
2. WHO Director-General's opening remarks at the media briefing on COVID-19. World Health Organization. 2020 Mar 11. URL: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2021-02-28]
3. Coronavirus disease (COVID-19) pandemic. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2021-02-28]
4. Araf Y, Akter F, Tang Y, Fatemi R, Parvez MSA, Zheng C, et al. Omicron variant of SARS-CoV-2: Genomics, transmissibility, and responses to current COVID-19 vaccines. J Med Virol 2022 May 23;94(5):1825-1832. [doi: 10.1002/jmv.27588] [Medline: 35023191]
5. Smith Rogers L. John Hopkins Bloomberg School of Mental Health. 2022 Jan 07. URL: https://publichealth.jhu.edu/2022/omicron-qa-making-some-sense-of-the-messiness-of-this-moment [accessed 2022-01-19]
6. Framework for implementation of COVID-19 community mitigation measures for lower-resource countries. Centers for Disease Control and Prevention. 2021 Jul 20. URL: https://www.cdc.gov/coronavirus/2019-ncov/global-covid-19/community-mitigation-measures.html [accessed 2022-01-19]
7. Tang S, Mao Y, Jones RM, Tan Q, Ji JS, Li N, et al. Aerosol transmission of SARS-CoV-2: evidence, prevention and control. Environ Int 2020 Nov;144:106039 [FREE Full text] [doi: 10.1016/j.envint.2020.106039] [Medline: 32822927]
8. Morawska L, Milton DK. It is time to address airborne transmission of Coronavirus Disease 2019 (COVID-19). Clin Infect Dis 2020 Dec 03;71(9):2311-2313 [FREE Full text] [doi: 10.1093/cid/ciaa939] [Medline: 32628269]
9. Centers for Disease Control and Prevention. 2021 Nov 29. URL: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/social-distancing.html [accessed 2022-01-19]
10. Van Dyke ME, Rogers TM, Pevzner E, Satterwhite CL, Shah HB, Beckman WJ, et al. Trends in county-level COVID-19 incidence in counties with and without a mask mandate - Kansas, June 1-August 23, 2020. MMWR Morb Mortal Wkly Rep 2020 Nov 27;69(47):1777-1781. [doi: 10.15585/mmwr.mm6947e2] [Medline: 33237889]
11. Scientific Brief: Community use of cloth masks to control the spread of SARS-CoV-2. Centers for Disease Control and Prevention. 2021 Dec 06. URL: https://www.cdc.gov/coronavirus/2019-ncov/more/masking-science-sars-cov2.html [accessed 2022-01-19]
12. Ventilation in buildings. Centers for Disease Control and Prevention. 2021 Jun 02. URL: https://www.cdc.gov/coronavirus/2019-ncov/community/ventilation.html [accessed 2022-01-19]
13. Morawska L, Tang JW, Bahnfleth W, Bluyssen PM, Boerstra A, Buonanno G, et al. How can airborne transmission of COVID-19 indoors be minimised? Environ Int 2020 Sep;142:105832 [FREE Full text] [doi: 10.1016/j.envint.2020.105832] [Medline: 32521345]

14. Hodgson SH, Mansatta K, Mallett G, Harris V, Emary KRW, Pollard AJ. What defines an efficacious COVID-19 vaccine? A review of the challenges assessing the clinical efficacy of vaccines against SARS-CoV-2. Lancet Infect Dis 2021 Feb;21(2):e26-e35 [FREE Full text] [doi: 10.1016/S1473-3099(20)30773-8] [Medline: 33125914]
15. Centers for Disease Control and Prevention. 2021 Dec 01. URL: https://www.cdc.gov/vaccines/covid-19/index.html [accessed 2022-01-19]
16. Fischhoff B. Making decisions in a COVID-19 world. JAMA 2020 Jul 14;324(2):139-140. [doi: 10.1001/jama.2020.10178] [Medline: 32496560]
17. Norheim O, Abi-Rached J, Bright L, Bærøe K, Ferraz OLM, Gloppen S, et al. Difficult trade-offs in response to COVID-19: the case for open and inclusive decision making. Nat Med 2021 Jan;27(1):10-13. [doi: 10.1038/s41591-020-01204-6] [Medline: 33340033]
18. Green L, Kreuter M. Health program planning: An educational and ecological approach. 4th edition. New York, NY: McGraw-Hill; 2005.
19. Prochaska J, Redding C, Evers K. The transtheoretical model stages of change. In: Glanz K, Rimer BK, Viswanath K, editors. Health behavior: theory, research, and practice. 3rd edition. San Francisco, CA: Jossey-Bass; 2002:97-123.
20. Rosenstock IM. Historical origins of the health belief model. Health Educ Monograph 1974 Dec 01;2(4):328-335. [doi: 10.1177/109019817400200403]
21. Rogers E. Diffusion of innovations, 5th edition. New York, NY: Simon and Schuster; 2003.
22. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. Annu Rev Public Health 2010;31:399-418. [doi: 10.1146/annurev.publhealth.012809.103604] [Medline: 20070207]
23. How COVID-19 spreads. Centers for Disease Control and Prevention. 2021 Jul 14. URL: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-covid-spreads.html [accessed 2022-01-20]
24. Goldman E. Exaggerated risk of transmission of COVID-19 by fomites. Lancet Infect Dis 2020 Aug;20(8):892-893 [FREE Full text] [doi: 10.1016/S1473-3099(20)30561-2] [Medline: 32628907]
25. CDC updates COVID-19 transmission webpage to clarify information about types of spread. Centers for Disease Control and Prevention. 2020 May 23. URL: https://www.cdc.gov/media/releases/2020/s0522-cdc-updates-covid-transmission.html [accessed 2022-01-20]
26. Jingnan H. Why there are so many different guidelines for face masks for the public. NPR Goats and Soda. 2020 Apr 10. URL: https://www.npr.org/sections/goatsandsoda/2020/04/10/829890635/why-there-so-many-different-guidelines-for-face-masks-for-the-public?t=1642795402943 [accessed 2022-01-20]
27. Yu IT, Li Y, Wong TW, Tam W, Chan AT, Lee JH, et al. Evidence of airborne transmission of the severe acute respiratory syndrome virus. N Engl J Med 2004 Apr 22;350(17):1731-1739. [doi: 10.1056/nejmoa032867]
28. Friedman U. Face masks are in. The Atlantic. 2020 Apr 02. URL: https://www.theatlantic.com/politics/archive/2020/04/america-asia-face-mask-coronavirus/609283/ [accessed 2022-01-20]
29. Felter C, Bussemaker N. Which countries are requiring face masks? Council on Foreign Relations. 2020 Aug 04. URL: https://www.cfr.org/in-brief/which-countries-are-requiring-face-masks [accessed 2022-01-20]
30. Manjoo F. How the world's richest country ran out of a 75-cent face mask. The New York Times. 2020 Mar 25. URL: https://www.nytimes.com/2020/03/25/opinion/coronavirus-face-mask.html [accessed 2022-01-20]
31. Types of masks and respirators. Centers for Disease Control and Prevention. 2022 Jan 14. URL: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/types-of-masks.html [accessed 2022-01-20]
32. Khaled F. Should Americans 'upgrade' face coverings to surgical masks amid Omicron spread? Newsweek. 2021 Dec 2. URL: https://www.newsweek.com/should-americans-upgrade-face-coverings-surgical-masks-amid-omicron-spread-1655505 [accessed 2022-01-20]
33. Caron C. Is It safe to travel? Can I still see my family on Christmas? New York Times. 2021 Feb 17. URL: https://www.nytimes.com/2021/12/17/well/live/omicron-holiday-travel-parties-testing.html [accessed 2022-01-20]
34. Godoy M, Doucleff M, Wroth C. Don't let omicron crash your holiday gathering. Here's how to keep your family safe. NPR. 2021 Dec 20. URL: https://www.npr.org/sections/health-shots/2021/12/04/1061069640/omicron-risk-holiday-travel [accessed 2022-01-20]
35. Kennedy B, Hefferon M. What Americans know about science. Pew Research Center. 2019 Mar 28. URL: https://www.pewresearch.org/science/2019/03/28/what-americans-know-about-science/ [accessed 2022-01-20]
36. Fox S. The social life of health information. Pew Research Center. 2014 Jan 15. URL: https://www.pewresearch.org/fact-tank/2014/01/15/the-social-life-of-health-information/ [accessed 2022-01-20]
37. The internet and health. Pew Research Center. 2013 Feb 12. URL: https://www.pewresearch.org/internet/2013/02/12/the-internet-and-health/ [accessed 2022-01-20]
38. Fidler DP. Disinformation and disease: social media and the Ebola epidemic in the Democratic Republic of the Congo. Council on Foreign Relations. 2019 Aug 20. URL: https://www.cfr.org/blog/disinformation-and-disease-social-media-and-ebola-epidemic-democratic-republic-congo [accessed 2022-01-24]
39. Timoneda J, Vallejo Vera S. Will I die of coronavirus? Google Trends data reveal that politics determine virus fears. PLoS One 2021;16(10):e0258189 [FREE Full text] [doi: 10.1371/journal.pone.0258189] [Medline: 34614032]

40. Cinarka H, Uysal M, Cifter A, Niksarlioglu EY, Çarkoğlu A. The relationship between Google search interest for pulmonary symptoms and COVID-19 cases using dynamic conditional correlation analysis. Sci Rep 2021 Jul 13;11(1):14387. [doi: 10.1038/s41598-021-93836-y] [Medline: 34257381]
41. O'Leary D, Storey V. A Google–Wikipedia–Twitter model as a leading indicator of the numbers of coronavirus deaths. Intell Sys Acc Fin Mgmt 2020 Sep 28;27(3):151-158. [doi: 10.1002/isaf.1482]
42. 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]
43. Carnegie Mellon University. 2020 Apr 16. URL: https://www.cmu.edu/ideas-social-cybersecurity/research/coronavirus.html [accessed 2022-01-20]
44. Hamel L, Lopes L, Kirzinger A, Sparks G, Stokes M, Brodie M. KFF COVID-19 Vaccine Monitor: media and misinformation. Kaiser Family Foundation. 2021 Nov 08. URL: https://www.kff.org/coronavirus-covid-19/poll-finding/kff-covid-19-vaccine-monitor-media-and-misinformation/ [accessed 2022-01-20]
45. Obiała J, Obiała K, Mańczak M, Owoc J, Olszewski R. COVID-19 misinformation: accuracy of articles about coronavirus prevention mostly shared on social media. Health Policy Technol 2021 Mar;10(1):182-186 [FREE Full text] [doi: 10.1016/j.hlpt.2020.10.007] [Medline: 33163352]
46. Igoe KJ. Establishing the truth: vaccines, social media, and the spread of misinformation. Harvard TH Chan School of Public Health. URL: https://www.hsph.harvard.edu/ecpe/vaccines-social-media-spread-misinformation/ [accessed 2022-01-18]
47. Understanding the infodemic and misinformation in the fight against COVID-19. Pan American Health Organization. URL: https://iris.paho.org/bitstream/handle/10665.2/52052/Factsheet-infodemic_eng.pdf?sequence=16 [accessed 2022-03-18]
48. Different COVID-19 vaccines. Centers for Disease Control and Prevention. 2022 Jan 21. URL: https://www.cdc.gov/coronavirus/2019-ncov/vaccines/different-vaccines.html [accessed 2022-01-21]
49. FDA issues emergency use authorization for third COVID-19 vaccine. US Food and Drug Administration. 2021 Feb 27. URL: https://www.fda.gov/news-events/press-announcements/fda-issues-emergency-use-authorization-third-covid-19-vaccine [accessed 2022-01-21]
50. Maragakis L, Kelen GB. Is the COVID-19 vaccine safe? Johns Hopkins Medicine. 2022 Jan 04. URL: https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/is-the-covid19-vaccine-safe [accessed 2022-01-21]
51. Owens C. Why the coronavirus vaccine rollout is behind schedule. Axios. 2021 Jan 06. URL: https://www.axios.com/coronavirus-vaccine-rollout-delays-a1de3fb6-4b13-4c1f-9647-de6d7c1f812c.html [accessed 2022-01-21]
52. Historical vaccine safety concerns. Centers for Disease Control and Prevention. 2020 Sep 04. URL: https://www.cdc.gov/vaccinesafety/concerns/concerns-history.html [accessed 2022-01-21]
53. Chiu A, Bever L. Are they experimental? Can they alter DNA? Experts tackle lingering coronavirus vaccine fears. The Washington Post. 2021 May 14. URL: https://www.washingtonpost.com/lifestyle/2021/05/14/safe-fast-vaccine-fear-infertility-dna/ [accessed 2022-01-21]
54. Cohen J. Further evidence supports controversial claim that SARS-CoV-2 genes can integrate with human DNA. Science. 2021 May 06. URL: https://www.science.org/content/article/further-evidence-offered-claim-genes-pandemic-coronavirus-can-integrate-human-dna [accessed 2022-01-21]
55. Hamel L, Lopes L, Sparks G, Kirzinger A, Kearney A, Stokes M, et al. KFF COVID-19 Vaccine Monitor: October 2021. Kaiser Family Foundation. 2021 Oct 28. URL: https://www.kff.org/coronavirus-covid-19/poll-finding/kff-covid-19-vaccine-monitor-october-2021/ [accessed 2022-01-21]
56. Kates J, Tolbert J, Orgera K. The red/blue divide in COVID-19 vaccination rates. Kaiser Family Foundation. 2021 Sep 14. URL: https://www.kff.org/policy-watch/the-red-blue-divide-in-covid-19-vaccination-rates/ [accessed 2022-01-21]
57. Beleche T, Ruhter J, Kolbe A, Marus J, Bush L, Sommers B. COVID-19 vaccine hesitancy: demographic factors, geographic patterns, and changes over time. Assistant Secretary for Planning and Evaluation (ASPE). 2021 May. URL: https://aspe.hhs.gov/sites/default/files/private/pdf/265341/aspe-ib-vaccine-hesitancy.pdf [accessed 2022-01-21]
58. Ndugga N, Hill L, Artiga S, Haldar S. Latest data on COVID-19 vaccinations by race/ethnicity. Kaiser Family Foundation. 2022 Jan 12. URL: https://www.kff.org/coronavirus-covid-19/issue-brief/latest-data-on-covid-19-vaccinations-by-race-ethnicity/ [accessed 2022-01-21]
59. Kirzinger A, Sparks G, Kearney A, Stokes M, Hamel L, Brodie M. KFF COVID-19 Vaccine Monitor: November 2021. Kaiser Family Foundation. 2021 Dec 02. URL: https://www.kff.org/coronavirus-covid-19/poll-finding/kff-covid-19-vaccine-monitor-november-2021/ [accessed 2022-01-21]
60. KFF COVID-19 Vaccine Monitor: Does the public want to get a COVID-19 vaccine? When? Kaiser Family Foundation. URL: https://www.kff.org/coronavirus-covid-19/dashboard/kff-covid-19-vaccine-monitor-dashboard/ [accessed 2022-01-21]
61. Bond S. Just 12 people are behind most vaccine hoaxes on social media, research shows. NPR. 2021 May 14. URL: https://www.npr.org/2021/05/13/996570855/disinformation-dozen-test-facebooks-twitters-ability-to-curb-vaccine-hoaxes [accessed 2022-01-21]
62. The disinformation dozen. Center for Countering Digital Hate. 2021 Mar 24. URL: https://252f2edd-1c8b-49f5-9bb2-cb57bb47e4ba.filesusr.com/ugd/f4d9b9_b7cedc0553604720b7137f8663366ee5.pdf [accessed 2022-01-21]

63. Vosloo S. Digital misinformation/disinformation and children. UNICEF. 2021 Aug 24. URL: https://www.unicef.org/globalinsight/stories/digital-misinformation-disinformation-and-children [accessed 2022-01-21]
64. Meeting COVID-19 misinformation/disinformation head-on. Johns Hopkins Bloomberg School of Public Health. URL: https://publichealth.jhu.edu/meeting-covid-19-misinformation-and-disinformation-head-on [accessed 2022-01-21]
65. Romer D, Jamieson KH. Patterns of media use, strength of belief in COVID-19 conspiracy theories, and the prevention of COVID-19 from March to July 2020 in the United States: survey study. J Med Internet Res 2021 Apr 27;23(4):e25215 [FREE Full text] [doi: 10.2196/25215] [Medline: 33857008]
66. Hornik R, Kikut A, Jesch E, Woko C, Siegel L, Kim K. Association of COVID-19 misinformation with face mask wearing and social distancing in a nationally representative US sample. Health Commun 2021 Jan;36(1):6-14. [doi: 10.1080/10410236.2020.1847437] [Medline: 33225745]
67. Resnicow K, Bacon E, Yang P, Hawley S, Van Horn ML, An L. Novel predictors of COVID-19 protective behaviors among US adults: cross-sectional survey. J Med Internet Res 2021 Apr 20;23(4):e23488 [FREE Full text] [doi: 10.2196/23488] [Medline: 33835930]
68. De Coninck D, Frissen T, Matthijs K, d'Haenens L, Lits G, Champagne-Poirier O, et al. Beliefs in conspiracy theories and misinformation about COVID-19: comparative perspectives on the role of anxiety, depression and exposure to and trust in information sources. Front Psychol 2021 Apr 16;12:646394. [doi: 10.3389/fpsyg.2021.646394] [Medline: 33935904]
69. Bridgman A, Merkley E, Loewen P, Owen T, Ruths D, Teichmann L, et al. The causes and consequences of COVID-19 misperceptions: Understanding the role of news and social media. Harvard Kennedy School Misinformation Review. 2020 Jun 18. URL: https://tinyurl.com/3m6fz95m [accessed 2022-01-21]
70. van Prooijen J, Etienne T, Kutiyski Y, Krouwel A. Conspiracy beliefs prospectively predict health behavior and well-being during a pandemic. Psychol Med 2021 Oct 13:1-25. [doi: 10.1017/s0033291721004438]
71. Chan H, Chiu CP, Zuo S, Wang X, Liu L, Hong Y. Not-so-straightforward links between believing in COVID-19-related conspiracy theories and engaging in disease-preventive behaviours. Humanit Soc Sci Commun 2021 May 04;8(1):104. [doi: 10.1057/s41599-021-00781-2]
72. Mosseri A. Working to stop misinformation and false news. Meta. 2021 Apr 17. URL: https://www.facebook.com/formedia/blog/working-to-stop-misinformation-and-false-news [accessed 2022-01-21]
73. YouTube to remove all anti-vaccine misinformation. BBC News. 2021 Sep 29. URL: https://www.bbc.co.uk/news/technology-58743252 [accessed 2022-01-21]
74. Hernandez J. Facebook suspends Marjorie Taylor Greene's account over COVID misinformation. NPR. 2022 Jan 03. URL: https://www.npr.org/2022/01/02/1069753102/twitter-bans-marjorie-taylor-greenes-personal-account-over-covid-misinformation [accessed 2022-01-21]
75. Cadier A. Misinformation 'superspreaders': Covid vaccine falsehoods still thriving on Facebook and Instagram. NewsGuard. 2021 Jan 07. URL: https://www.newsguardtech.com/wp-content/uploads/2021/10/WHO-FB-insta-thriving-report-ENG.pdf [accessed 2022-01-21]
76. Social media. Global Handwashing Partnership. 2021 Oct 15. URL: https://globalhandwashing.org/global-handwashing-day/social-media/ [accessed 2022-01-21]
77. Brown A. Coronavirus: The World Health Organization is becoming the planet's most important social media influencer. Forbes. 2020 Mar 16. URL: https://tinyurl.com/mr3552d8 [accessed 2022-01-21]
78. Marks M. How state officials are trying to use TikTok to stop the spread Of COVID-19. Texas Standard. 2020 Apr 20. URL: https://www.texasstandard.org/stories/how-state-officials-are-trying-to-use-tiktok-to-stop-the-spread-of-covid-19/ [accessed 2022-01-21]
79. Basch CH, Hillyer GC, Meleo-Erwin ZC, Jaime C, Mohlman J, Basch CE. Preventive behaviors conveyed on YouTube to mitigate transmission of COVID-19: cross-sectional study. JMIR Public Health Surveill 2020 Apr 2;6(2):e18807. [doi: 10.2196/18807]
80. Basch CE, Basch CH, Hillyer GC, Jaime C. The role of YouTube and the entertainment industry in saving lives by educating and mobilizing the public to adopt behaviors for community mitigation of COVID-19: successive sampling design study. JMIR Public Health Surveill 2020 Apr 21;6(2):e19145 [FREE Full text] [doi: 10.2196/19145] [Medline: 32297593]
81. Hillyer GC, Basch CH, Basch CE. Coverage of transmission of COVID-19 information on successive samples of YouTube videos. J Community Health 2021 Aug 05;46(4):817-821 [FREE Full text] [doi: 10.1007/s10900-020-00956-z] [Medline: 33400077]
82. Basch C, Hillyer G, Jaime C. COVID-19 on TikTok: harnessing an emerging social media platform to convey important public health messages. Int J Adolesc Med Health. Preprint posted online on August 10, 2020. [FREE Full text] [doi: 10.1515/ijamh-2020-0111] [Medline: 32776899]
83. Basch CH, Fera J, Pierce I, Basch CE. Promoting mask use on TikTok: descriptive, cross-sectional study. JMIR Public Health Surveill 2021 Feb 12;7(2):e26392 [FREE Full text] [doi: 10.2196/26392] [Medline: 33523823]
84. Basch CH, Mohlman J, Fera J, Tang H, Pellicane A, Basch CE. Community mitigation of COVID-19 and portrayal of testing on TikTok: descriptive study. JMIR Public Health Surveill 2021 Jun 10;7(6):e29528. [doi: 10.2196/29528]

85. Basch CH, Fera J, Pellicane A, Basch CE. Handwashing videos on TikTok during the COVID-19 pandemic: potential for disease prevention and health promotion. Infect Dis Health 2022 Feb;27(1):31-37 [FREE Full text] [doi: 10.1016/j.idh.2021.09.039] [Medline: 34690108]
86. Basch CH, Meleo-Erwin Z, Fera J, Jaime C, Basch CE. A global pandemic in the time of viral memes: COVID-19 vaccine misinformation and disinformation on TikTok. Hum Vaccin Immunother 2021 Aug 03;17(8):2373-2377 [FREE Full text] [doi: 10.1080/21645515.2021.1894896] [Medline: 33764283]
87. Jacobsen KH, Vraga EK. Improving communication about COVID-19 and emerging infectious diseases. Eur J Clin Invest 2020 May 09;50(5):e13225 [FREE Full text] [doi: 10.1111/eci.13225] [Medline: 32294248]
88. 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 2022-01-23]
89. Lima D, Lopes M, Brito A. Social media: friend or foe in the COVID-19 pandemic? Clinics (Sao Paulo) 2020;75:e1953 [FREE Full text] [doi: 10.6061/clinics/2020/e1953] [Medline: 32428114]
90. Atehortua NA, Patino S. COVID-19, a tale of two pandemics: novel coronavirus and fake news messaging. Health Promot Int 2021 Apr 15;36(2):524-534 [FREE Full text] [doi: 10.1093/heapro/daaa140] [Medline: 33450022]

Abbreviations

CDC: Centers for Disease Control and Prevention
WHO: World Health Organization

Edited by T Leung, G Eysenbach; submitted 26.01.22; peer-reviewed by D He, A Mavragani, I Mircheva; comments to author 04.03.22; revised version received 25.03.22; accepted 31.03.22; published 12.04.22.

Please cite as:
Basch CH, Basch CE, Hillyer GC, Meleo-Erwin ZC
Social Media, Public Health, and Community Mitigation of COVID-19: Challenges, Risks, and Benefits
J Med Internet Res 2022;24(4):e36804
URL: https://www.jmir.org/2022/4/e36804
doi:10.2196/36804
PMID:35380539

Viewpoint

Defining Telehealth for Research, Implementation, and Equity

Joy Roy[1*], BS, BA; Deborah R Levy[2*], MD, MPH; Yalini Senathirajah[1*], PhD

[1]Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, PA, United States

[2]Department of Medical Informatics and Clinical Epidemiology, Oregon Health Sciences University, Portland, OR, United States

[*]all authors contributed equally

Corresponding Author:
Joy Roy, BS, BA
Department of Biomedical Informatics
School of Medicine
University of Pittsburgh
5607 Baum Boulevard, Suite 500
Pittsburgh, PA, 15206
United States
Phone: 1 412 648 9203
Email: jor115@pitt.edu

Abstract

When the COVID-19 pandemic spurred a disruption in health care delivery, the role of telehealth shifted from an option to a near necessity to maintain access when in-person care was deemed too risky. Each state and many organizations developed temporary telehealth policies for the COVID-19 emergency, each policy with its own definitions, coverage, government cases, and regulations. As pandemic-era policies are now being replaced with more permanent guidelines, we are presented with an opportunity to reevaluate how telehealth is integrated into routine health care delivery. We believe that the timing and nature of the sequential steps for redefining telehealth are critical and that it is important to develop a clear and agreed-on definition of telehealth and its components at this time. We further suggest a necessary preliminary step is to support clear communication and interoperability throughout the development of this definition. Precise and standardized definitions could create an unambiguous environment for clinical care for both patients and providers while enabling researchers to have more precise control over their investigations of telehealth. A consensus when defining telehealth and its derivatives at this critical stage could create a consistent expectation of care for all patients and those who set the standards of care, as it has for other clinical scenarios with clear guidelines.

(J Med Internet Res 2022;24(4):e35037) doi:10.2196/35037

KEYWORDS

telehealth; telemedicine; standards; health equity; public health; digital health; delivery of health care

Introduction

The pandemic accelerated the swift adoption of telehealth [1-4]. As the emergency mode of health care delivery during the pandemic draws to a close for many entities and organizations, questions regarding sustainability, definitions of equitable access, and the future of telehealth arise [3]. Between June 6 and November 6, 2020, nearly one-third of all health visits were conducted remotely as telehealth visits, a considerable expansion compared to prior years [1]. Although 43% of hospitals reported the capacity of supporting telemedicine in 2019, 95% of health centers reported using telehealth during the COVID-19 pandemic, which reflected a rapid increase in use [1]. However, the emergency adoption of telehealth has led to widely differing practices, varying definitions of what constitutes telehealth or telemedicine, and a spectrum of state rules; this variability may be linked to the rapid adoption of telemedicine during the pandemic emergency phase. This new and expanded adoption has at times included exemptions from the usual regulatory processes; these exemptions are now being reevaluated [5,6]. The end of the emergency adoption period opens an opportunity to reevaluate the standards and definitions of telehealth and its components and may reveal the more subtle risk of perpetuating suboptimal practices adopted during the emergency phase of the pandemic.

We will employ the World Health Organization (WHO) definition of equity which explains that equity “is the absence of unfair, avoidable, or remediable differences among groups of people, whether those groups are defined socially, economically, demographically, or geographically or by other dimensions of inequality (e.g., sex, gender, ethnicity, disability, or sexual orientation)” [7]. We propose that a clear and

standardized definition of telehealth, ideally developed through consensus across a wide variety of organizations, must be established to support better implementation and evaluation and to contribute to greater equity in health care access overall.

Definition

To define telemedicine, some sources distinguish between telehealth and telemedicine, while other sources use the terms interchangeably [8-14]. Telemedicine definitions commonly agree that it encompasses the remote diagnosis and treatment of patients by telecommunications infrastructure, while telehealth is commonly considered a superset of telemedicine, defined as any services used to provide care remotely (eg, live conferencing, remote patient monitoring, and personal health apps) [8,9,11]. Already, the terms telemedicine and telehealth both appear to leave room for interpretation and have a broad scope; telemedicine is often the narrower of the 2 terms, though still vaguely defined. With the introduction of derivative terms such as mHealth and eHealth [15], coupled with the different technologies (eg, video visit, audio-only visit, and email- and text-based correspondence) and communication methods (synchronous or asynchronous) that may be used for a telemedicine visit, the expectations for what types of care telemedicine encompasses are further obfuscated.

Standardized definitions in clinical disciplines, as in other disciplines, can help categorize information and encourage more precise communication relating to the area of interest [16-18]. We argue that this situation presents an opportune time to establish a standardized definition of telehealth, with explicit classifications of its modalities, applications, scope, and relationships to other modes of health care. Prior to the expansion of telehealth during the pandemic, telehealth had limited applications in both urban and rural settings, including but not limited to tele–critical care and tele-neurology applications [4,19,20]. Patients communicating with providers and having clear expectations when doing so is known to improve engagement in care; this could be an indirect impact of improving definitions in the field of telehealth [21,22]. As telehealth expands, knowing what the term constitutes will enable more precise communication between providers, patients, policy makers, and researchers.

Clear definitions could support patients in scheduling and requesting visits with greater clarity of the scope of telehealth and expectations of what will occur during their visit as well as potential outcomes to anticipate at the end of the visit. Providers can also benefit by recommending services that stand to offer the best anticipated utility given the combination of patients' conditions, needs, and expectations. For example, video visits are not always possible or even desired by patients [23]. However, phone calls or chat box correspondence may not be enough for clinicians to gauge a patient's health properly for a specific condition. Another factor can be what is covered by a payor or state's coverage for telehealth-related services. Hence, with greater control of the nomenclature used to describe telehealth services, providers can work with patients to recommend the most appropriate type of care, which can lead to greater outreach to patients, quality of care, and return on investments for both patients and physicians. Westby et al [24] provide a case of a patient who had a telehealth visit that can offer an example of the importance of matching expectations with patient understanding of what is offered. Because the patient had concerns regarding his ability to use the technology for a video visit, the clinic staff scheduled the patient for a telephone visit. This resulted in several potentially unanticipated benefits to both the patient and provider, including the patient being comfortable disclosing his challenges with reading, which the providers were unaware of, and the patient being able to spell his medication names directly from the bottles to the provider. The providers were able to clarify the instructions for his medications, which the patient misread due to his difficulty with reading. Audio-only visits are generally not considered telehealth visits in every state or institution, whereas video visits are accepted as telehealth universally [11]. Westby et al [24] note that the Center for Medicare and Medicaid, at the time, had a narrower definition of telehealth than the WHO and excluded telephone visits from reimbursement. Had the clinic's staff not been precise about describing the nature of the visit at the time of scheduling with the patient, and instead simply scheduled a "telehealth visit," the patient could have assumed it to be a video visit and may have chosen to cancel the appointment without the barriers to his participation being brought to light. Patients may not be aware of the options and accessibilities that are offered to them, especially when every state, institution, and payor may have different stipulations of what constitutes a telehealth visit.

Similar arguments regarding the need for a standardized interpretation of telehealth have been made previously, with attempts to develop a common "taxonomy of telemedicine," but this has not gained widespread adoption [18]. Regulation and credentialing are governed by state governments, where each state defines telehealth and telemedicine and its coverage laws independently, and no 2 states are alike in their definitions and regulations [11]. While some states alternate in their use of terms such as telehealth and telemedicine, other states use them interchangeably and even add a variety of terms with a tele- prefix to refer to a remotely delivered version of a term [11].

We found it instructive to compare definitions across US states when examining how significant the framing of telehealth vernacular can be. The term telehealth may be effective in Missouri, defined in a way that encompasses a variety of visit types, but because Maryland might have a different definition for the term, it may not be as effective there. For example, the State Telehealth Laws and Reimbursement Policies Report 2021 shows only 15 states cover audio-only visits, whereas all 50 states reimburse for live video visits [11].

These nonstandard definitions and regulations lead to difficulty in designing and implementing telehealth studies across states. An effectiveness study on telehealth in Missouri, for example, would not be easily compared to a separate study done in Maryland. This discrepancy also creates confusion among patients and providers and makes policies that are needed across states or on a national level challenging to implement. Cross-state billing, for example, is not approved in many cases and poses a barrier for many clinicians who practice outside of a distant patient's state health care network [3].

Evaluations

Researchers across different disciplines interfacing with telehealth in their work could benefit from standardization and precision in the associated definitions [25,26]. We propose that a consensus in definitions could allow investigators to better communicate both in proposals and sharing the products of their work related to telehealth. Because telehealth comprises a large variety of interventions, it is difficult to generalize its effectiveness as different interventions can yield varied results [25]. A precise nomenclature can help researchers identify intended interventions and independent variables, leading to clearer and more precise conclusions [16-18,27,28].

Current studies underway on new telehealth programs vary in scope and outcome measures, ranging from measuring use data, such as adherence measures that count the number of app downloads, to data on the number of appointments scheduled by patients and providers, and the number of completed telemedicine visits [29]. We note that these quantitative analyses might not include qualitative perspectives that could correlate to clinical outcomes. Namely, there has been a lack of evaluations comparing the cost-effectiveness of telehealth to usual care and examining the patient experience [29-31]. Once standardized telehealth definitions are established, there will be an opportunity to better compare and understand the outcomes of in-person care compared to telehealth since the beginning of the COVID-19 pandemic. Standardized definitions have been found to be useful in other clinical research disciplines, such as for cardiovascular outcomes (eg, for metabolic syndrome and primary cardiovascular disease prevention), where studies with differing scopes can be analyzed using agreed-on basic characteristics and shared definitions [32].

The 2017 National Quality Forum report Creating a Framework to Support Measure Development for Telehealth proposed a more quality of care–driven measurement framework made to guide outcome measures about overall experience and care delivery outcomes [33]. This quality of care–motivated framework includes 4 key domains: access to care, financial cost and impact, experience, and effectiveness. Subdomains were also proposed, which include a measurement framework to correlate with clinical and quality of care outcomes. The adoption of this framework would support our aim of standardizing investigations and linking them to care delivery. These domains have a direct impact on quality of care, but this framework has yet to be used as a standard practice [29].

Telehealth is evolving and expanding rapidly with new research, facets, and associated technology. Telehealth today is not the same as it was last year, especially after its expansion in the era of COVID-19. For example, with the rise of smartphone technology, mHealth today is not the same as mHealth 10 years ago [15]. For studies of telehealth to have a fair comparison, normalized definitions and classifications of the different types of care provided via telehealth are required because this can make comparisons across studies easier, both geographically and through time [29].

A large number of organizations and institutions are defining telehealth and doing so on terms that suit their stakeholders, which has created a unique challenge. It is much easier to change a single normative definition than several definitions all at once. Now that organizations such as the US Centers for Disease Control and Prevention, the Office of the National Coordinator for Health Information Technology, the American Academy of Family Physicians, the American Medical Association, state and federal payors, as well as local private payors and institutions are all developing their own definitions, this exacerbates the complexity of developing unified definitions [9,11,13,34,35]. We risk definitions not being updated on the same cycle and further diverging over time.

Standardized definitions will help further to evaluate the system and measure and follow impacts on equity. There are observed trends in the adoption of telehealth from a patient perspective [4,23,36-38]. As Rodriguez shows, there are differences in the use of telehealth services across various ethnic groups [23]. The motivation or explanation from a patient stakeholder perspective warrants further investigation to guide equitable approaches. Failure to account for the patient perspective, especially in health disparity populations and underrepresented populations, can lead to further marginalization and exacerbation of health inequalities [39]. On an enterprise level, institutional resources can vary, which can directly impact the resources available to clinicians and providers and patterns of technology adoption.

There are many different facets of inequity, including structural racism, wealth, sexism, geographic location, and more, which contribute to overall health inequity and social determinants of health [40-42]. We propose that clear definitions of available telehealth services, though they may not solve all forms of inequity, have the potential to reduce misunderstandings, miscommunications, and confusion, all of which contribute to a lack of access to telehealth services. More attention is needed to understand the impact that interventions addressing patient preferences regarding digital health, digital literacy, and technological access will have on equity.

As telehealth rapidly evolves, we propose that attention should be paid to maintaining equity in part through clear definitions and acknowledging the access differentials that may vary across institutions and populations.

Conclusions

Consensus is critical in terms of both equitable access and clinical and other research on the growing field of telemedicine and telehealth. Our scope of interest is in these areas, recognizing the importance of definitions across the institutional level, including interactions with third-party payors, which we have not addressed in this work. Looking to the future, health care organizations can consider consistent and agreed-on standardized definitions to benefit all stakeholders by providing clearer communications, comparable policies, precise and controlled evaluations, and by supporting the delivery of equitable care. This consensus is especially important now as telehealth could be taking on a larger role as part of routine clinical access to care. It is important that we agree on a common

nomenclature soon to make full use of its benefits and prevent future backtracking.

Acknowledgments

JR was supported by the National Library of Medicine (grant 5T15-LM007059-35).

Authors' Contributions

JR initiated the conceptualization of this viewpoint, but all 3 authors contributed substantially. All authors contributed to the methodology. The original draft was written by JR, then heavily modified by all 3 authors. Review and editing were done primarily by JR but all 3 authors contributed substantially.

Conflicts of Interest

None declared.

References

1. Demeke HB, Merali S, Marks S, Pao LZ, Romero L, Sandhu P, et al. Trends in use of telehealth among health centers during the COVID-19 pandemic - United States, June 26-November 6, 2020. MMWR Morb Mortal Wkly Rep 2021 Mar 19;70(7):240-244 [FREE Full text] [doi: 10.15585/mmwr.mm7007a3] [Medline: 33600385]
2. Koonin LM, Hoots B, Tsang CA, Leroy Z, Farris K, Jolly B, et al. Trends in the use of telehealth during the emergence of the COVID-19 pandemic - United States, January-March 2020. MMWR Morb Mortal Wkly Rep 2020 Oct 30;69(43):1595-1599 [FREE Full text] [doi: 10.15585/mmwr.mm6943a3] [Medline: 33119561]
3. Shachar C, Engel J, Elwyn G. Implications for telehealth in a postpandemic future: regulatory and privacy issues. JAMA 2020 May 18:2375-2376. [doi: 10.1001/jama.2020.7943] [Medline: 32421170]
4. Chu C, Cram P, Pang A, Stamenova V, Tadrous M, Bhatia RS. Rural telemedicine use before and during the COVID-19 pandemic: a repeated cross-sectional study. J Med Internet Res 2021 Mar 24:e26960 [FREE Full text] [doi: 10.2196/26960] [Medline: 33769942]
5. Policy changes during COVID-19. U.S. Department of Health & Human Services. URL: https://telehealth.hhs.gov/providers/policy-changes-during-the-covid-19-public-health-emergency/ [accessed 2022-03-30]
6. Poll: voters overwhelmingly support urgent action to permanently protect access to telehealth. Telehealth Access for America. URL: https://telehealthaccessforamerica.org/poll-voters-overwhelmingly-support-urgent-action-to-permanently-protect-access-to-telehealth/ [accessed 2022-03-30]
7. Health equity. World Health Organization. URL: https://www.who.int/health-topics/health-equity#tab=tab_1 [accessed 2022-03-20]
8. What is telehealth? NEJM Catalyst. 2018 Feb 1. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.18.0268 [accessed 2022-03-30]
9. Telehealth and telemedicine. American Academy of Family Physicians. URL: https://www.aafp.org/about/policies/all/telehealth-telemedicine.html [accessed 2022-03-30]
10. How to make the most of a telemedicine visit. Mayo Clinic. URL: https://www.mayoclinic.org/diseases-conditions/coronavirus/in-depth/questions-about-telemedicine/art-20485831 [accessed 2022-03-30]
11. State telehealth laws and reimbursement policies report, fall 2021. Center for Connected Health Policy. 2021 Oct. URL: https://www.cchpca.org/resources/state-telehealth-laws-and-reimbursement-policies-report-spring-2021/ [accessed 2022-03-30]
12. Chaet D, Clearfield R, Sabin JE, Skimming K, Council on Ethical and Judicial Affairs American Medical Association. Ethical practice in telehealth and telemedicine. J Gen Intern Med 2017 Oct;32(10):1136-1140 [FREE Full text] [doi: 10.1007/s11606-017-4082-2] [Medline: 28653233]
13. AMA telehealth quick guide. American Medical Association. URL: https://www.ama-assn.org/practice-management/digital/ama-telehealth-quick-guide [accessed 2022-03-30]
14. Practice management. American Academy of Pediatrics. URL: http://www.aap.org/en-us/professional-resources/practice-transformation/telehealth/Pages/What-is-Telehealth.aspx [accessed 2022-03-30]
15. Ali EE, Chew L, Yap KYL. Evolution and current status of mhealth research: a systematic review. BMJ Innov 2016 Jan 05;2(1):33-40. [doi: 10.1136/bmjinnov-2015-000096]
16. Andrews J, Bogliatto F, Lawson H, Bornstein J. Speaking the same language: using standardized terminology. J Low Genit Tract Dis 2016 Jan;20(1):8-10. [doi: 10.1097/LGT.0000000000000157] [Medline: 26579837]
17. Awaysheh A, Wilcke J, Elvinger F, Rees L, Fan W, Zimmerman K. A review of medical terminology standards and structured reporting. J Vet Diagn Invest 2018 Jan;30(1):17-25 [FREE Full text] [doi: 10.1177/1040638717738276] [Medline: 29034813]

18. Bashshur R, Shannon G, Krupinski E, Grigsby J. The taxonomy of telemedicine. Telemed J E Health 2011 Jul;17(6):484-494. [doi: 10.1089/tmj.2011.0103] [Medline: 21718114]
19. Weigel G, Ramaswamy A, Sobel L, Salganicoff A, Cubanski J, Freed M. Opportunities and barriers for telemedicine in the U.S. during the COVID-19 emergency and beyond. Kaiser Family Foundation. 2020 May 11. URL: https://www.kff.org/womens-health-policy/issue-brief/opportunities-and-barriers-for-telemedicine-in-the-u-s-during-the-covid-19-emergency-and-beyond/ [accessed 2022-03-30]
20. Herasevich V, Subramanian S. Tele-ICU technologies. Crit Care Clin 2019 Jul;35(3):427-438. [doi: 10.1016/j.ccc.2019.02.009] [Medline: 31076043]
21. Flickinger TE, Saha S, Moore RD, Beach MC. Higher quality communication and relationships are associated with improved patient engagement in HIV care. J Acquir Immune Defic Syndr 2013 Jul 01;63(3):362-366 [FREE Full text] [doi: 10.1097/QAI.0b013e318295b86a] [Medline: 23591637]
22. Dickinson JK, Guzman SJ, Maryniuk MD, O'Brian CA, Kadohiro JK, Jackson RA, et al. The use of language in diabetes care and education. Diabetes Care 2017 Dec;40(12):1790-1799. [doi: 10.2337/dci17-0041] [Medline: 29042412]
23. Rodriguez JA, Betancourt JR, Sequist TD, Ganguli I. Differences in the use of telephone and video telemedicine visits during the COVID-19 pandemic. Am J Manag Care 2021 Jan;27(1):21-26 [FREE Full text] [doi: 10.37765/ajmc.2021.88573] [Medline: 33471458]
24. Westby A, Nissly T, Gieseker R, Timmins K, Justesen K. Achieving equity in telehealth: "centering at the margins" in access, provision, and reimbursement. J Am Board Fam Med 2021 Feb;34(Suppl):S29-S32 [FREE Full text] [doi: 10.3122/jabfm.2021.S1.200280] [Medline: 33622814]
25. Portnoy JM, Wu AC. Is telemedicine as effective as usual care? J Allergy Clin Immunol Pract 2019 Jan;7(1):217-218. [doi: 10.1016/j.jaip.2018.09.002] [Medline: 30598177]
26. Lee KK, Thomas RC, Tan TC, Leong TK, Steimle A, Go AS. The Heart Failure Readmission Intervention by Variable Early Follow-up (THRIVE) study: a pragmatic randomized trial. Circ Cardiovasc Qual Outcomes 2020 Oct;13(10):e006553 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.120.006553] [Medline: 32967439]
27. Standardization of medical devices nomenclature: international classification, coding and nomenclature of medical devices. World Health Organization. 2019 Apr 30. URL: https://apps.who.int/gb/ebwha/pdf_files/EB145/B145_3-en.pdf?ua=1 [accessed 2022-03-30]
28. Kay C, Clifford M, Mena P, McDougall G, Andres-Lacueva C, Cassidy A, et al. Recommendations for standardizing nomenclature for dietary (poly)phenol catabolites. Am J Clin Nutr 2020 Oct 01;112(4):1051-1068 [FREE Full text] [doi: 10.1093/ajcn/nqaa204] [Medline: 32936878]
29. Hollander J, Neinstein A. Maturation from adoption-based to quality-based telehealth metrics. NEJM Catalyst. 2020 Sep 9. URL: https://catalyst.nejm.org/doi/pdf/10.1056/CAT.20.0408 [accessed 2022-03-30]
30. Lee M, Wang M, Liu J, Holbrook A. Do telehealth interventions improve oral anticoagulation management? A systematic review and meta-analysis. J Thromb Thrombolysis 2018 Apr;45(3):325-336. [doi: 10.1007/s11239-018-1609-2] [Medline: 29350322]
31. So CF, Chung JW. Telehealth for diabetes self-management in primary healthcare: a systematic review and meta-analysis. J Telemed Telecare 2018 Jun;24(5):356-364. [doi: 10.1177/1357633X17700552] [Medline: 28463033]
32. Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Executive summary of the third report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (adult treatment panel III). JAMA 2001 May 16;285(19):2486-2497. [doi: 10.1001/jama.285.19.2486] [Medline: 11368702]
33. Creating a framework to support measure development for telehealth. National Quality Forum. 2017 Aug 31. URL: https://www.qualityforum.org/publications/2017/08/creating_a_framework_to_support_measure_development_for_telehealth.aspx [accessed 2022-03-30]
34. What is telemedicine in a non-US Setting. US Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/global-covid-19/telemedicine.html [accessed 2022-03-30]
35. What is telehealth? How is telehealth different from telemedicine? Office of the National Coordinator for Health Information Technology. URL: https://www.healthit.gov/faq/what-telehealth-how-telehealth-different-telemedicine [accessed 2022-03-30]
36. Franciosi AN, Quon BS. TeleHealth or TeleWealth? Equity challenges for the future of cystic fibrosis care (Commentary). J Cyst Fibros 2021 Dec;20 Suppl 3:55-56 [FREE Full text] [doi: 10.1016/j.jcf.2021.08.025] [Medline: 34507897]
37. Rodriguez S. SDOH limit mHealth app use for older patients, care management. PatientEngagementHIT. 2022 Feb 16. URL: https://patientengagementhit.com/news/sdoh-limit-mhealth-app-use-for-older-patients-care-management [accessed 2022-03-30]
38. The Lancet Haematology. Telehealth and digital equity for older people. Lancet Haematol 2021 Nov;8(11):e777. [doi: 10.1016/S2352-3026(21)00314-8] [Medline: 34715040]
39. Valdez RS, Rogers CC, Claypool H, Trieshmann L, Frye O, Wellbeloved-Stone C, et al. Ensuring full participation of people with disabilities in an era of telehealth. J Am Med Inform Assoc 2021 Feb 15;28(2):389-392 [FREE Full text] [doi: 10.1093/jamia/ocaa297] [Medline: 33325524]

40. 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] [Medline: 32639914]
41. Khairat S, Haithcoat T, Liu S, Zaman T, Edson B, Gianforcaro R, et al. Advancing health equity and access using telemedicine: a geospatial assessment. J Am Med Inform Assoc 2019 Aug 01;26(8-9):796-805 [FREE Full text] [doi: 10.1093/jamia/ocz108] [Medline: 31340022]
42. Braveman P, Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. Public Health Rep 2014;129 Suppl 2:19-31 [FREE Full text] [doi: 10.1177/00333549141291S206] [Medline: 24385661]

Abbreviations

WHO: World Health Organization

Edited by R Kukafka; submitted 17.11.21; peer-reviewed by K Fultz Hollis, K Wrona; comments to author 09.02.22; revised version received 01.03.22; accepted 13.03.22; published 13.04.22.

Please cite as:
Roy J, Levy DR, Senathirajah Y
Defining Telehealth for Research, Implementation, and Equity
J Med Internet Res 2022;24(4):e35037
URL: https://www.jmir.org/2022/4/e35037
doi:10.2196/35037
PMID:35416778

Viewpoint

Framework and Practical Guidance for the Ethical Use of Electronic Methods for Communication With Participants in Medical Research

Atsushi Kogetsu[1], MD, PhD; Kazuto Kato[1], PhD

Department of Biomedical Ethics and Public Policy, Graduate School of Medicine, Osaka University, Suita, Osaka, Japan

Corresponding Author:
Atsushi Kogetsu, MD, PhD
Department of Biomedical Ethics and Public Policy
Graduate School of Medicine
Osaka University
2-2 Yamadaoka
Suita, Osaka, 565-0871
Japan
Phone: 81 6 6879 3688
Fax: 81 6 6879 3681
Email: kogetsu@eth.med.osaka-u.ac.jp

Abstract

Online communication with participants, including online recruitment, electronic informed consent, and data communication, is one of the fields to which information and communication technology (ICT) has been applied in medical research. Online communication provides various benefits, especially for genome research and rare disease research. However, ethical challenges that are derived from or exacerbated by online communication need to be addressed. Here, we present an overview of such ethical issues and provide practical guidance for the ethical implementation of ICT. We specify the ethical issues in the context of using online communication for medical research by an analysis based on the eight ethical principles for clinical research. Informed by this ethical context, we then develop a novel framework for the governance of medical research involving ICT, which consists of eight categories: five research processes (ie, design of research, recruitment, informed consent, data communication, and dissemination and return of results) and three overarching perspectives related to multiple processes of research (ie, access to research and online dialog, community involvement, and independent review). Finally, we present a practical guidance chart for researchers, patient partners, independent reviewers, and funding agencies. We believe that our study will contribute to the ethical implementation of online communication in medical research.

(J Med Internet Res 2022;24(4):e33167) doi:10.2196/33167

KEYWORDS

online communication; electronic methods; online recruitment; electronic informed consent; e-IC; digital consent; online consent; data communication; digital health

Introduction

One of the applications of information and communication technology (ICT) in medical research is a variety of online communications with participants, including recruitment, informed consent, and sharing of research results [1-4]. The use of the internet is expected to enable continuous communication without spatial restrictions [3]. This is especially beneficial in rare disease research because it is difficult to recruit a sufficient number of participants in geographically limited areas. As clarified by the Rare and Undiagnosed Diseases Study (RUDY) in the United Kingdom and Japan, digital platforms enable the effective recruitment of participants [5-7]. Additionally, the RUDY project demonstrated that digital technologies facilitate dialog and collaboration between patients and researchers [5-7]. This type of active patient involvement encouraged by digital tools is referred to as participant-centric initiatives [8,9]. Another research area where ICT use is expected to bring about significant change is human genome research. Since genome cohort studies such as the “All of US” study in the United States aim to recruit a large number of participants, registering and answering an online questionnaire should prove to be efficient and helpful for participants [10]. Moreover, long-term communication with participants is more important for genome

research projects because the results of individual analyses can change as data are accumulated over time [11].

At the same time, online communication can raise new ethical challenges such as the digital divide, concerns about invasion of privacy, how to assure understanding of information for consent, and methods for authenticating participants [12-15]. To maximize the benefits of ICT, it is crucial to deal with such ethical challenges. Previous studies focused on individual ethical issues; however, there has been no attempt to provide an overview of the ethical issues that need to be addressed in implementing online communication with participants. Although some countries already have regulations or recommendations regarding electronic methods, such as guidance on the use of electronic consent released by the US Food and Drug Administration [16], their scope is limited. In Japan, the newly revised ethical guidelines for medical research established in 2021 include descriptions about online informed consent, but few details are specified. Therefore, the purpose of this viewpoint is to provide an overview of ethical issues and present a practical guide for implementing online communication in medical research.

We initially performed an analysis based on the eight ethical principles for clinical research proposed by Emanuel et al [17], referred to as the Emanuel Framework (EF). We specified the ethical issues in the context of online communication with participants through each principle of the EF: collaborative partnership, social value, scientific validity, fair participant selection, favorable risk-benefit ratio, independent review, informed consent, and respect for participants. The lead author (AK) carried out a literature search and both authors performed analyses. We then developed a new framework composed of the issues specified by the analysis, after receiving feedback from the researchers of the medical and genomics research projects under the same research program that funded this study, sponsored by the Japan Agency for Medical Research and Development (AMED). Finally, we propose a practical guide for researchers, patient partners, independent reviewers, and funding agencies.

Ethical Analysis Based on the EF

Collaborative Partnership

The first key principle in the EF emphasizes that during the course of medical research, creating a good partnership among medical researchers, participants, and other members of the project is crucial. Several key issues emerge when examining situations to conduct medical research using ICT.

First, there is a possibility that research projects have difficulties in finding representatives of participating communities, although ICT enables existing community members to communicate more easily. For example, for large-scale research projects that target multiple countries, it is a challenge to decide who represents the target population. It may be necessary to redefine our understanding of “community” as well as community “representatives.” This newly understood community would include online patient networks that have been attracting attention in recent years [18-21].

Second, a community’s distinct values, circumstances, culture, and social practices should be respected even when employing electronic methods. If most of the community members are not familiar with the internet or are reluctant to send their information online, electronic methods should not be implemented.

Third, digital tools make collaboration more diverse, enabling more casual patient involvement such as feedback online [22]. One challenge is how to share responsibility with such casual involvement. A fair distribution of the tangible and intangible rewards of research among the partners based on each contribution is also a challenge.

Social Value

When conducting medical research, assessing and enhancing its social value are essential as well as protecting participants. Using ICT can enhance the social value of medical research. Performing large-scale studies with electronic methods leads to an increase in the number of beneficiaries. The value of the research can also be enhanced with the secondary usage of the data online, which may be especially true for studies using genomic data. Research results can be disseminated more widely with the internet [23]. Moreover, the use of multimedia can improve accessibility and understanding.

However, caution is needed to prevent adverse impacts on existing health care infrastructure and its sustainability when communicating directly with participants across countries through online systems. This is especially true when returning individual genetic analysis results because novel therapeutic strategies based on these results may create an additional burden on existing health care systems.

Scientific Validity

Scientific validity, the third principle of the EF, is one of the fundamental factors for ethical research. When utilizing ICT for medical research, it should be taken into account that the use of ICT can positively influence the scientific validity and reliability of the research design. Regarding study designs, a sufficient sample size is important to ensure scientific validity. For rare disease research, recruiting a sufficient number of participants online makes the studies more scientifically valid.

However, the ease of this approach varies depending on the frequency of ICT use and differing values toward privacy. For instance, online recruitment may be ineffective for people who do not have access to the internet, are unfamiliar with it, or do not want to enter their personal information online. This may cause selection bias [2,19,24-27]. At the time of registration, especially when there is a reward for participation, preventing duplicate registration of participants is also important to ensure scientific validity [28].

Fair Participant Selection

Research participants must be selected primarily based on scientific objectivity. The online platform allows individualized recruitment, which makes it possible to achieve a fairer selection of participants by ensuring accountability for the selection [29].

For research projects that target vulnerable people, careful consideration is necessary. It is conceivable that the use of electronic methods to facilitate participation in research will enable vulnerable individuals to participate in more research owing to their medical or economic circumstances, which could cause overexploitation or insufficient understanding of the projects. Measures should be taken to protect the potential participants with such vulnerabilities from the risks associated with research participation.

Favorable Risk-Benefit Ratio

When carrying out research, potential risks and benefits for individual participants should be assessed and explained. Based on this principle, the risks associated with using electronic methods should be identified, described, and minimized. These risks include data leakage during online data transfer and impersonation by others in the digital authentication system for participants [15]. Another risk is miscommunication caused by the lack of nonverbal information when using certain kinds of online tools [30].

In the process of returning individual genetic analysis results, sufficient medical, psychological, and social care for the patients are absolutely required. When returning results through online systems, it should be noted that there may be an increased risk of failing to ensure that such considerations are made.

Independent Review

Independent review mechanisms are usually determined by laws and regulations, and vary depending on countries or institutions. In traditional research projects, researchers generally recruit participants from their own countries and therefore only need to follow the rules of their countries. However, online recruitment may encourage participation across countries and institutions. Therefore, there is often a lack of laws and regulations concerning the detailed procedures for independent review of research protocols when the research is planned in one country and participants are recruited from other countries, as in so-called direct-to-participant (DTP) recruitment. A detailed analysis of these issues is beyond the scope of this paper; however, some possible solutions have already been proposed [31].

In an independent review, reviewers should be competent; however, it is currently very difficult to adequately assess research using online communications from scientific, ethical, and technical perspectives due to the lack of research or guidance on practices for implementing communications using electronic methods. In particular, questions concerning how a secure research system can be built and what kind of privacy risks associated with online recruitment may occur cannot be fully assessed without a high level of expertise [32]. Therefore, the availability of reports and guidance, as well as expert advice, is crucial.

Informed Consent

Informed consent requires demonstration of respect for the autonomy of individuals, and online consent is no exception. Online informed consent enables prospective participants to make their participation decisions at their own pace, which is a major advantage in reducing psychological impact [33]. It is also important to ensure individual authentication for voluntary research participation [15].

In terms of recruitment procedures, it is necessary to assess whether online recruitment is culturally, politically, and socially acceptable. For example, due consideration should be given to the recruitment of prospective participants when using online behavior histories or social networks, if these are considered to be potential invasions of privacy [27].

It is important that the information presented online is complete, accurate, and not overwhelming. Adjusting the amount of information presented at one time and using multimedia are expected to improve the understanding of the prospective participants [33,34]. However, it is necessary to provide information tailored to individual literacy rather than providing a uniform explanation. Furthermore, to obtain informed consent with the full understanding of the participants, the opportunity to ask questions must also be guaranteed in online informed consent. Unless face-to-face informed consent is provided, such as through a video conferencing system, a more careful assessment is required as to whether candidates fully understand the information presented [13,33]. In addition, it should be noted whether the means of symbolizing consent, such as electronic signatures and clicks for online informed consent, are sufficiently accepted in the communities to which the candidates belong. Finally, as indicated in the US Guidance, the method of obtaining informed consent from a participant's legally authorized representative should also be determined in advance [16].

Respect for Participants

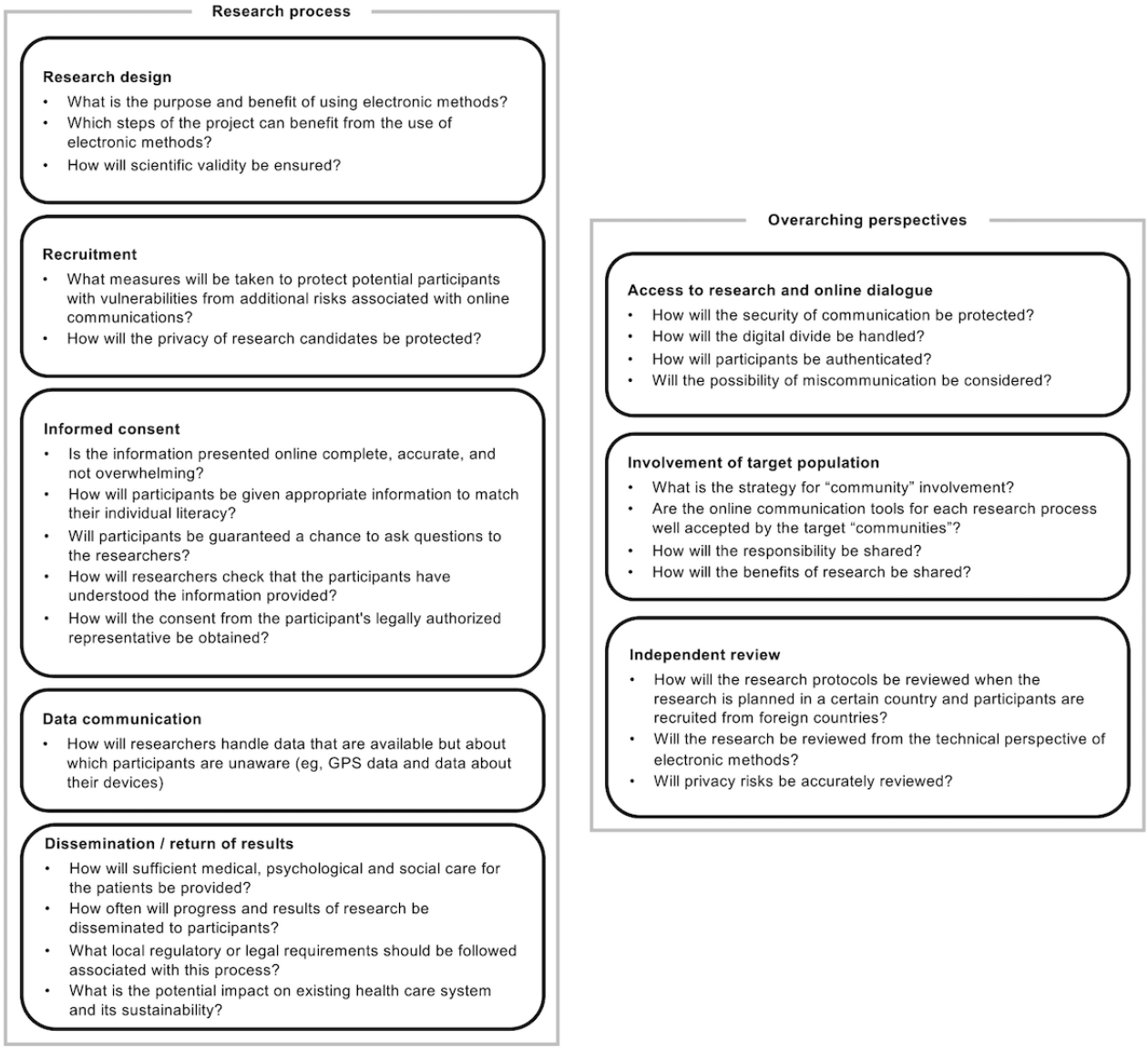
Communication with participants online can help with monitoring health and well-being so as to minimize harm, particularly through subjective assessments. Online methods may help to facilitate communication in terms of providing participants with information about research progress and results, related health care, and treatment. However, when, how, and what information is appropriate to provide may vary from person to person and, as such, these issues should be considered to suit individual participants where possible.

Regarding confidentiality, various security challenges raised by online communications must be addressed [14]. It should be noted that additional data that can be collected, such as location data, may infringe on the privacy of participants [35]. To ensure security, it is necessary to establish a reliable method of authentication [15].

Overview of the Ethical Issues in Online Communication With Participants

Informed by the EF ethical analysis, we have schematically presented the interrelationships among the issues based on the actual processes of medical research and formulated a novel framework for the governance of medical research involving ICT (Figure 1).

Figure 1. Framework for the governance of genomics research involving information and communication technology. The five boxes on the left show the research process and the three boxes on the right show overarching perspectives related to the five processes. Each box contains the main benchmarks for that issue or process.



The framework is designed for any research involving communication with participants, not for studies that involve digital technology in general (eg, research drawing on electronic medical records without any online communication with participants). The framework covers eight elements. The first five elements (left side of Figure 1) are the processes involved in medical research using ICT, which cover the design of research, recruitment, informed consent, data communication, and dissemination of progress and results, including the return of individual genomic analysis results. The other three elements (right side of Figure 1) are overarching perspectives. They cover access to research, community involvement, and independent review. In each category, benchmarks are described to show what issues are involved.

The "design of research" category includes the purpose and benefit of the use of electronic methods. In addition, it should be taken into account that online communication would affect scientific validity. This is especially true when recruitment and data communication through online systems are employed. The category of "recruitment" includes the issues related to risks when recruiting participants online. Researchers must consider that online recruitment would have additional risks for participants with vulnerabilities and that some of their data may be available before the participants have given consent. The category of "informed consent" includes the benchmarks associated with information presented online, the understanding of participants, and consent from the participant's legally authorized representative. The category of "data communication" covers the situation in which participants manage their own data or when data are generated by the participants' own devices. The category of "dissemination and return of results" includes the benchmarks related to not only individual participants but also the target community.

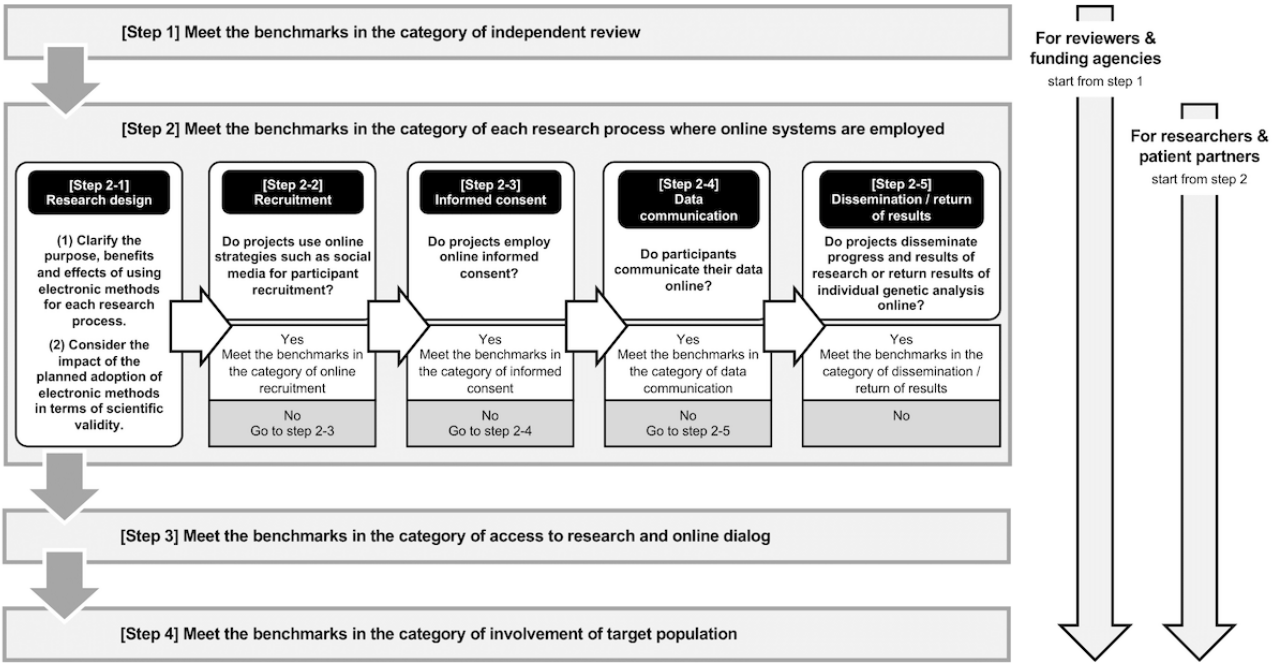
Security issues, digital divide, authentication of participants, and the risk of miscommunication are important factors in considering access to the research system and online dialog between researchers and participants. The "involvement of target population" category includes the benchmarks associated with

the strategy and tools for "community" involvement, as well as sharing responsibility and benefits. The category of "independent review" includes the benchmarks regarding the procedures for independent review of DTP recruitment and competency of the reviewers.

Implications for Practice

Based on this framework, we present a practical guide in the form of a chart for researchers, patient partners, independent reviewers, and funding agencies (Figure 2). This guidance chart mainly consists of four steps: (1) meet the benchmarks in the category of independent review, (2) meet the benchmarks in the category of each research process where online systems are employed, (3) meet the benchmarks in the category of access to research, and (4) meet the benchmarks in the category of patient involvement.

Figure 2. Guidance chart for ethical research using online communication with participants. This guidance chart mainly consists of four steps. Step 1 is for independent reviewers and funding agencies, and steps 2, 3, and 4 are for reviewers, funding agencies, researchers, and patient partners.



Depending on the users of the chart, the first step is either step 1 or step 2. For independent reviewers and funding agencies, the first step is to ensure that the reviewers are competent to review this type of research using online communications with participants, especially from the perspectives of technology and privacy risks. In cases where they are not competent, they need to ask ICT experts for help and review the literature to address new ethical issues such as the means of symbolizing consent in the process of informed consent.

Steps 2 to 4 are for independent reviewers, funding agencies, researchers, and patient partners. In the second step, researchers should clarify the purpose, benefits, and effects of using electronic methods for each research process, as well as the impact of the planned adoption of electronic methods in terms of scientific validity. Subsequently, researchers and patient partners should address the ethical issues related to each research process where they plan to employ online systems for communication with participants, from recruitment to dissemination of research results. Concrete issues that need to be addressed can be found in the benchmarks in the framework categories shown in Figure 1. Independent reviewers should review each step by carefully examining the content of research protocols that address these ethical aspects.

The third and fourth steps focus on the overarching perspectives that need to be ensured in all of the processes of research projects. Even if a research project plans to employ online communication only for one process such as online informed consent, the ethical issues included in the category of access to research and online dialog, and involvement of the target population should be addressed.

For funding agencies, in addition to making sure that the relevant benchmarks are met, it is desirable to ensure that adequate funding is available for the costs of addressing ethical issues in each step, particularly in steps 1, 3, and 4. Regarding step 1, funding may be required to ask ICT experts for help and access to useful literature. In step 3, large investments are sometimes required to address security issues and to implement appropriate authentication systems.

Conclusion

As ICT becomes implemented and applied more widely, it is hoped that it will be used actively in medical research. In some countries, electronic methods are increasingly being employed in research projects. Online systems should be utilized carefully while keeping in mind the ethical considerations that have been described in this article. We believe that our ethical framework and the guidance chart can help to make electronic strategies ethically sound and acceptable.

Acknowledgments

We thank Dr Sumio Sugano from Tokyo Medical and Dental University, and members of the Japan Agency for Medical Research and Development (AMED) project on “Leading the way to Genomic Medicine by ELSI Research Program” for their valuable discussions and comments on this work. We also thank the researchers of the research program “The Advanced Genome Research and Bioinformatics Study to facilitate Medical Innovation (GRIFIN)” funded by AMED for providing valuable feedback on the framework. This research was supported by AMED under grant 18km0405301h0003.

Authors' Contributions

KK conceived of the idea for the study, and AK and KK designed the study. AK carried out the literature search. Both authors performed the analyses. The draft written by AK was improved through discussions and editing by both authors, who read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. 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]
2. Eysenbach G, Wyatt J. Using the internet for surveys and health research. J Med Internet Res 2002;4(2):E13 [FREE Full text] [doi: 10.2196/jmir.4.2.e13] [Medline: 12554560]
3. De Sutter E, Zaçe D, Boccia S, Di Pietro ML, Geerts D, Borry P, et al. Implementation of electronic informed consent in biomedical research and stakeholders' perspectives: systematic review. J Med Internet Res 2020 Oct 08;22(10):e19129 [FREE Full text] [doi: 10.2196/19129] [Medline: 33030440]
4. Budin-Ljøsne I, Teare HJA, Kaye J, Beck S, Bentzen HB, Caenazzo L, et al. Dynamic consent: a potential solution to some of the challenges of modern biomedical research. BMC Med Ethics 2017 Jan 25;18(1):4 [FREE Full text] [doi: 10.1186/s12910-016-0162-9] [Medline: 28122615]
5. Javaid MK, Forestier-Zhang L, Watts L, Turner A, Ponte C, Teare H, et al. The RUDY study platform - a novel approach to patient driven research in rare musculoskeletal diseases. Orphanet J Rare Dis 2016 Nov 08;11(1):150 [FREE Full text] [doi: 10.1186/s13023-016-0528-6] [Medline: 27825362]
6. Teare HJA, Hogg J, Kaye J, Luqmani R, Rush E, Turner A, et al. The RUDY study: using digital technologies to enable a research partnership. Eur J Hum Genet 2017 Jun;25(7):816-822 [FREE Full text] [doi: 10.1038/ejhg.2017.57] [Medline: 28443622]
7. Hamakawa N, Kogetsu A, Isono M, Yamasaki C, Manabe S, Takeda T, et al. The practice of active patient involvement in rare disease research using ICT: experiences and lessons from the RUDY JAPAN project. Res Involv Engagem 2021 Feb 01;7(1):9 [FREE Full text] [doi: 10.1186/s40900-021-00253-6] [Medline: 33526087]
8. Kaye J, Curren L, Anderson N, Edwards K, Fullerton SM, Kanellopoulou N, et al. From patients to partners: participant-centric initiatives in biomedical research. Nat Rev Genet 2012 Apr 03;13(5):371-376 [FREE Full text] [doi: 10.1038/nrg3218] [Medline: 22473380]
9. Hamakawa N, Nakano R, Kogetsu A, Coathup V, Kaye J, Yamamoto BA, et al. Landscape of participant-centric initiatives for medical research in the United States, the United Kingdom, and Japan: scoping review. J Med Internet Res 2020 Aug 04;22(8):e16441 [FREE Full text] [doi: 10.2196/16441] [Medline: 32749228]
10. All of Us Research Program Investigators. The “All of Us” research program. N Engl J Med 2019 Aug 15;381(7):668-676 [FREE Full text] [doi: 10.1056/nejmsr1809937]
11. Aronson SJ, Rehm HL. Building the foundation for genomics in precision medicine. Nature 2015 Oct 15;526(7573):336-342 [FREE Full text] [doi: 10.1038/nature15816] [Medline: 26469044]
12. Casañas i Comabella C, Wanat M. Using social media in supportive and palliative care research. BMJ Support Palliat Care 2015 Jun;5(2):138-145. [doi: 10.1136/bmjspcare-2014-000708] [Medline: 25311211]
13. Vayena E, Haeusermann T, Adjekum A, Blasimme A. Digital health: meeting the ethical and policy challenges. Swiss Med Wkly 2018;148:w14571 [FREE Full text] [doi: 10.4414/smw.2018.14571] [Medline: 29376547]
14. Henderson EM, Law EF, Palermo TM, Eccleston C. Case study: ethical guidance for pediatric e-health research using examples from pain research with adolescents. J Pediatr Psychol 2012;37(10):1116-1126 [FREE Full text] [doi: 10.1093/jpepsy/jss085] [Medline: 22851643]
15. Kogetsu A, Ogishima S, Kato K. Authentication of patients and participants in health information exchange and consent for medical research: a key step for privacy protection, respect for autonomy, and trustworthiness. Front Genet 2018;9:167 [FREE Full text] [doi: 10.3389/fgene.2018.00167] [Medline: 29910822]

16. Use of electronic informed consent: Questions and answers: Guidance for institutional review boards, investigators, and sponsors. US Food and Drug Administration. 2016 Dec. URL: https://www.fda.gov/media/116850/download [accessed 2021-08-26]
17. Emanuel EJ, Wendler D, Grady JO. An ethical framework for biomedical research. In: Emanuel EJ, Grady C, Crouch RA, Lie RK, Miller FG, Wendler D, editors. The Oxford textbook of clinical research ethics. Oxford, UK: Oxford University Press; 2008:123-135.
18. Demiris G. The diffusion of virtual communities in health care: concepts and challenges. Patient Educ Couns 2006 Aug;62(2):178-188. [doi: 10.1016/j.pec.2005.10.003] [Medline: 16406472]
19. Ravoire S, Lang M, Perrin E, Participants of Giens XXXII Round Table No. 6. Advantages and limitations of online communities of patients for research on health products. Therapie 2017 Feb;72(1):135-143. [doi: 10.1016/j.therap.2016.11.058] [Medline: 28139283]
20. Lamas E, Salinas R, Vuillaume D. A new challenge to research ethics: patients-led research (PLR) and the role of internet based social networks. Stud Health Technol Inform 2016;221:36-40. [Medline: 27071872]
21. Frost JH, Massagli MP. Social uses of personal health information within PatientsLikeMe, an online patient community: what can happen when patients have access to one another's data. J Med Internet Res 2008 May 27;10(3):e15 [FREE Full text] [doi: 10.2196/jmir.1053] [Medline: 18504244]
22. Kim KK, Khodyakov D, Marie K, Taras H, Meeker D, Campos H, et al. A novel stakeholder engagement approach for patient-centered outcomes research. Med Care 2018 Oct;56(10 Suppl 1):S41-S47 [FREE Full text] [doi: 10.1097/MLR.0000000000000790] [Medline: 30074950]
23. Schnitzler K, Davies N, Ross F, Harris R. Using Twitter™ to drive research impact: A discussion of strategies, opportunities and challenges. Int J Nurs Stud 2016 Jul;59:15-26. [doi: 10.1016/j.ijnurstu.2016.02.004] [Medline: 27222446]
24. Volkova E, Michie J, Corrigan C, Sundborn G, Eyles H, Jiang Y, et al. Effectiveness of recruitment to a smartphone-delivered nutrition intervention in New Zealand: analysis of a randomised controlled trial. BMJ Open 2017 Jul 02;7(6):e016198 [FREE Full text] [doi: 10.1136/bmjopen-2017-016198] [Medline: 28674144]
25. Karlson EW, Boutin NT, Hoffnagle AG, Allen NL. Building the Partners HealthCare Biobank at Partners Personalized Medicine: informed consent, return of research results, recruitment lessons and operational considerations. J Pers Med 2016 Jan 14;6(1):2 [FREE Full text] [doi: 10.3390/jpm6010002] [Medline: 26784234]
26. Puhan MA, Steinemann N, Kamm CP, Müller S, Kuhle J, Kurmann R, Swiss Multiple Sclerosis Registry SMSR. A digitally facilitated citizen-science driven approach accelerates participant recruitment and increases study population diversity. Swiss Med Wkly 2018;148:w14623 [FREE Full text] [doi: 10.4414/smw.2018.14623] [Medline: 29767828]
27. Schwab-Reese LM, Hovdestad W, Tonmyr L, Fluke J. The potential use of social media and other internet-related data and communications for child maltreatment surveillance and epidemiological research: Scoping review and recommendations. Child Abuse Negl 2018 Nov;85:187-201 [FREE Full text] [doi: 10.1016/j.chiabu.2018.01.014] [Medline: 29366596]
28. Teitcher JEF, Bockting WO, Bauermeister JA, Hoefer CJ, Miner MH, Klitzman RL. Detecting, preventing, and responding to "fraudsters" in internet research: ethics and tradeoffs. J Law Med Ethics 2015;43(1):116-133 [FREE Full text] [doi: 10.1111/jlme.12200] [Medline: 25846043]
29. Parra-Calderón CL, Kaye J, Moreno-Conde A, Teare H, Nuñez-Benjumea F. Desiderata for digital consent in genomic research. J Community Genet 2018 Apr;9(2):191-194 [FREE Full text] [doi: 10.1007/s12687-017-0355-z] [Medline: 29363051]
30. Plasek JM, Pieczkiewicz DS, Mahnke AN, McCarty CA, Starren JB, Westra BL. The role of nonverbal and verbal communication in a multimedia informed consent process. Appl Clin Inform 2017 Dec 16;02(02):240-249 [FREE Full text] [doi: 10.4338/aci-2011-02-ra-0016] [Medline: 23616873]
31. Rothstein MA, Zawati MH, Beskow LM, Brelsford KM, Brothers KB, Hammack-Aviran CM, et al. Legal and ethical challenges of international direct-to-participant genomic research: conclusions and recommendations. J Law Med Ethics 2019 Dec;47(4):705-731. [doi: 10.1177/1073110519898297] [Medline: 31957580]
32. Bender JL, Cyr AB, Arbuckle L, Ferris LE. Ethics and privacy implications of using the internet and social media to recruit participants for health research: a privacy-by-design framework for online recruitment. J Med Internet Res 2017 Apr 06;19(4):e104 [FREE Full text] [doi: 10.2196/jmir.7029] [Medline: 28385682]
33. Chalil Madathil K, Koikkara R, Obeid J, Greenstein JS, Sanderson IC, Fryar K, et al. An investigation of the efficacy of electronic consenting interfaces of research permissions management system in a hospital setting. Int J Med Inform 2013 Sep;82(9):854-863 [FREE Full text] [doi: 10.1016/j.ijmedinf.2013.04.008] [Medline: 23757370]
34. Nishimura A, Carey J, Erwin PJ, Tilburt JC, Murad MH, McCormick JB. Improving understanding in the research informed consent process: a systematic review of 54 interventions tested in randomized control trials. BMC Med Ethics 2013 Jul 23;14(1):28 [FREE Full text] [doi: 10.1186/1472-6939-14-28] [Medline: 23879694]
35. Perakslis ED. Using digital health to enable ethical health research in conflict and other humanitarian settings. Confl Health 2018;12(1):23 [FREE Full text] [doi: 10.1186/s13031-018-0163-z] [Medline: 29785204]

Abbreviations

AMED: Japan Agency for Medical Research and Development
DTP: direct-to-participant
EF: Emanuel Framework
GRIFIN: The Advanced Genome Research and Bioinformatics Study to facilitate Medical Innovation
ICT: information and communication technology
RUDY: Rare and Undiagnosed Diseases Study

Edited by A Mavragani; submitted 26.08.21; peer-reviewed by T Boughtwood, A Garcia; comments to author 11.12.21; revised version received 02.01.22; accepted 06.02.22; published 20.04.22.

Please cite as:
Kogetsu A, Kato K
Framework and Practical Guidance for the Ethical Use of Electronic Methods for Communication With Participants in Medical Research
J Med Internet Res 2022;24(4):e33167
URL: https://www.jmir.org/2022/4/e33167
doi:10.2196/33167
PMID:35442208

Viewpoint

A Framework for Femtech: Guiding Principles for Developing Digital Reproductive Health Tools in the United States

Tamar Krishnamurti[1,2], PhD; Mehret Birru Talabi[2,3], MD, PhD; Lisa S Callegari[2,4,5], MD, MPH; Traci M Kazmerski[2,6], MD, MSc; Sonya Borrero[1,2,7], MD, MSc

[1]Division of General Internal Medicine, University of Pittsburgh, Pittsburgh, PA, United States
[2]Center for Innovative Research on Gender Health Equity, University of Pittsburgh, Pittsburgh, PA, United States
[3]Division of Rheumatology and Clinical Immunology, University of Pittsburgh, Pittsburgh, PA, United States
[4]Department of Obstetrics and Gynecology, University of Washington School of Medicine, Seattle, WA, United States
[5]Health Services Research and Development, Veterans Affairs Puget Sound Health Care System, Seattle, WA, United States
[6]Department of Pediatrics, University of Pittsburgh, Pittsburgh, PA, United States
[7]Center for Health Research and Promotion, Veterans Affairs Pittsburgh Healthcare System, Pittsburgh, PA, United States

Corresponding Author:
Tamar Krishnamurti, PhD
Division of General Internal Medicine
University of Pittsburgh
200 Meyran Avenue
Suite 200
Pittsburgh, PA, 15213
United States
Phone: 1 4126924855
Email: tamark@pitt.edu

Abstract

The United States has abysmal reproductive health indices that, in part, reflect stark inequities experienced by people of color and those with preexisting medical conditions. The growth of "femtech," or technology-based solutions to women's health issues, in the public and private sectors is promising, yet these solutions are often geared toward health-literate, socioeconomically privileged, and/or relatively healthy white cis-women. In this viewpoint, we propose a set of guiding principles for building technologies that proactively identify and address these critical gaps in health care for people from socially and economically marginalized populations that are capable of pregnancy, as well as people with serious chronic medical conditions. These guiding principles require that such technologies: (1) include community stakeholders in the design, development, and deployment of the technology; (2) are grounded in person-centered frameworks; and (3) address health disparities as a strategy to advance health equity and improve health outcomes.

(J Med Internet Res 2022;24(4):e36338) doi:10.2196/36338

KEYWORDS

United States; North America; femtech; mHealth; health equity; pregnancy; women's health; preterm birth; contraception; family planning; reproductive care; sterilization; cystic fibrosis; rheumatic disease; eHealth; mobile health; reproductive health; digital health; health technology; health outcomes

Background

Black birthing people living in the United States are four times more likely to die during the peripartum period than those who are white, regardless of income, education, or severity of complications [1,2]. Black Americans, Indigenous Americans, and Latinas residing in the United States all experience significantly higher rates of preterm delivery than those who are white [3]. Drivers of reproductive health disparities, including personally mediated and institutional racism, manifest in myriad ways, from housing inequities and unequal access to safe environments to education, opportunity, and health care [4,5]. In addition, individuals with serious chronic medical conditions—particularly those from socially marginalized groups—may be less likely than healthy people to receive critical health interventions, including access to high-quality contraceptives and other reproductive health care [6] and prenatal services [7]. Research by our team members identified

critical gaps in current reproductive health care delivery models, especially for people from socially and economically marginalized populations, those with serious chronic medical conditions, and those with intersectional identities that are subject to both social and medical marginalization. These gaps in care undermine proactive approaches to identify reproductive health needs and thereby address preventable adverse outcomes.

A host of new software and digital tools have emerged from the private and public sectors to address women's reproductive health needs. Colloquially referred to as "femtech," these technologies range from menstrual cycle–tracking apps to medical devices for pelvic floor strengthening. In the private sector, femtech is expected to constitute a US $50 billion industry by 2025 [8]. While femtech represents a promising opportunity to address existing gaps in women's health care, the intended recipients of femtech innovations largely appear to be healthy, affluent, white, cis women. In the current model, an opportunity is missed to engage populations who have been historically underserved and bear the largest burden of poor pregnancy and perinatal outcomes.

This viewpoint outlines three key principles for developing scientifically grounded digital tools that bridge critical gaps in women's health care among individuals who have been subjected to social, economic, and medical marginalization. We illustrate the implementation of these principles with examples taken from our own development of tools that span reproductive health transitions or decisions, including family planning, pregnancy, and sterilization. These principles are equally applicable to the development of research and commercial tools, and offer a set of guidelines to broaden femtech to more equitably address women's health needs.

Here, "women's health" refers to the area of research dedicated to the treatment and diagnosis of diseases and conditions that affect those with female physiology and, in this instance, female reproductive health. As such, we use the terms "woman" and "women" throughout this paper. However, we acknowledge that these are gendered terms and the intended users of femtech do not necessarily identify as women. An important future consideration for our framework and the femtech field more broadly will be enhancing gender inclusivity in both the language and methods we use.

Our Principles

Collaborative Development Strategy

Our collaborative of university-based researchers first joined to garner intellectual support and collaboration around our shared interests across diverse medical and scientific subspecialties. We developed the following shared principles to formalize guidance for our own reproductive digital health tool designs. Lessons learned were obtained through trial and error.

Create Interdisciplinary Stakeholder–Inclusive Teams

Cross-disciplinary collaboration is required to ensure that any reproductive health tool is comprehensive and accurate. Clinical experts offer scientifically grounded content and appropriate clinical actions. Social scientists explain human behavior and the social structures that constrain or shape behaviors. Visual designers capture the clinical and social scientists' input through a user interface. Technical experts realize the teams' vision into a functional and scalable tool.

However, academic and technical partnerships are not sufficient. Echoing tenets of user-centered design, those who are building tools must have a solid understanding of who the users will be and how they will interact with the tool [9]. We take this a step further by suggesting that representatives of all stakeholder groups, particularly patient stakeholders, must be included in the entire process from design to implementation. Patients, for example, should serve not only as testers but as active participants in shaping the conceptual and pragmatic underpinnings of the project, and are thus a critical part of an interdisciplinary team. The specific stakeholders' needs and the methods in which they are engaged in the development process will inevitably differ from tool to tool and by the specific health context being addressed.

Facilitate a Person-Centered Approach

Over the past 20 years, person-centeredness has increasingly been recognized as an indicator of high-quality health care [10]. While patients are now acknowledged as active partners in the clinical decision-making process [11], a power asymmetry intrinsically exists between clinicians and patients [12]. Clinicians may control patients' access to reproductive services by selecting which procedures they conduct or which referrals they order; they are thereby positioned to support or undermine an individual's ability to actualize their reproductive goals and preferences [13]. Furthermore, women experiencing social marginalization may have different decision-making constraints, values, or choice architecture, leading to different, but equally valid, decisions [14].

It is an ethical imperative that individuals who are the desired users of these tools have the power to be active participants in their health care decisions and that their right to make their own reproductive decisions is honored, regardless of the context. Thus, tools must aim to facilitate meaningful understanding of the risks, benefits, and uncertainties associated with various reproductive health decisions, helping users to clarify their personal preferences as they make their choices. Some of our tools aim to help patients make decisions that are aligned with their preferences and values. Some tools offer personalized feedback on patients' specific health needs. Other tools are designed to support shared decision-making between users and their clinicians. All of our tools serve the purpose of supporting a patient's autonomy and self-efficacy to achieve their desired reproductive health goals.

The success of a digital reproductive health tool must, in part, be judged on its outcomes. We believe that these outcomes, much like the approach to tool design and functionality, must be person-centered. In a clinical setting, this includes ensuring that changes in clinical as well as patient-reported outcomes are aligned with the values of the patients themselves. While the success of femtech is usually evaluated on widespread adoption or a revenue stream if commercialized, our mission is to build and implement tools that improve health care experiences and support health outcomes that are desirable to the patient. Such

outcomes may range from interpersonal (eg, feeling respected) to psychosocial (eg, decreased decisional conflict) to clinical (eg, less disparity in preterm births). The key is that these outcomes must be grounded in what patients value as individual *people* and not what affects the fiscal or other priorities of the health care establishment serving them. Ideally, tools that can improve reproductive health in a person-centered manner will also be widely adopted and consequently improve the efficiency and quality of health services.

Advance Reproductive Health Equity

To advance reproductive health equity and address the needs of populations whose reproductive health and well-being have been harmed or neglected by our current societal structures, femtech work must have a race-, class-, and gender-conscious approach from its inception. We suggest specific frameworks that guide our development process, including Intersectionality Theory [15], Critical Race Theory [16], the R4P Framework [17], and the Reproductive Justice Framework [18]. These specific frameworks are grounded in racial justice and US civil rights; nevertheless, considerations of intersectionality and the larger message of these frameworks are globally relevant [19]. Moreover, we urge constant self-reflection on each step of the process [20].

Practical Application of These Principles

Here, we offer additional details about our collaborative's suite of tools, each of which were built using our guiding principles to address a unique decision-making period in the reproductive journey: reproductive care for nonpregnant people, pregnancy, and sterilization. Table 1 outlines how to incorporate these guiding principles into the different phases of tool development. As we describe each tool, we offer practical illustrations.

Table 1. Ways to incorporate principles into phases of femtech development.

Phase of development	Principles		
	Interdisciplinary stakeholder–inclusive teams	Person-centered approach	Advancing reproductive health equity
Conceptualization and content development	Conduct semistructured qualitative work (one-on-one interviews or focus groups) with women, health care providers (including subspecialty providers), and other expert stakeholders	Structure interview guides around evidence-based best practices to identify gaps in knowledge, and understand experiences and preferences related to reproductive care	Incorporate historical and theoretical frameworks in conceptualizing the tool and its content to ensure an equity lens from the start of any work
Design implementation	Review content and functionality iteratively with members of key stakeholder groups such as patients, medical experts, human-computer interaction specialists, bioethicists, social scientists, and relevant community organization leaders (eg, reproductive justice advocacy groups, church leaders, women's shelters, doulas)	Design features and content to incorporate clinical best practices, yet focus on users' informational needs and personal values	Structure advisory or expert panels to include content and lived-experience experts; seek diverse perspectives within each category of stakeholder
Testing	Prioritize patient and other stakeholder goals for the tool	Plan acceptability metrics around patient-centered/patient-defined outcomes	Power trials to identify differences in outcomes for diverse patient populations based on preplanned equity-driven hypotheses

Reproductive Care for Nonpregnant People With MyPath

National guidelines recommend that patient-centered reproductive services be routinely provided in preventive care settings to help individuals optimize health and well-being prior to desired pregnancies and to prevent unwanted pregnancy and births [21]. Despite these recommendations, contraception and abortion counseling is frequently absent from preventive health encounters [22,23] and, when it does occur, often fails to prioritize individual preferences, values, and goals [24,25]. Individuals in socially marginalized groups, including people of color and those with low incomes, perceive a lower quality of care [26] and report experiences of discrimination and pressure to use contraception, to choose certain methods, or to limit family size [24,27]. Our understanding of what constitutes high-quality reproductive care has shifted in recent years, driven by advancements in research and social movements [28,29]. This evolution has included moving from approaches focused on strict pregnancy planning and reducing individual risk behaviors toward approaches that include assessment of individuals' goals and needs, acknowledge the structural factors (eg, racism and poverty) underlying poor reproductive health outcomes, and prioritize reproductive autonomy [28]. The benefits of person-centered counseling strategies such as shared decision-making, which acknowledge the complexity of contraceptive preferences and help individuals match those preferences with their choices, are increasingly being recognized [27]. At the same time, the potential harms of directive counseling approaches, which advocate for one contraceptive method or group of methods over others, have also been brought to light, particularly in communities that face historic and ongoing reproductive oppression in the United States [30,31].

Figure 1 shows MyPath, a patient-facing web-based reproductive decision support tool that we developed to promote high-quality, person-centered discussions about reproductive needs in the Veterans Health Administration (VA). Veterans capable of pregnancy who use VA Health Care services are a highly diverse

population, with nearly half identifying as racial or ethnic minoritized groups [32], and face elevated risks of adverse pregnancy and birth outcomes compared to their civilian counterparts due to a higher prevalence of medical, mental health, and psychosocial factors [33]. Drawing on our first principle (*Principle 1: Interdisciplinary stakeholder–inclusive teams*), MyPath development was informed by formative qualitative work to understand veterans' preferences and needs [25] and followed user-centered design principles, guided by patient, provider, and scientific expert input [34]. Designed to be used prior to primary care visits, MyPath empowers patients by helping them clarify and share their reproductive goals with their primary care providers, augmenting their knowledge, and building self-efficacy in communicating about their reproductive needs (*Principle 2: Facilitating a patient-centered approach*). By centering individuals' preferences, needs, and goals, MyPath aims to safeguard reproductive autonomy and address inequities in reproductive health care related to poor provider-patient communication (*Principle 3: Advancing reproductive health equity*).

In a nonrandomized pilot of the final prototype (control group, n=28; intervention group, n=30), we found that a greater proportion of intervention participants reported having discussions about their reproductive needs in the visit compared to controls (93% vs 68%; *P*=.02) [34]. A multisite hybrid effectiveness-implementation randomized controlled trial is currently ongoing to evaluate the impact of the text message–delivered MyPath on various person-centered and clinical outcomes, with the goal of ultimately scaling the intervention nationally within VA if found to be effective.

Figure 1. MyPath online decision support tool.

Pregnancy Support With MHP MyHealthyPregnancy

Preterm births, those that occur prior to 37 weeks of gestation, are the leading direct cause of neonatal mortality and morbidity [35]. More than 1 in 9 births in the United States are preterm [36], and Black women are roughly 60% more likely than white women to have a baby born prematurely. The causes of preterm birth are complex, with a wide range of risk factors identified in the literature, some of which may never be identified by the provider during routine care, such as intimate partner violence (IPV) [37], depression [38], or chronic toxic stress [39]. Even when risk factors are identifiable and addressable, physicians may find it difficult to communicate recommendations effectively enough to support risk mitigation [40]. Physician-patient communication challenges are particularly prevalent for Black patients, who, compared to non-Hispanic white patients, report that they are unlikely to ask questions of their physicians and that their physicians are less likely to listen when they do [41]. Even for those physicians that are aware of patients' particular informational needs, time constraints may prevent more in-depth discussion. Addressing these adverse events requires both accurate identification of an individual's risk factors and actionable communication of those risk factors to women and their health care providers.

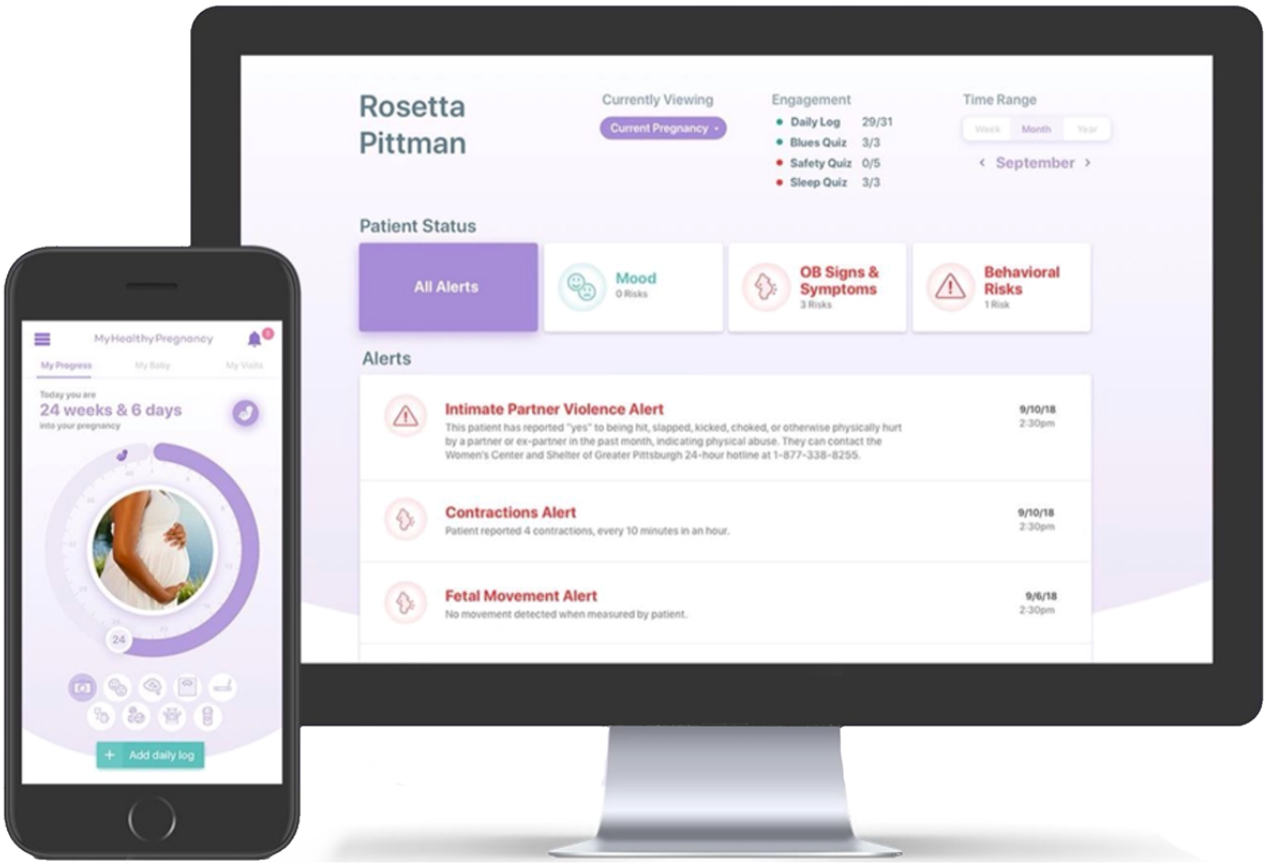
The MHP MyHealthyPregnancy mobile health platform shown in Figure 2 comprises a patient-facing smartphone app and an electronic health record–integrated provider portal. MHP MyHealthyPregnancy uses real-time data collection from pregnant people via the app [42] and applies statistical machine learning algorithms [43] to those data to identify modifiable precursor risks to preterm birth between routine prenatal care visits. As risk factors are identified, they are communicated to health care providers in real time through a provider portal. The app, in turn, offers patients sensitive, respectful, and actionable risk minimization strategies and resources.

MHP MyHealthyPregnancy content and flow were initially drafted by an interdisciplinary team (*Principle 1: Interdisciplinary stakeholder–inclusive teams*), including medical experts in the field of maternal-fetal medicine as well

as human-computer interaction specialists and community informants (eg, church leaders, nonprofit organizations, women's shelters, doula groups). All experts had a voice in determining how to assess and intervene on the specific risk factors that fell within their own area of expertise. Next, we worked directly with pregnant and recently pregnant women, purposively recruiting a socioeconomically and racially diverse group, to identify their current beliefs, values, and constraints with respect to using such a pregnancy tool. The features we incorporated into the app were designed iteratively with the women we interviewed to specifically address expert-identified risks while remaining sensitive to the users' needs (*Principle 2: Facilitating a patient-centered approach*). For example, our medical experts highlighted a need to provide education about symptomatic bleeding; yet, many of the peripartum women we spoke with reported confusion between spotting, miscarrying, and menstruation. As such, the content we produced had to be acceptable and understandable to both types of stakeholders. Our app-based solution was a daily symptom assessment with feedback on the need for immediate medical care when appropriate [42].

MHP MyHealthyPregnancy was tested in a quality-improvement initiative in UPMC (formerly known as the University of Pittsburg Medical Center) health care system from September 23, 2019, to September 1, 2021, with over 5600 English-speaking pregnant people; evaluation analyses are ongoing. Of those offered MHP MyHealthyPregnancy by their provider, 81.5% of patients initiated use of the tool. Initial findings show promising results for MHP MyHealthyPregnancy for filling gaps in clinical risk detection and, ultimately, prevention. In an analysis of 959 patients who used the app for reporting IPV, 100% of those reporting a current physical risk of IPV had no mention of IPV in their medical charts, despite treatment for injuries that should have prompted in-person screening administration [44]. MHP MyHealthyPregnancy was similarly able to successfully implement screening for preeclampsia risk factors among 2563 patients, with more than half of those app users who met the highest preeclampsia risk criteria reporting no preeclampsia prophylactic recommendation from their provider [40]. Patients with baseline reports of certain clinical risk factors such as a history of depression or prior preeclampsia have engaged with the app in significantly higher numbers than those without a baseline history of pregnancy risk factors. To address challenges identified with provider bandwidth in responding to notifications in the provider portal, we built in an alert to be directly sent to a dedicated clinical care team consisting of nursing staff. This approach allowed for an increased response to app-based detection of depression, facilitating connection to behavioral health services.

Figure 2. The MHP MyHealthyPregnancy smartphone app (left) and provider portal (right).



For any digital tool that addresses reproductive health needs, a one-size-fits-all approach will likely fail. For example, in creating a Spanish-language version of MyHealthyPregnancy (*MHP Embarazo Saludable)*, our first step was not a straight translation of the existing content into Spanish, but rather to reframe and then translate the content to address the unique needs of Latinas not born in the United States, who may be navigating a system and approach to pregnancy health care that differ substantially from their prior experience. For example, many Latinas we spoke with who were recent immigrants

discussed their concerns about social isolation, particularly given the supportive role that female relatives traditionally play in transitioning to motherhood [45]. Therefore, for the Latina population, explicitly facilitating a connection to relevant social supports (eg, Spanish-language doulas) based on when a need is identified through MHP MyHealthyPregnancy is a person-centered approach that may affect equitable health outcomes, as identified by the patients themselves (*Principle 3: Advancing reproductive health equity*). A multisite pilot study, funded by the Centers for Disease Control and Prevention, is currently examining the effectiveness of person-centered approaches for psychosocial screening and support with both Spanish and English versions of the tool at two demographically diverse sites.

Sterilization Decision-Making With MyDecision

Female sterilization, which is currently the most commonly used contraceptive method in the United States, is disproportionately used by those with social disadvantages, including those with low incomes, with public or no insurance, with lower educational levels, and from racial/ethnic minority groups [46,47]. The reasons for this are complex and exist within a broader social context that includes a history of sterilization abuses as well as ongoing devaluation of socially marginalized women's reproduction. Research findings illuminate various tensions and potentially countervailing forces that exist in the decision to use sterilization as a contraceptive method and the ability to execute such decisions. For example, women report that sterilization decisions are largely driven by personal preferences and prior reproductive experiences [48,49]. However, among those who have undergone the procedure, there is a significant level of misunderstanding about sterilization (eg, the permanence of the procedure) and alternative contraception methods, as well as a relatively high prevalence of poststerilization regret [50], suggesting suboptimal counseling and decision-making. Additionally, although rates of sterilization are higher among those with low incomes and people of color, many from these groups also report restricted access to the procedure due to provider reluctance to perform it for various reasons (eg, concern that the patient will ultimately regret the decision) as well as barriers posed by stringent Medicaid sterilization consent policies, and these issues are perceived as undermining their reproductive autonomy [51].

The content of the MyDecision tool, as shown in Figure 3, was informed by foundational, in-depth interviews with racially diverse, low-income women who had ever considered tubal sterilization, as well as with health care providers who perform female sterilization to understand informational and decision support preferences and needs. Several of the women participants were also included on a Steering Committee that helped guide the design and structure of the tool throughout the development process. As the history of sterilization and its consent among low-income women have been socially and ethically fraught, Steering Committee members also included representation from women of color and reproductive justice advocacy groups, the American College of Obstetricians and Gynecologists (ACOG) Ethics Committee, and the ACOG Committee on Health Care for Underserved Women. To ensure that the voices of our "patient" participants were heard, we held separate meetings with them in addition to their participation in the larger Steering Committee meetings (*Principle 1: Interdisciplinary stakeholder–inclusive teams*). Once a prototype of the tool was approved by the Steering Committee, cognitive testing with end users (ie, women with low incomes and considering tubal sterilization) was performed to assess comprehensibility and usability of the tool and make subsequent refinements, consistent with human-centered design principles.

Figure 3. The MyDecision web-based tool.

Dismantling the potential for sterilization abuses will require grappling with the complicated social and political forces that stratify reproductive value. However, the MyDecision patient-facing, web-based decision aid seeks to address a more feasible, yet necessary, objective: to empower low-income English- and Spanish-speaking women who are contemplating undergoing a sterilization procedure by providing them with unbiased, relevant information, and a process with which to *independently* engage in informed and value-concordant decision-making and communicate their preferences to their health care providers. The overarching goal of the tool is to better support low-income women's bodily and reproductive

autonomy and help them achieve their reproductive goals (*Principle 2: Facilitating a patient-centered approach* and *Principle 3: Advancing reproductive health equity*).

The MyDecision tool is currently being tested in a National Institutes of Health–funded multisite randomized controlled trial to determine its efficacy in improving decision quality. If found to improve sterilization decision quality and help ensure informed and voluntary consent, MyDecision could potentially offer a scalable and evidence-based alternative to the current problematic Medicaid sterilization consent process.

Tools to Address Considerations Related to Chronic Medical Conditions With MyVoice:Rheum and MyVoice:CF

People who have chronic medical conditions may have reproductive health concerns that are both general and disease-specific. Our team developed two patient-directed tools to support family planning care for women with rheumatic and musculoskeletal diseases (RMDs) and cystic fibrosis (CF); these conditions are independently associated with pregnancy-related concerns, including potentially higher rates of maternal morbidity and mortality and/or adverse neonatal outcomes [52]. Approximately 7 million women in the United States have RMDs, including autoimmune and connective tissue diseases, and many of these diseases are diagnosed during reproductive age [53]. Owing to the advent of new and highly effective medications targeting the underlying defect in CF, people with CF are anticipating longer, healthier lives, and many are considering and pursuing pregnancy [54]. Among women with RMDs and CF, family planning care is therefore essential for supporting informed decision-making about: (1) the benefits and risks of pregnancy, (2) medication safety in the context of pregnancy and lactation, (3) pregnancy intentions and preferred timing for pregnancy and/or parenthood, and (4) selection of safe and acceptable contraceptive methods [55,56].

However, family planning discussions rarely occur in either the RMD or CF subspecialty care contexts [57,58]. Many subspecialty clinicians are inadequately trained to provide consistent, accurate, or high-quality family planning care. Obstetricians and primary care providers often lack adequate disease-related knowledge to provide family planning care that comprehensively addresses patients' information needs [59]. Furthermore, clinicians may have social biases that lead them to counsel patients about their reproductive options differently with consideration of their social and economic backgrounds in addition to their health [60].

We developed two decisions aids (see Figure 4)—MyVoice:Rheum and MyVoice:CF—to help women (1) conceptualize the benefits and potential health risks of parenting, pregnancy, and/or pregnancy avoidance in the context of their conditions; (2) recognize and refine their reproductive goals; and (3) communicate these goals to their health care teams. Tool development centered on engagement with patients, clinicians, and reproductive health specialists to ensure that the content was accurate (*Principle 1: Interdisciplinary stakeholder–inclusive teams*), reflected patients' specific preferences and information priorities (*Principle 2: Supporting a patient-centered approach*), and helped to prepare patients for shared reproductive decision-making with clinicians in ways that enable them to receive the reproductive health care that meets their needs and preferences (*Principle 3: Advancing reproductive health equity*) [61,62]. Although the content of MyVoice:Rheum and MyVoice:CF is different and disease-specific, we developed the tools in tandem to maximize efficiencies related to time and expense.

We are currently undertaking feasibility testing for MyVoice:Rheum and MyVoice:CF.

Figure 4. MyVoice:Rheum and MyVoice:CF tools for reproductive health decision-making for women with specific chronic illnesses.

Avenues and Challenges to Ethical Dissemination

Our principles are, in theory, equally applicable to the design of femtech in both the academic and commercial spheres, and our own work straddles the spectrum from early prototypes to commercialization. However, the academic and commercial spheres operate on very different timelines and incentives. Academic research is, by nature, slow-moving. Its overarching goal is the creation of knowledge. The pace with which such knowledge is created is dependent on rigorous scientific design and is often at the mercy of federal or nonprofit funding cycles. Therefore, an idea is often formed many years before the work can even begin to be realized. The benefit of such a timeline is the luxury of a careful and scientifically grounded approach.

However, the risk is not moving at a pace that is necessary for real-world impact. Commercial femtech, while often borne from academic research, can, and indeed must, move on a faster timeline, which can result in faster realization of a product that meets a pressing reproductive health need. The risk is prioritizing profit over patient-centeredness, especially in instances where the needs of patients and the desires of investors are not aligned, or if certain patients are not considered to constitute a profitable consumer base.

Within our own collaborative, we have tried different models. Most of us are currently evaluating our work with either federal or nonprofit funding (MyPath, MyDecision, MyVoice:Rheum, and MyVoice:CF). In contrast, MHP MyHealthyPregnancy has licensed intellectual property from the universities where it was built, leading to the development of a start-up company. The commercialization approach allows for continuity in delivering the product to the patient, preventing the need to cut off tool functionality due to a lapse in grant funding. For example, the MHP MyHealthyPregnancy platform was able to rapidly integrate triaging and education in response to the COVID-19 pandemic, which may not have been possible if relying solely on federal funding. However, forming a commercial entity may lead to the perception that the entity was created to monetize health disparities, regardless of how mission-driven product development is. As commercial products grow and scale, that skepticism can evolve into a true tension between social good and return on investment. Moreover, once a tool has moved out of the university setting, the researchers/developers no longer have access to an internal institutional review board to provide ethical oversight. This oversight may be particularly important for the protection of patients who are already socially or medically vulnerable. Several commercial femtech companies without such oversight have, without user consent, sold user data to third parties, including to the companies at which users work. Therefore, researchers wishing to offer sustainable and scalable services through commercialization must build the ethical imperative for responsible femtech into the driving mission of any commercial entity from the outset to ensure that the principles of equitable and high-quality care are upheld if and when transitioning from the academic to the market sphere.

Another important consideration for femtech dissemination is equitable outreach that extends beyond the research context. In each of our studies, we were able to proactively recruit diverse groups of participants, oftentimes partnering with community-based organizations or advocacy groups to do so. In some instances, we were able to provide phones or tablets for use by those with limited digital access. However, the question remains for how to ensure that these femtech tools continue to serve the needs and interests of those experiencing health, societal, and digital inequities once they are launched on a larger scale. While the principles proposed herein will ideally lead to high levels of engagement among populations that are typically underresourced, the technology still needs to get into the hands of those who may need it most. To achieve this, we advocate for a community-based partnership model of dissemination, including seeking consultation and partnership on continued product design and distribution with professional and nonprofit expert organizations. Women currently comprise the majority of Medicaid participants and Medicaid coverage is the highest among demographic groups experiencing the greatest health inequities [63]. Therefore, another promising dissemination pathway is state-based coverage, whereby individual states mandate the utilization of specific femtech innovations by Managed Care Organizations. Lastly, it is vital to make sure that the technology itself poses minimal access issues. In the United States, use of smartphones and similar digital devices is ubiquitous, yet the quality and consistency of digital service is not. For people living in rural areas or who have financial constraints around data plans, software development should ensure that the key features still work offline and on a variety of devices (eg, building in native code).

It is essential to take stock of what can and cannot be addressed by health technology. We believe that a limitation of emerging technology innovations is the conviction that technology presents an all-encompassing solution to societal ills. Experiences of societal, institutional, and interpersonal discrimination or marginalization require much more than a single technology-based solution. However, individuals facing health disparities cannot also be sacrificed to a digital divide, which may further potentiate health disparities. If digital technologies are developed by centering the unique needs of those who face the greatest health disparities, then femtech can begin to rectify some of the disproportionately experienced risk factors for poor outcomes. Without following these principles, even the most well-intentioned femtech risks further exacerbating inequities at the intersection of gender, race, and health.

Looking Ahead

We recognize there will be missteps as we struggle to document truths and address gaps in care for underserved populations. We continue to explore the best ways to scale and disseminate our tools to reach those women with the greatest reproductive health disparities. We hope that others may contribute their own guiding principles and implementation strategies to the discussion and continue to refine our proposed approaches, so that the femtech community can aspire to enhance the health care quality and equity of all women.

Acknowledgments

TK's work was supported, in part, by grants from the Centers for Disease Control and Prevention (R44DP006417), the National Institute of Mental Health (R21 MH119450), and the Kuno Award for Applied Science for Social Good from the S&R Foundation. MBT's work was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (K23 AR075057-01A1) and the Robert Wood Johnson Harold Amos Medical Faculty Development Award. LSC was supported by a VA Health Services Research and Development Service Career Development Award (CDA 14-412) and Merit Review Award (IIR 19-387). TMK's

work was supported by a Harry S Shwachman Award from the Cystic Fibrosis Foundation (KAZMER18A0-Q). SB's work was supported, in part, by the National Institute on Minority Health and Health Disparities (R01 MD011678-01). We thank Miriam Bernard, Iris Olson, and Alison Decker for editorial assistance in preparing this manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, the Centers for Disease Control and Prevention, or VA.

Authors' Contributions

TK, MBT, TMK, LSC, and SB all conceptualized the framework, contributed original writing, and provided critical feedback on the manuscript. TK drafted the first version of the manuscript.

Conflicts of Interest

TK is a cofounder and equity-holder of Naima Health LLC, whose flagship product, MyHealthyPregnancy, was licensed from Carnegie Mellon University and the University of Pittsburgh, where she holds academic appointments. TK completed this manuscript as part of her academic research and did not receive compensation from Naima Health LLC for this work. Her role as both an academic and company cofounder offered her unique intellectual insights for this manuscript. However, there is no anticipated financial benefit to TK or Naima Health from this manuscript. The other authors have no conflicts of interest to declare.

References

1. Howell EA. Reducing disparities in severe maternal morbidity and mortality. Clin Obstet Gynecol 2018 Jun;61(2):387-399 [FREE Full text] [doi: 10.1097/GRF.0000000000000349] [Medline: 29346121]
2. Tucker MJ, Berg CJ, Callaghan WM, Hsia J. The Black–White disparity in pregnancy-related mortality from 5 conditions: differences in prevalence and case-fatality rates. Am J Public Health 2007 Feb;97(2):247-251. [doi: 10.2105/ajph.2005.072975]
3. Matthews TJ, MacDorman MF. Infant mortality statistics from the 2010 period linked birth/infant death data set. Natl Vital Stat Rep 2013 Dec 18;62(8):1-26 [FREE Full text] [Medline: 24735562]
4. Lorch SA, Enlow E. The role of social determinants in explaining racial/ethnic disparities in perinatal outcomes. Pediatr Res 2016 Jan;79(1-2):141-147. [doi: 10.1038/pr.2015.199] [Medline: 26466077]
5. Alhusen JL, Bower KM, Epstein E, Sharps P. Racial discrimination and adverse birth outcomes: an integrative review. J Midwifery Womens Health 2016 Nov;61(6):707-720 [FREE Full text] [doi: 10.1111/jmwh.12490] [Medline: 27737504]
6. Borrero S, Schwarz EB, Creinin M, Ibrahim S. The impact of race and ethnicity on receipt of family planning services in the United States. J Womens Health (Larchmt) 2009;18(1):91-96 [FREE Full text] [doi: 10.1089/jwh.2008.0976] [Medline: 19072728]
7. Bengiamin MI, Capitman JA, Ruwe MB. Disparities in initiation and adherence to prenatal care: impact of insurance, race-ethnicity and nativity. Matern Child Health J 2010 Jul;14(4):618-624. [doi: 10.1007/s10995-009-0485-y] [Medline: 19557508]
8. Femtech - Time for a Digital Revolution in the Women's Health Market. Frost & Sullivan. 2018 Jan 31. URL: https://ww2.frost.com/frost-perspectives/femtechtime-digital-revolution-womens-health-market/ [accessed 2020-09-17]
9. Gould JD, Lewis C. Designing for usability: key principles and what designers think. Commun ACM 1985 Mar 01;28(3):300-311. [doi: 10.1145/3166.3170]
10. Institute of Medicine (US) Committee on Quality of Health Care in America, Institute of Medicine. Crossing the quality chasm: a new health system for the 21st century. In: Institute of Medicine Report. Washington DC: National Academy Press; 2001.
11. NEJM Catalyst Staff. What is patient-centered care? NEJM Catalyst 2017;3(1):1-1 [FREE Full text] [doi: 10.1056/CAT.17.0559]
12. Nimmon L, Stenfors-Hayes T. The "Handling" of power in the physician-patient encounter: perceptions from experienced physicians. BMC Med Educ 2016 Apr 18;16:114 [FREE Full text] [doi: 10.1186/s12909-016-0634-0] [Medline: 27091146]
13. Bending ZJ. Reconceptualising the doctor-patient relationship: recognising the role of trust in contemporary health care. J Bioeth Inq 2015 Jun;12(2):189-202. [doi: 10.1007/s11673-014-9570-z] [Medline: 25124984]
14. Geronimus AT, Bound J, Waidmann TA. Health inequality and population variation in fertility-timing. Soc Sci Med 1999 Dec;49(12):1623-1636. [doi: 10.1016/s0277-9536(99)00246-4]
15. Carbado DW, Crenshaw KW, Mays VM, Tomlinson B. Intersectionality. Du Bois Rev 2014 Jan 03;10(2):303-312. [doi: 10.1017/s1742058x13000349]
16. Cabrera NL. Where is the racial theory in Critical Race Theory?: a constructive criticism of the Crits. Rev High Educ 2018;42(1):209-233. [doi: 10.1353/rhe.2018.0038]
17. Hogan V, Rowley DL, White SB, Faustin Y. Dimensionality and R4P: A Health Equity Framework for Research Planning and Evaluation in African American Populations. Matern Child Health J 2018 Feb;22(2):147-153. [doi: 10.1007/s10995-017-2411-z] [Medline: 29392541]

18. Ross LJ. Reproductive justice as intersectional feminist activism. Souls 2018 Jan 16;19(3):286-314. [doi: 10.1080/10999949.2017.1389634]
19. Kapilashrami A, Hankivsky O. Intersectionality and why it matters to global health. Lancet 2018 Jun 30;391(10140):2589-2591. [doi: 10.1016/S0140-6736(18)31431-4] [Medline: 30070211]
20. Thomas SB, Quinn SC, Butler J, Fryer CS, Garza MA. Toward a fourth generation of disparities research to achieve health equity. Annu Rev Public Health 2011;32:399-416 [FREE Full text] [doi: 10.1146/annurev-publhealth-031210-101136] [Medline: 21219164]
21. Gavin L, Pazol K. Update: Providing Quality Family Planning Services - Recommendations from CDC and the U.S. Office of Population Affairs, 2015. MMWR Morb Mortal Wkly Rep 2016 Mar 11;65(9):231-234. [doi: 10.15585/mmwr.mm6509a3] [Medline: 26963363]
22. Bello JK, Rao G, Stulberg DB. Trends in contraceptive and preconception care in United States ambulatory practices. Fam Med 2015 Apr;47(4):264-271 [FREE Full text] [Medline: 25853596]
23. Oza-Frank R, Gilson E, Keim SA, Lynch CD, Klebanoff MA. Trends and factors associated with self-reported receipt of preconception care: PRAMS, 2004-2010. Birth 2014 Dec;41(4):367-373. [doi: 10.1111/birt.12122] [Medline: 24995805]
24. Becker D, Tsui A. Reproductive health service preferences and perceptions of quality among low-income women: racial, ethnic and language group differences. Perspect Sex Reprod Health 2008 Dec;40(4):202-211. [doi: 10.1363/4020208] [Medline: 19067933]
25. Callegari LS, Tartaglione EV, Magnusson SL, Nelson KM, Arteburn DE, Szarka J, et al. Understanding women veterans' family planning counseling experiences and preferences to inform patient-centered care. Womens Health Issues 2019;29(3):283-289. [doi: 10.1016/j.whi.2019.03.002] [Medline: 30981559]
26. Thorburn S, Bogart LM. African American women and family planning services: perceptions of discrimination. Women Health 2005 Dec 05;42(1):23-39. [doi: 10.1300/j013v42n01_02]
27. Downing RA, LaVeist TA, Bullock HE. Intersections of ethnicity and social class in provider advice regarding reproductive health. Am J Public Health 2007 Oct;97(10):1803-1807. [doi: 10.2105/ajph.2006.092585]
28. Callegari LS, Aiken AR, Dehlendorf C, Cason P, Borrero S. Addressing potential pitfalls of reproductive life planning with patient-centered counseling. Am J Obstet Gynecol 2017 Feb;216(2):129-134. [doi: 10.1016/j.ajog.2016.10.004] [Medline: 27776920]
29. Dehlendorf C, Krajewski C, Borrero S. Contraceptive counseling: best practices to ensure quality communication and enable effective contraceptive use. Clin Obstet Gynecol 2014 Dec;57(4):659-673 [FREE Full text] [doi: 10.1097/GRF.0000000000000059] [Medline: 25264697]
30. Gomez AM, Fuentes L, Allina A. Women or LARC first? Reproductive autonomy and the promotion of long-acting reversible contraceptive methods. Perspect Sex Reprod Health 2014 Sep;46(3):171-175 [FREE Full text] [doi: 10.1363/46e1614] [Medline: 24861029]
31. Gubrium AC, Mann ES, Borrero S, Dehlendorf C, Fields J, Geronimus AT, et al. Realizing reproductive health equity needs more than Long-Acting Reversible Contraception (LARC). Am J Public Health 2016 Jan;106(1):18-19. [doi: 10.2105/ajph.2015.302900]
32. Callegari L, Edmonds S, Borrero S, Ryan G, Cusack C, Zephyrin L. Preconception care in the Veterans Health Administration. Semin Reprod Med 2018 Nov;36(6):327-339. [doi: 10.1055/s-0039-1678753] [Medline: 31003248]
33. Frayne S, Phibbs C, Saechao F, Friedman SA, Shaw JG, Romodan Y, et al. Sourcebook: Women Veterans in the Veterans Health Administration Volume 4: Longitudinal Trends in Sociodemographics, Utilization, Health Profile, and Geographic Distribution. Women's Health Evaluation Initiative, Women's Health Services, Veterans Health Administration, Department of Veterans Affairs. URL: https://www.womenshealth.va.gov/WOMENSHEALTH/docs/WHS_Sourcebook_Vol-IV_508c.pdf [accessed 2020-09-17]
34. Callegari LS, Nelson KM, Arterburn DE, Dehlendorf C, Magnusson SL, Benson SK, et al. Development and pilot testing of a patient-centered web-based reproductive decision support tool for primary care. J Gen Intern Med 2021 Oct;36(10):2989-2999. [doi: 10.1007/s11606-020-06506-6] [Medline: 33538956]
35. Reich ES. Pre-term births on the rise. Nature 2012 May 02;485(7396):20. [doi: 10.1038/485020a] [Medline: 22552070]
36. Ely DM, Driscoll AK. Infant mortality in the United States, 2018: data from the Period Linked Birth/Infant Death File. Natl Vital Stat Rep 2020 Jul;69(7):1-18. [Medline: 32730740]
37. Lipsky S, Holt VL, Easterling TR, Critchlow CW. Impact of police-reported intimate partner violence during pregnancy on birth outcomes. Obstet Gynecol 2003 Sep;102(3):557-564. [doi: 10.1016/s0029-7844(03)00573-8] [Medline: 12962943]
38. Grote NK, Bridge JA, Gavin AR, Melville JL, Iyengar S, Katon WJ. A meta-analysis of depression during pregnancy and the risk of preterm birth, low birth weight, and intrauterine growth restriction. Arch Gen Psychiatry 2010 Oct;67(10):1012-1024 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2010.111] [Medline: 20921117]
39. Witt WP, Cheng ER, Wisk LE, Litzelman K, Chatterjee D, Mandell K, et al. Maternal stressful life events prior to conception and the impact on infant birth weight in the United States. Am J Public Health 2014 Feb;104(Suppl 1):S81-S89 [FREE Full text] [doi: 10.2105/AJPH.2013.301544] [Medline: 24354829]

40. Krishnamurti T, Davis AL, Rodriguez S, Hayani L, Bernard M, Simhan HN. Use of a smartphone app to explore potential underuse of prophylactic Aspirin for preeclampsia. JAMA Netw Open 2021 Oct 01;4(10):e2130804 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.30804] [Medline: 34714341]
41. Collins KS, Hughes DL, Doty MM, Ives BL, Edwards JN, Tenney K. Diverse communities, common concerns: assessing health care quality for minority Americans. New York: Commonwealth Fund; Mar 01, 2022. URL: https://tinyurl.com/3j4kkzm7
42. Krishnamurti T, Davis AL, Wong-Parodi G, Fischhoff B, Sadovsky Y, Simhan HN. Development and testing of the MyHealthyPregnancy app: a behavioral decision research-based tool for assessing and communicating pregnancy risk. JMIR Mhealth Uhealth 2017 Apr 10;5(4):e42 [FREE Full text] [doi: 10.2196/mhealth.7036] [Medline: 28396302]
43. Mesner O, Davis A, Casman E, Simhan H, Shalizi C, Keenan-Devlin L, et al. Using graph learning to understand adverse pregnancy outcomes and stress pathways. PLoS One 2019;14(9):e0223319 [FREE Full text] [doi: 10.1371/journal.pone.0223319] [Medline: 31568495]
44. Krishnamurti T, Davis AL, Quinn B, Castillo AF, Martin KL, Simhan HN. Mobile remote monitoring of intimate partner violence among pregnant patients during the COVID-19 shelter-in-place order: quality improvement pilot study. J Med Internet Res 2021 Feb 19;23(2):e22790 [FREE Full text] [doi: 10.2196/22790] [Medline: 33605898]
45. Mora AC, Krishnamurti T, Davis A, Simhan H. 930: A culturally appropriate mobile health application for pregnancy risk communication to latino women. Am J Obstet Gynecol 2020 Jan;222(1):S575-S576. [doi: 10.1016/j.ajog.2019.11.941]
46. Daniels K, Daugherty J, Jones J, Mosher W. Current contraceptive use and variation by selected characteristics among women aged 15-44: United States, 2011-2013. Natl Health Stat Report 2015 Nov 10(86):1-14 [FREE Full text] [Medline: 26556545]
47. Borrero S, Schwarz EB, Reeves MF, Bost JE, Creinin MD, Ibrahim SA. Race, insurance status, and tubal sterilization. Obstet Gynecol 2007 Jan;109(1):94-100. [doi: 10.1097/01.AOG.0000249604.78234.d3] [Medline: 17197593]
48. Schmidt JE, Hillis SD, Marchbanks PA, Jeng G, Peterson HB. Requesting information about and obtaining reversal after tubal sterilization: findings from the U.S. Collaborative Review of Sterilization. Fertil Steril 2000 Nov;74(5):892-898. [doi: 10.1016/s0015-0282(00)01558-2] [Medline: 11056229]
49. Borrero S, Nikolajski C, Rodriguez KL, Creinin MD, Arnold RM, Ibrahim SA. "Everything I know I learned from my mother...Or not": perspectives of African-American and white women on decisions about tubal sterilization. J Gen Intern Med 2009 Mar;24(3):312-319 [FREE Full text] [doi: 10.1007/s11606-008-0887-3] [Medline: 19107540]
50. Gilliam M, Davis SD, Berlin A, Zite NB. A qualitative study of barriers to postpartum sterilization and women's attitudes toward unfulfilled sterilization requests. Contraception 2008 Jan;77(1):44-49. [doi: 10.1016/j.contraception.2007.09.011] [Medline: 18082666]
51. Hahn TA, McKenzie F, Hoffman SM, Daggy J, Tucker Edmonds B. A prospective study on the effects of Medicaid regulation and other barriers to obtaining postpartum sterilization. J Midwifery Womens Health 2019 Mar;64(2):186-193. [doi: 10.1111/jmwh.12909] [Medline: 30411465]
52. Jain R, Kazmerski TM, Zuckerwise LC, West NE, Montemayor K, Aitken ML, et al. Pregnancy in cystic fibrosis: Review of the literature and expert recommendations. J Cyst Fibros 2021 Aug 26:online ahead of print [FREE Full text] [doi: 10.1016/j.jcf.2021.07.019] [Medline: 34456158]
53. Cooper GS, Stroehla BC. The epidemiology of autoimmune diseases. Autoimmun Rev 2003 May;2(3):119-125. [doi: 10.1016/s1568-9972(03)00006-5] [Medline: 12848952]
54. Kazmerski TM, West NE, Jain R, Uluer A, Georgiopoulos AM, Aitken ML, et al. Family-building and parenting considerations for people with cystic fibrosis. Pediatr Pulmonol 2022 Feb;57(Suppl 1):S75-S88. [doi: 10.1002/ppul.25620] [Medline: 34407321]
55. Birru Talabi M, Clowse MEB, Schwarz EB, Callegari LS, Moreland L, Borrero S. Family planning counseling for women with rheumatic diseases. Arthritis Care Res 2018 Feb;70(2):169-174. [doi: 10.1002/acr.23267] [Medline: 28437837]
56. Wolgemuth T, Stransky OM, Chodoff A, Kazmerski TM, Clowse MEB, Birru Talabi M. Exploring the preferences of women regarding sexual and reproductive health care in the context of rheumatology: a qualitative study. Arthritis Care Res 2021 Aug;73(8):1194-1200. [doi: 10.1002/acr.24249] [Medline: 32374928]
57. Kazmerski TM, Borrero S, Sawicki GS, Abebe KZ, Jones KA, Tuchman LK, et al. Provider attitudes and practices toward sexual and reproductive health care for young women with cystic fibrosis. J Pediatr Adolesc Gynecol 2017 Oct;30(5):546-552. [doi: 10.1016/j.jpag.2017.01.009] [Medline: 28163120]
58. Birru Talabi M, Clowse MEB, Blalock SJ, Hamm M, Borrero S. Perspectives of adult rheumatologists regarding family planning counseling and care: a qualitative study. Arthritis Care Res 2020 Mar 27;72(3):452-458 [FREE Full text] [doi: 10.1002/acr.23872] [Medline: 30875455]
59. Toomey D, Waldron B. Family planning and inflammatory bowel disease: the patient and the practitioner. Fam Pract 2013 Feb;30(1):64-68. [doi: 10.1093/fampra/cms035] [Medline: 22843639]
60. Birru Talabi M, Callegari LS, Borrero S. Redefining to include reproductive autonomy: a new paradigm in subspecialty medicine. Womens Health Rep 2021;2(1):497-499 [FREE Full text] [doi: 10.1089/whr.2021.0079] [Medline: 34970654]

61. Leech MM, Stransky OM, Talabi MB, Borrero S, Roe AH, Kazmerski TM. Exploring the reproductive decision support needs and preferences of women with cystic fibrosis. Contraception 2021 Jan;103(1):32-37. [doi: 10.1016/j.contraception.2020.10.004] [Medline: 33068614]
62. Stransky OM, Pam M, Ladores SL, Birru Talabi M, Borrero S, Godfrey EM, et al. Engaging stakeholders in the development of a reproductive goals decision aid for women with cystic fibrosis. J Patient Exp 2022;9:23743735221077527 [FREE Full text] [doi: 10.1177/23743735221077527] [Medline: 35128042]
63. Gomez I, Rnaji U, Salganicoff A, Frederiksen B. Medicaid Coverage for Women. KFF. 2022 Feb 17. URL: https://www.kff.org/womens-health-policy/issue-brief/medicaid-coverage-for-women/ [accessed 2022-03-14]

Abbreviations

ACOG: American College of Obstetricians and Gynecologists
CF: cystic fibrosis
IPV: Intimate Partner Violence
RMD: rheumatic and musculoskeletal disease
VA: Veterans Affairs

Edited by G Eysenbach, T Leung; submitted 11.01.22; peer-reviewed by S Panahi, SS Amritphale, N Levitz; comments to author 21.02.22; revised version received 14.03.22; accepted 01.04.22; published 28.04.22.

Please cite as:
Krishnamurti T, Birru Talabi M, Callegari LS, Kazmerski TM, Borrero S
A Framework for Femtech: Guiding Principles for Developing Digital Reproductive Health Tools in the United States
J Med Internet Res 2022;24(4):e36338
URL: https://www.jmir.org/2022/4/e36338
doi:10.2196/36338
PMID:35482371

Tutorial

Methodological Guidelines for Systematic Assessments of Health Care Websites Using Web Analytics: Tutorial

Edwin Lauritz Fundingsland Jr[1], BA, DO; Joseph Fike[2], BSc, DO; Joshua Calvano[3], BSc, DO; Jeffrey Beach[4], BSc, DO; Deborah Lai[5], BA; Shuhan He[6], MD

[1]Department of Emergency Medicine, Homer Stryker MD School of Medicine, Western Michigan University, Kalamazoo, MI, United States

[2]Department of Physical Medicine and Rehabilitation, Medical Center, University of Nebraska, Omaha, NE, United States

[3]Department of Anesthesiology and Critical Care, University of New Mexico Hospital, Albuquerque, NM, United States

[4]College of Osteopathic Medicine, Rocky Vista University, Parker, CO, United States

[5]Division of Psychology and Language Sciences, University College London, London, United Kingdom

[6]Center for Innovation in Digital HealthCare, Lab of Computer Science, Massachusetts General Hospital, Boston, MA, United States

Corresponding Author:
Shuhan He, MD
Center for Innovation in Digital HealthCare
Lab of Computer Science
Massachusetts General Hospital
55 Fruit St.
Boston, MA, 02114
United States
Phone: 1 (617) 726 2000
Email: she@mgh.harvard.edu

Abstract

With the growing importance of communicating with the public via the web, many industries have used web analytics to provide information that organizations can use to better achieve their goals. Although the importance of health care websites has also grown, the health care industry has been slower to adopt the use of web analytics. Web analytics are the measurement, collection, analysis, and reporting of internet data used to measure direct user interaction. Our objective is to provide generalized methods for using web analytics as key performance metrics to evaluate websites and outline actionable recommendations for improvement. By deconstructing web analytic categories such as engagement, users, acquisition, content, and platform, we describe how web analytics are used to evaluate websites and how improvements can be made using this information. *Engagement* is how a user interacts with a website. It can be evaluated using the daily active users to monthly active users (DAU/MAU) ratio, bounce rate, pages viewed, and time on site. Poor engagement indicates potential problems with website usability. *Users* pertains to demographic information regarding the users interacting with a website. This data can help administrators understand who is engaging with their website. *Acquisition* refers to the overall website traffic and the method of traffic, which allows administrators to see how people are accessing their website. This information helps websites expand their methods of attracting users. *Content* refers to the overall relevancy, accuracy, and trustworthiness of a website's content. If a website has poor content, it will likely experience difficulty with user engagement. Finally, *platform* refers to the technical aspects of how people access a website. It includes both the internet browsers and devices used. By providing detailed descriptions of these categories, we have identified how web administrators can use web analytics to systematically assess their websites. We have also provided generalized recommendations for actionable improvements. By introducing the potential of web analytics to augment usability and the conversion rate, we hope to assist health care organizations in better communicating with the public and therefore accomplishing the goals of their websites.

(J Med Internet Res 2022;24(4):e28291) doi:10.2196/28291

KEYWORDS

Google Analytics; website usability; conversion rate; website engagement; user demographics; website traffic; website content; internet browsers; healthcare websites; web analytics; healthcare industry; usability

Introduction

Background

With the continually growing global importance of the World Wide Web, websites have become a crucial communication channel for corporations, political groups, and organizations because of their capability to rapidly disseminate information to various audiences at a low cost [1]. Web analytics has become a mainstay of commercial industries and even a commercial industry itself. The web analytics market was valued at US $2.63 billion in 2018 and is projected to reach US $10.73 billion by 2026, growing at a compound annual growth rate of 19.3% from 2019 to 2026 [2]. The field of medicine, however, remains hindered as stakeholders in health care have been slow to adopt digital innovations. Studies have shown that the adoption of digital technologies can improve the performance of health care processes, increase efficiency, and enable the delivery of higher-quality care and reduced response times, with many benefits for several stakeholders, such as national health systems, clinicians, and patients [3]. For organizations to achieve their goals in use and impact, their website's communication capacity is key. A website that cannot effectively communicate is not serving its purpose. Communication capacity can be measured through the usability and conversion rate of a website [1].

Studies have shown a relationship between the usability of health care websites and the credibility ascribed by its users [4]. The International Organization for Standardization defines usability as "the extent to which a system, product or service can be used by specific users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" [5]. Measures of effectiveness, efficiency, and satisfaction can be viewed as key web analytic metrics and if optimized, can lead to increased website success. By augmenting usability, a website can reach a higher level of engagement and achieve its desired objectives. A lack of design errors, following established design conventions, and ease of navigation are important features emphasized in the literature [6]. When users have difficulty accessing or using a website, they are likely to move on to another resource, while a website that uses usability metrics is more likely to retain users.

Other industries have established user expectations for their respective websites; health care websites are facing the need to conform [7,8]. Studies have been conducted which evaluate usability in areas such as e-commerce, e-government, mobile news apps, and library websites [9-12]. More recently, there have been increasing usability studies focusing on websites within the health care sector, such as websites for emergency medicine residency programs, digital health care centers, hospitals, and cancer centers [13-16]. With the growing importance of website usability and the conversion rate in the health care sector, web analytics can provide health care stakeholders with an easily accessible tool to assist their evaluation of usability and measure conversion rate.

The conversion rate is closely intertwined with usability. It measures the number of users who perform the desired goal of the page (ie, buying a product or filling out a form) relative to the total users [17]. A high conversion rate separates a successful website from an unsuccessful website.

Web analytics refers to the collection, analysis, and reporting of internet data for the purposes of understanding and optimizing web use [18]. On-site web analytics are used to measure direct user interaction, such as the number of visitors, time spent on a website, and click path [19]. Overall, web analytics can contribute to determining a website's usability and conversion rates [1]. This collection of data is used to analyze the performance of a site and can allow websites to improve their persuasion and relevance [20]. It is important to note, however, that web analytics do not provide a comprehensive measurement of website usability. Measurement of usability consists of additional variables that are outside the scope of this publication.

To the authors' knowledge, there is no recent paper outlining how web analytics can be applied broadly to the field of health care. By evaluating the categories of engagement, users, acquisition, content, and platform, we aim to create a universal framework of web analytics that can be applied to health care websites to improve the quality and effectiveness of these websites.

Objectives

This aim of this tutorial is to (1) provide a basic understanding of definitions and methods pertaining to web analytics, (2) create a framework for using web analytics to evaluate the effectiveness of health care websites specifically, and (3) outline the actionable implications of web analytics to assist health care websites in achieving their goals.

Methods

Google Analytics

Google Analytics (GA) is a web analytics service that has been offered by Google since 2005. It is the most widely used web analytics tool, with 84.1% of the market share [21]. It can be used for both websites and apps, across iOS and Android devices. As of August 2013, GA was reportedly used by 66.2% of the 10,000 most popular websites [22]. GA offers a free version that can be used by those with a graphical user interface and without software engineering skills. Any owner of a website or app can sign up for a GA account.

Engagement

An engagement analysis evaluates user activity and is one of the most used analytic tools. It describes how users interact with websites [19]. Factors that are often addressed include how often visitors return to the site, how often new visitors become returning visitors, pages visited per session, and duration of visits [19].

Daily Active User to Monthly Active User (DAU/MAU) Ratio

When evaluating overall engagement, *1-day active users* refers to users who have been active at least once in the previous day, *7-day active users* refers to users who have been active at least once in the previous 7 days, and so on for 14- and 28-day active users. 1-day active users are referred to as daily active users

(DAU) and 28-day active users are referred to as monthly active users (MAU). The ratio of DAUs to MAUs, DAU/MAU, can be expressed as a percentage to understand user engagement; this measure was first popularized by Facebook and has since become a popular key performance indicator (KPI), with some venture capitalists considering a ratio of over 20% favorable and over 50% excellent [23]. Assessing the DAU/MAU ratio can indicate whether a site is attracting users at its intended or expected frequency. For instance, a rideshare website may expect to see a high DAU/MAU ratio. On the other hand, a flight booking site may see a lower DAU/MAU ratio. Regarding health care websites, many sites have goals of continued user involvement. A low DAU/MAU ratio can be used as an indication that the relevancy of the content and usability of a website can be improved. It is important to view this metric in the context of a website's desired goals. For health care websites that do not desire continued engagement, this may not be a relevant metric and therefore would not correlate with usability issues.

Bounce Rate

The *bounce rate* is another metric of engagement referring to the percentage of single-session users (ie, users who visit the site and "bounce" without interacting further, as opposed to users who interact with at least 1 additional page). A session is recorded by GA each time a user visits the site, beginning as soon as the site is first loaded and ending after 30 minutes of inactivity. Using this metric, the website host is provided with insight into the user's engagement with their product. Navigating to other pages of a website or application is typically viewed as an active event triggered by the user. Similar to the DAU/MAU ratio, the bounce rate indicates users who are not achieving the desired interaction with the website. A high bounce rate may suggest a usability issue steering users away from the page. More than simply content, many things can cause users to avoid visiting additional pages. One example is slow loading speeds. If a website is loading too slowly, users may leave the site before viewing any of the content. According to a recent Google study, a website that takes longer than 3 seconds to load on a mobile device loses approximately 53% of its users and the average mobile website speed is around 18 seconds [24]. This issue can be addressed simply by reducing conflicting technology on the back-end server [13,14]. As with the DAU/MAU ratio, web administrators must view this metric in the proper context. If they do not desire continual engagement within a single session, the bounce rate is not a useful metric.

Page Views

The number of pages per session is the number of pages within the site that a user visits during a single session and indicates how thoroughly a user is engaging with a website. A page view is counted every time a website is loaded, and this can be tracked using GA [1]. Similar to the bounce rate, if users are accessing a website but not interacting with additional pages, there may be an issue with its usability. The goal of many websites is for people to view subpages with additional content but various issues could interrupt this, one being front-end web page design. If users are not easily finding links to subpages, they may lose interest and bounce. By working with marketing specialists, web administrators can improve their webpage design and ease user navigation. As with the bounce rate, if continual engagement within a session is not a desired outcome, this metric is not a helpful measure of engagement.

Time on Site

As the name suggests, *time on site* refers to the duration of time a user spends on a website. If the same visitor comes back several hours later or the next day, a new session is counted. This is considered a key indicator of how successfully a website is engaging visitors. It has been suggested that time on site is an indication of website usability. However, this is operating under the notion that the greater the usability of a website, the more time a user will spend on it. A long session duration may suggest that users are spending more time reviewing the detail of a website's content, while a short duration may suggest poor usability. It is important to analyze time on site and page views together to dissect whether users are spending increased time on the site due to difficulty navigating it [1].

Users

Analytics can help health care centers understand who the users of their websites are. GA provides limited demographic information about users, including age and gender distribution and location. If a website has a target audience, they can monitor if they are reaching that demographic. If they are targeting a diverse population of users, they can also use these demographics to monitor their success. Using this information, web creators can better focus their efforts on the population viewing their website or target those who are not using the site. For example, if it is discovered that males over the age of 65 years are primarily accessing a men's health website, the web administrators would know they are reaching part of their target demographic. However, they may want to make efforts in marketing to younger users as well.

Acquisition

By employing use data to understand consumer needs, websites can increase their user acquisition [19]. *Acquisition* refers to the amount of traffic a website receives. *Sources* refer to the origins of a user's traffic to the site. If the overall user volume of a website is low, the method of traffic can be an important variable to address to reach more users. By using acquisition data, administrators can see where there is room for improvement in reaching potential users.

Direct Traffic

Direct traffic refers to visitors who arrive on the site directly by typing the URL into the browser address bar, clicking on a bookmark, or clicking on a link in an email, text message, or chat. Direct traffic can be a strong indicator of brand strength as well as success in email, text message, and offline marketing. If a website is experiencing low volumes of direct traffic, they can increase their efforts in these forms of marketing and in improving their overall brand strength.

Referral Traffic

Referral traffic refers to visitors who arrived at the site via another website. This occurs when outside websites contain links to a given site. If referral traffic is low, websites can place

more emphasis on promotion via other websites. For example, if an organization has multiple websites under their umbrella, they can use their content on one website to direct users to their other. Referral traffic also encourages organizations to increase website partnerships that mutually benefit both parties.

Organic Traffic

Organic traffic refers to visitors who arrived at the site via a search result page (eg, Google or Bing) and can be an indicator of strong content or search engine optimization (SEO). SEO is a method for increasing organic traffic that has gained popularity in many industries. Many users find websites by simply entering keywords into a search engine and choosing the website that seems most appropriate. By strategically strengthening a website's content, web administrators can help move their website closer to the top of a search engine results page (SERP) and therefore increase traffic. The nuances of SEO are outside the scope of this publication.

Social Traffic

Social traffic is similar to referral traffic, but it refers to traffic from social media platforms as opposed to traffic from other websites. Websites that are receiving low levels of social traffic can seek to implement, improve, and promote their social media presence on platforms such as Facebook, Twitter, and Instagram. Improved social media presence can also improve overall brand awareness.

Content

Assessments of a website's content can refer to the relevancy of information, the quality of multimedia content, and even grammar and spelling [13,14]. One of the most obvious reasons a website may not achieve its goals is its content not meeting the needs of users.

Relevancy

Concerning relevancy, the following questions should be posed: is a website's information up to date and fact-driven, and does it provide answers people are seeking [13,14]? If the answer to any of these questions is no, users will not engage with a website. If a site is concerned with relevance, a solution may be to increase the frequency of content updates to ensure the information provided is not out of date. Additionally, it is particularly important for health care–related sites to have accurate and fact-driven content. Especially pertaining to health-related information, users will not engage with a website they believe to contain inaccurate information.

Multimedia Content

Multimedia content can be evaluated by quantity and quality of resolution [13,14]. Seeking to further augment their content, websites can use multimedia to make their content more dynamic. Increasing the quality and quantity of videos, graphics, and animations has been shown to increase user engagement [13,14].

Spelling and Grammar

Spelling and grammar are important aspects of content quality. Even if a website's content is up to date and accurate, users still may not trust it if there are obvious spelling and grammatical errors. There are easily accessible spelling and grammar tools available for websites to avoid this issue.

Platform

To better understand potential areas of improvement for a website, engagement can be evaluated on each page. One can also assess the different browsers and devices through which users access a website to identify technical areas of improvement.

Browser

A *browser* is the software application used to access the internet. Common browsers include Google Chrome (Google LLC), Internet Explorer (Microsoft Corp), and Safari (Apple Inc). If a website is not easily accessible on all major internet browsers, web administrators are automatically eliminating potential users. We have already discussed the benefits of SEO to improve the placement of a site on a SERP, but the improvement discussed here is made from a technological perspective. For example, if a website uses Java, a Google Chrome browser will not be able to support it [25]. It is wise for web developers to tailor their websites to the browsers of their users. GA data can provide information about which browsers are being used to access the website.

Device

Users are accessing websites on various devices, namely desktop computers or laptops, mobile phones, and tablets. Similar to browsers, if a website cannot be accessed on all devices, this eliminates an entire category of potential users. Tablets and smartphones are more commonly being used to access the internet; therefore, it is important that websites are mobile friendly. Administrators have the option to make separate mobile websites, but with mobile devices becoming more sophisticated, new methods have developed. One new and simpler method known as responsive design allows for the creation of one web page, then uses multiple sets of CSS rules to adjust formatting of the website to fit the size of the browser window [26].

Discussion

Main Recommendations

By providing detailed descriptions of categories such as engagement, users, acquisition, content, and platform, we have identified how web administrators can use web analytics to systematically assess their websites using tools such as GA. We have also provided generalized recommendations for actionable improvements that can be made to address website weaknesses.

Although web analytics may be at an infant stage in the world of health care, it is very prevalent in other industries. By introducing the potential benefits of web analytics in the health care sector, we hope to continue the standardization of web practices that users have become accustomed to. Using web analytic tools in the proper context, health care website administrators can gain more information on user engagement and use this information to make improvements.

With the health care industry being slow to adapt to standards for website usability, we hope that the outlined methods and

recommendations for using web analytics can be directed toward areas in need of improvement and increase the websites' conversion rates. These recommendations can make a significant impact for health care organizations because they are actionable at a low cost. The potential of a website to improve persuasion and relevance has been established and by using web analytics, web administrators can easily expand upon this potential with a smaller financial burden compared to other methods.

Limitations

The approaches outlined in this paper are intended to be broadly generalizable to health care–related websites such that they can be used by a wide spectrum of web administrators in the health care industry. However, each organization should tailor this approach to their unique objectives and considerations. This content serves primarily as an introduction to the potential benefits and methods of using web analytics, and future studies may focus on more specific use cases, such as applications for subfields in health care.

Key web analytic metrics are not a comprehensive method for evaluating website usability. In certain cases, a degree of inference must be made to use web analytics as a reflection of a website's usability. For example, the conversion rate can be used as a measure of a website's effectiveness. However, if those viewing a website are not its targeted users, a poor conversion rate does not necessarily reflect poor usability. This underscores the importance of using various web analytic measures to gain a comprehensive perspective of user interaction. In the given scenario, administrators could examine the demographic characteristics of their websites users to determine if there is in fact an issue with usability. Similarly, metrics like the DAU/MAU ratio, bounce rate, and page views are used as a measure of website engagement, but it remains important to consider these measures within the context of a website's targeted users and objectives. If it is not a website's goal to promote continual access, the DAU/MAU ratio is not a useful measure for usability. Similarly, if it is not a website's goal to foster continual engagement within each session, page views and the bounce rate are not useful.

Finally, these metrics are only one aspect of the overall capabilities of website usability analysis. Other methods to evaluate usability include user interviews and on-page heat mapping. Future studies delving into these methods would help improve our understanding of website usability in health care–related websites.

Conclusions

Websites continue to be a primary method by which health care organizations interact with their consumers; however, the health care sector lags behind many other industries in using accepted and standardized website usability practices. With evidence pointing to the efficacy of using web analytics to augment the usability and conversion rate, health care organizations can benefit from adopting these practices to better accomplish the goals of their websites.

Authors' Contributions

ELF wrote the manuscript. JB contributed to the manuscript revisions. JF contributed to the final version of the manuscript. JC conceived the original idea and contributed to writing the manuscript. DL conceived the original idea and contributed to writing the manuscript. SH conceived original idea and supervised the project. All authors reviewed the final manuscript.

Conflicts of Interest

SH serves on the advisory board of Covid Act Now and Safeter. SH is Cofounder and Executive Board Director of GetUsPPE (unpaid) and ConductScience. SH serves on the American College of Emergency Physicians Supply Chain Task Force. SH has received research funding from the Foundation for Opioid Response Efforts, royalties from MazeEngineers, and personal fees from Withings Inc, the Boston Globe, the American College of Emergency Physicians, ConductScience, Curative Medical Associates, and VIO Med Spa New England. SH is a volunteer at Emojination. The other authors report no conflicts of interest.

References

1. Beri B, Singh P. Web analytics: increasing website's usability and conversion rate. Int J Comput Appl 2013 Jun 26;72(6):35-38. [doi: 10.5120/12501-8420]
2. Global web analytics market to reach $10.73 billion by 2026: allied market research. Allied Market Research. 2020. URL: https://www.globenewswire.com/news-release/2020/07/02/2057046/0/en/Global-Web-Analytics-Market-to-Reach-10-73-Billion-by-2026-Allied-Market-Research.html#:~:text=Portland%2C [accessed 2022-04-01]
3. Laurenza E, Quintano M, Schiavone F, Vrontis D. The effect of digital technologies adoption in healthcare industry: a case based analysis. Bus Process Manag J 2018 Sep 03;24(5):1124-1144. [doi: 10.1108/bpmj-04-2017-0084]
4. Theng YL, Soh ES. An asian study of healthcare web portals: implications for healthcare digital libraries. In: Proceedings of the 8th International Conference on Asian Digital Libraries: Implementing Strategies and Sharing Experiences. Digital Libraries: Implementing Strategies and Sharing Experiences. Springer Berlin Heidelberg; 2005 Presented at: 8th International Conference on Asian Digital Libraries; December 12-15, 2005; Bangkok, Thailand p. 52-61. [doi: 10.1007/11599517_7]
5. ISO 9241-11:2018 ergonomics of human-system interaction — part 11: usability: definitions and concepts. International Organization for Standardization. URL: https://www.iso.org/standard/63500.html [accessed 2022-04-01]
6. Cappel J, Huang Z. A usability analysis of company websites. J Comput Inf Syst 2007;48(1):117-123. [doi: 10.1080/08874417.2007.11646000]

7. Mittler J, Volmar K, Shaw B, Christianson J, Scanlon D. Using websites to engage consumers in managing their health and healthcare. Am J Manag Care 2012 Sep;18(6 Suppl):s177-s184. [Medline: 23286713]
8. Liang CJ, Chen HJ. A study of the impacts of website quality on customer relationship performance. Total Qual Manag Bus Excell 2009 Sep;20(9):971-988. [doi: 10.1080/14783360903181784]
9. Kuan HH, Bock GW, Vathanophas V. Comparing the effects of usability on customer conversion and retention at e-commerce websites. In: Proceedings of the 38th Annual Hawaii International Conference on System Sciences. 2005 Presented at: 38th Annual Hawaii International Conference on System Sciences; January 3-6, 2005; Big Island, HI p. 174a. [doi: 10.1109/hicss.2005.155]
10. Huang Z, Benyoucef M. Usability and credibility of e-government websites. Gov Inf Q 2014 Oct;31(4):584-595. [doi: 10.1016/j.giq.2014.07.002]
11. Jeong W, Jung Han H. Usability study on newspaper mobile websites. OCLC Systems & Services 2012 Oct 26;28(4):180-198. [doi: 10.1108/10650751211279120]
12. Harinarayana N, Raju NV. An analysis of usability features of library web sites. Ann Libr Inf Stud 2008;55:111-122.
13. Fundingsland E, Fike J, Calvano J, Raja A, Lai D, Silacci S, et al. Website usability analysis of United States emergency medicine residencies. AEM Educ Train 2021 Jul;5(3):e10604. [doi: 10.1002/aet2.10604] [Medline: 34124531]
14. Calvano JD, Fundingsland EL, Lai D, Silacci S, Raja AS, He S. Applying website rankings to digital health centers in the United States to assess public engagement: website usability study. JMIR Hum Factors 2021 Mar 29;8(1):e20721 [FREE Full text] [doi: 10.2196/20721] [Medline: 33779564]
15. Huerta TR, Hefner JL, Ford EW, McAlearney AS, Menachemi N. Hospital website rankings in the United States: expanding benchmarks and standards for effective consumer engagement. J Med Internet Res 2014 Feb 25;16(2):e64 [FREE Full text] [doi: 10.2196/jmir.3054] [Medline: 24568892]
16. Huerta TR, Walker DM, Ford EW. Cancer center website rankings in the USA: expanding benchmarks and standards for effective public outreach and education. J Cancer Educ 2017 Jun;32(2):364-373. [doi: 10.1007/s13187-015-0931-z] [Medline: 26472325]
17. Stamoulis N. Definition of a conversion rate. Brick Marketing. URL: http://www.brickmarketing.com/define-conversion-rate.htm [accessed 2022-04-01]
18. Web analytics basics. Usability.gov. URL: https://www.usability.gov/what-and-why/web-analytics.html#:~:text=Web [accessed 2022-04-01]
19. Zheng G, Peltsverger S. Web analytics overview. In: Khosrow-Pour M, editor. Encyclopedia of Information Science and Technology, Third Edition. Hershey, PA: Information Science Reference; 2015:7674-7683.
20. Waisberg D, Kaushik A. Web analytics 2.0: empowering customer centricity. SEMJ 2009;2(1):7.
21. Usage statistics of traffic analysis tools for websites. W3Techs - World Wide Web Technology Surveys. URL: https://w3techs.com/technologies/overview/traffic_analysis [accessed 2022-04-01]
22. Google Analytics usage statistics. Built With. URL: https://trends.builtwith.com/analytics/Google-Analytics [accessed 2022-04-01]
23. Chen A. DAU/MAU is an important metric to measure engagement, but here's where it fail. andrewchen.com. URL: https://andrewchen.co/dau-mau-is-an-important-metric-but-heres-where-it-fails [accessed 2022-04-01]
24. An D. Find out how you stack up to new industry benchmarks for mobile page speed. Think With Google. 2018. URL: https://www.thinkwithgoogle.com/marketing-resources/data-measurement/mobile-page-speed-new-industry-benchmarks/ [accessed 2022-04-01]
25. Today I learned: why some websites don't work properly in different browsers. University of Pittsburgh Information Technology. URL: https://www.fis.pitt.edu/today-i-learned-why-some-websites-dont-work-properly-in-different-browsers/#:~:text=Each [accessed 2022-04-01]
26. Warner J. What's the difference between adaptive and responsive web design? Digital Family. 2013. URL: http://www.digitalfamily.com/tutorials/css-article/whats-the-difference-between-adaptive-and-responsive-web-design/ [accessed 2022-04-01]

Abbreviations

DAU: daily active users
DAU/MAU: daily active users to monthly active users
GA: Google Analytics
MAU: monthly active users
SEO: Search Engine Optimization
SERP: search engine results page

Edited by A Mavragani; submitted 27.02.21; peer-reviewed by B Chaudhry, J Brooke; comments to author 10.05.21; revised version received 27.05.21; accepted 16.03.22; published 15.04.22.

Please cite as:
Fundingsland Jr EL, Fike J, Calvano J, Beach J, Lai D, He S
Methodological Guidelines for Systematic Assessments of Health Care Websites Using Web Analytics: Tutorial
J Med Internet Res 2022;24(4):e28291
URL: https://www.jmir.org/2022/4/e28291
doi:10.2196/28291
PMID:35436216

Tutorial

From the United Kingdom to Australia—Adapting a Web-Based Self-management Education Program to Support the Management of Type 2 Diabetes: Tutorial

Jenny Olson[1,2,3], BSc (Hons), PhD; Michelle Hadjiconstantinou[4], BSc, MSc, PhD, CPsychol; Carly Luff[1], BSc, Grad Cert; Karen Watts[1], MBA; Natasha Watson[1], BSc (Hons), Grad Dip; Venus Miller[5], BSc (Hons); Deborah Schofield[1], BAppSc, Grad Dip; Kamlesh Khunti[4], MD, PhD; Melanie J Davies[4], MD, ChB, MB; Sara Calginari[1], BSc (Hons), MBA

[1]Diabetes WA, Subiaco, Australia

[2]Department of Kinesiology, The Pennsylvania State University, University Park, PA, United States

[3]College of Medicine, The Pennsylvania State University, Hershey, PA, United States

[4]Diabetes Research Centre, University of Leicester, Leicester, United Kingdom

[5]Catalysis Research and Evaluation Consultancy, Perth, Australia

Corresponding Author:
Jenny Olson, BSc (Hons), PhD
Department of Kinesiology
The Pennsylvania State University
Recreation Hall
University Park, PA, 16801
United States
Phone: 1 8143083411
Email: Jenny.Olson@psu.edu

Abstract

Diabetes self-management education and support can improve outcomes in people with diabetes. Providing health interventions via digital modes of delivery can extend the reach of programs delivered through traditional means. The web-based version of the Diabetes Education and Self-Management for Ongoing and Newly Diagnosed (MyDESMOND) is a digital diabetes education and support program for people with type 2 diabetes. The program was originally developed in the United Kingdom and is evidence-based, grounded in behavioral theory, and designed through a rigorous process of intervention mapping. As such, MyDESMOND was considered an ideal candidate for adaptation to the Australian setting. Program content and the digital platform were modified to suit the local context to increase the likelihood that the revised version of MyDESMOND will deliver similar outcomes to the original program. The aim of this paper is to describe the systematic processes undertaken to adapt the digital MyDESMOND diabetes education and support program for people with type 2 diabetes to the Australian setting. The adaptation involved a multidisciplinary group with a diverse range of skills and expertise—a governance structure was established, a skilled project team was appointed, and stakeholder engagement was strategically planned. The adaptation of the program content included modifications to the clinical recommendations, the inclusion of local resources, practical changes, and revisions to optimize readability. A 2-stage independent review of the modified content was enacted. Digital adaptations were informed by relevant standards, local legislative requirements, and considerations of data sovereignty. The digital platform was extensively tested before deployment to the production setting. MyDESMOND is the first evidence-based digital diabetes education and support program for Australians with type 2 diabetes. This paper provides a road map for the adaptation of digital health interventions to new contexts.

(J Med Internet Res 2022;24(4):e26339) doi:10.2196/26339

KEYWORDS

diabetes mellitus; type 2; technology; self-management

Introduction

Background

Diabetes is a chronic condition defined by high plasma glucose levels that can lead to an increased risk of serious health complications and premature mortality [1]. The incidence of type 2 diabetes mellitus (T2DM) is increasing worldwide [2]. In Australia, >1.2 million people are estimated to have diabetes [3]. T2DM is the most common form of diabetes [4], with almost 1 million Australian adults diagnosed with this condition [5]. Moreover, it is estimated that there are an additional 500,000 undiagnosed cases of T2DM in Australia and a further 2 million people at high risk of developing the condition in the future [3].

A high proportion of people with T2DM experience long-term complications, including kidney disease, retinopathy, amputations, heart attacks, and stroke, associated with reduced life expectancy [6]. Optimal management of T2DM to reduce the risk of long-term complications requires the adoption and maintenance of self-care behaviors, including healthy eating, physical activity, blood glucose monitoring, medication adherence, and behaviors that reduce the risk of complications (eg, foot checks), in addition to the application of cognitive strategies to facilitate problem solving and healthy coping [7].

Diabetes self-management education and support (DSMES) programs empower people with diabetes to be actively involved in their own self-care by supporting informed decision-making and the adoption and maintenance of self-care strategies and behaviors [8]. DSMES has shown to be effective in reducing glycated hemoglobin A_{1c} (HbA_{1c}) levels [9] and decreasing the risk of all-cause mortality [10], as well as being cost-effective compared with usual care [11]. One such program is the Diabetes Education and Self-Management for Ongoing and Newly Diagnosed (DESMOND) program available in the United Kingdom for people with T2DM, which was evaluated in a multicenter randomized controlled trial and was also found to be cost-effective [12,13]. This program was adapted and developed for the Australian population and found to be effective in increasing patient activation (ie, active involvement in self-management of health conditions) among people with T2DM in regional Western Australia (WA) [14]. DESMOND has since been incorporated into the suite of DSMES programs delivered and evaluated nationally throughout Australia [15].

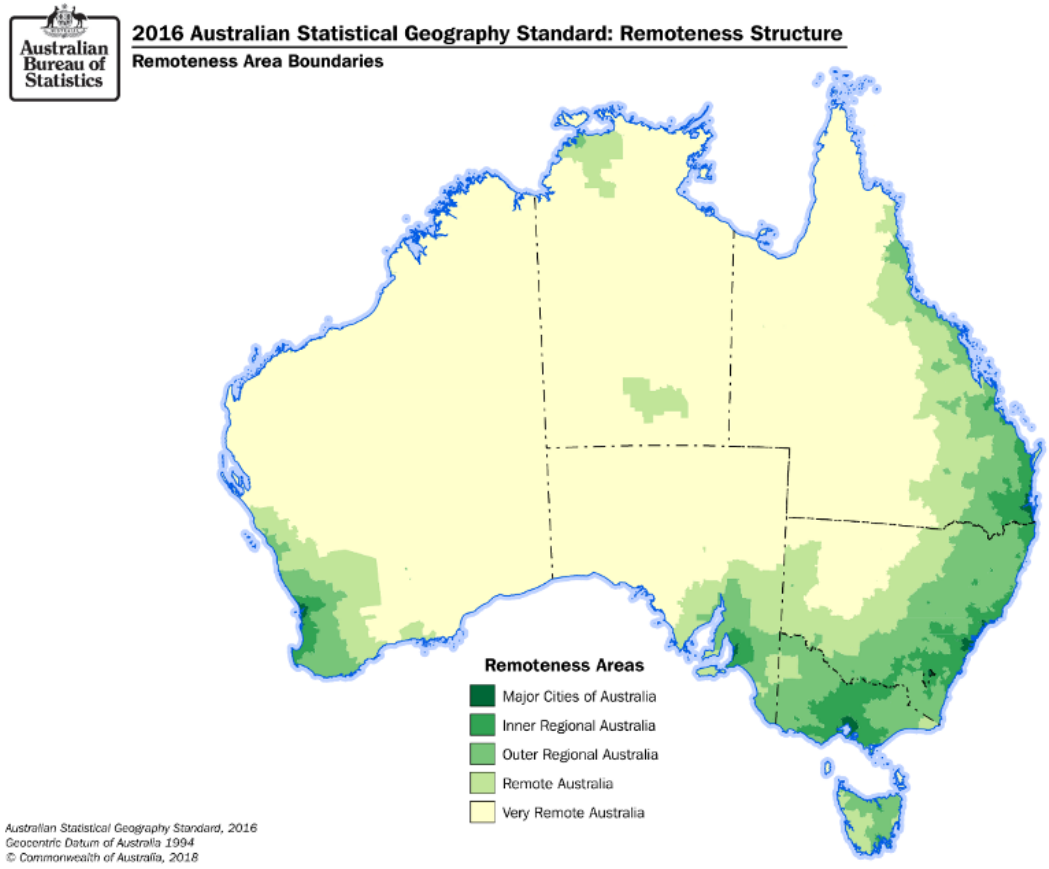
Despite the utility and cost-effectiveness of DSMES programs such as DESMOND, face-to-face programs may not be accessible or suitable for all people with diabetes. In addition, particularly during COVID-19, web-based delivery as an option has been very attractive [16,17]. Common barriers to participation include logistical (eg, time and lack of transport or parking), medical (eg, a person has another health condition that affects their ability to attend), financial (eg, costs associated with getting to a venue), emotional (eg, negative feelings with regard to groups), and cultural barriers (eg, beliefs and language) in addition to a lack of knowledge or no perceived benefit of attending such a program [15]. Moreover, people with diabetes cannot be assumed to be homogeneous in terms of their engagement preferences when seeking participation in a health service [18]. Individualized preferences should be taken into account in the design of support programs and services [19]. The provision of alternative modes of participation is critical when the goal is to maximize participation rates and improve diabetes care.

In Australia in particular, access to programs can be affected by geographic remoteness, which affects the availability of health resources, accessibility, and the financial viability of facilitating programs and services in remote areas. Australia's population is dispersed over almost 7.7 million km^2 [20]. To demonstrate the potential impact of geographic remoteness on face-to-face service delivery, a map of Australia by geographic remoteness is presented in Figure 1. Around one-third of the Australian population lives outside major cities [21], with age-standardized ratios of diabetes often higher in regional and remote areas than in major cities [22].

Web-based delivery of DSMES offers an opportunity to provide comprehensive coverage of education and support to otherwise difficult to reach populations. Such programs can mitigate many of the barriers associated with attendance to face-to-face programs, particularly in the Australian context where the population is dispersed over a vast geographical area.

Participation in DSMES via web-based modes of delivery has increased significantly in recent years [23]. Such programs are effective in improving knowledge and glycemic control when compared with usual care [24] and have been associated with improved well-being outcomes in people with T2DM [25]. Hadjiconstantinou et al [26] recently adapted the face-to-face version of DESMOND to a digital platform in accordance with the intervention-mapping approach [27]. The adoption and implementation of this digital program, named MyDESMOND, in the UK setting are currently under evaluation. Early findings suggest that the program is effective in reducing diabetes distress and improving self-efficacy in diabetes self-management [17]. MyDESMOND is an interactive digital program to support people living with T2DM. The digital platform has functionalities including educational material; booster sessions; peer support chat forum; *ask the expert*; integration with commercially available activity trackers; a decision maker (ie, personalized goal setting); and tracking of HbA_{1c}, blood pressure, cholesterol, weight, and waist measurements.

Figure 1. Map of Australia by geographic remoteness.



Objective

The aim of this paper is to describe the systematic process undertaken to adapt MyDESMOND to the Australian setting in readiness for consumer pilot-testing. This systematic approach considered the relevant social, cultural, environmental, and policy-level factors of influence to develop the first web-based DSMES program for people with T2DM in Australia. The description of this work provides a road map for others to follow when adapting comprehensive digital self-management programs to local contexts to ensure that such programs meet the needs and priorities of all stakeholders (eg, consumers, service providers, health systems, and governments) and are fit for purpose for implementation in a new setting.

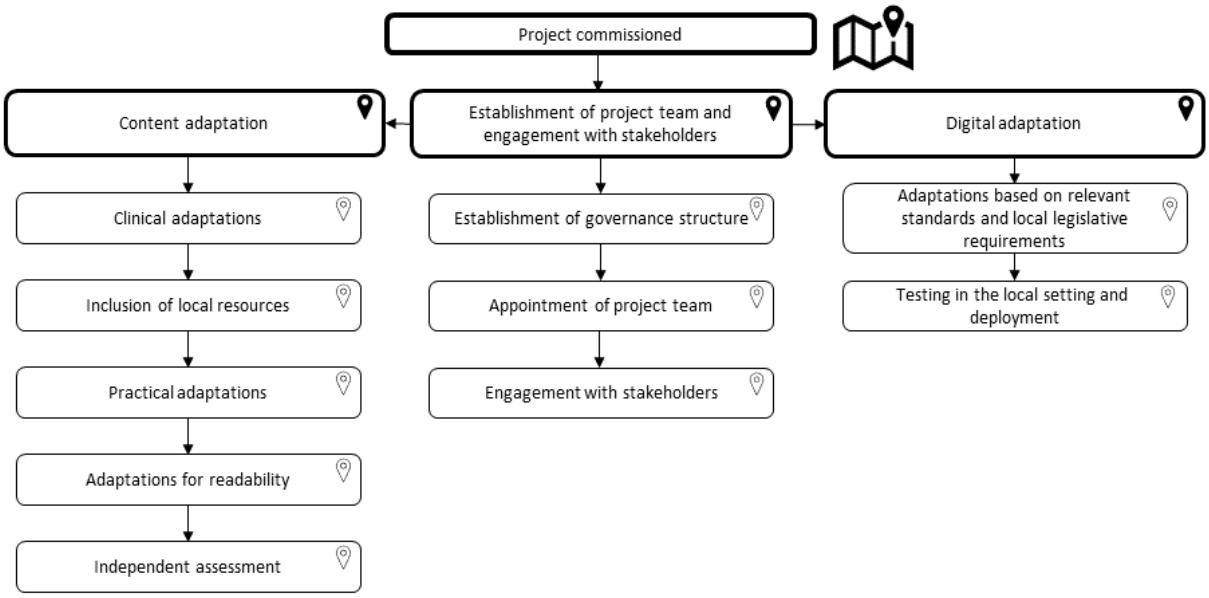
Methods

Overview

In December 2018, following a *request-for-quote* process, the Australian Government Department of Health (hereon referred to as *the Department*) commissioned Diabetes WA, National Diabetes Services Scheme (NDSS) Agent for WA, to adapt, pilot, evaluate, and implement MyDESMOND in the Australian setting. The process undertaken to adapt the program to the Australian context is the focus of this paper.

The adaptation of MyDESMOND to the Australian context involved three key areas of focus: (1) establishment of the project team and carefully planned engagement with stakeholders, (2) content adaptation, and (3) digital adaptation. These areas of focus are summarized in the project road map presented in Figure 2 and described in further detail below.

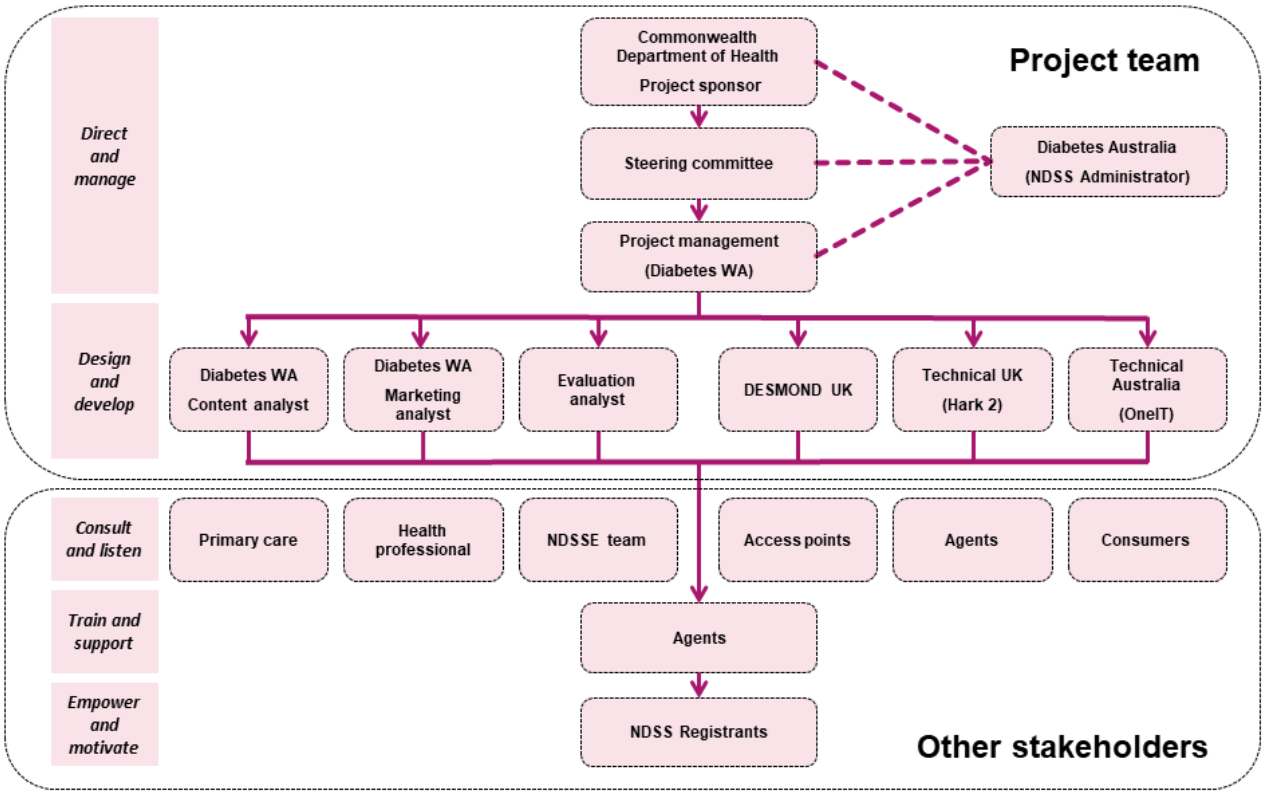
Figure 2. Project roadmap for the adaptation of MyDESMOND.



Establishment of Project Team and Engagement With Stakeholders

The first stage of the project to adapt MyDESMOND involved the establishment of a governance structure, appointment of the project team, and planning for stakeholder engagement. An organizational chart portraying the roles and responsibilities of the governing bodies, project management team, and key stakeholders is presented in Figure 3 and described in further detail below.

Figure 3. MYDESMOND governance, project management, and stakeholders.



Establishment of a Governance Structure

A steering committee of senior stakeholders was established to provide overall governance of the project, including the Chief Executives from Diabetes WA, Diabetes Australia, and Diabetes Tasmania; the General Manager of the NDSS; and the Directors of Diabetes Products, NDSS Enhancements, and Chronic Disease Policy from the Department. The role of the steering committee was to make recommendations in relation to project plans, budgets, and timelines and to broadly monitor progress and risks throughout the life of the project. This included recommendations as to how MyDESMOND (1) integrated with

and complemented without duplicating existing NDSS services delivered by Diabetes Australia and Agents of the NDSS across Australia; (2) complied with relevant government policy, the overarching NDSS service agreement, and individual NDSS Agent agreements; (3) met relevant standards for NDSS content, review, and approvals; (4) complied with local data privacy, cyber security, and other internet technology security requirements of the NDSS and government policy; (5) managed risks, defined as any event or condition with the potential to affect achievement of project objectives; and (6) could be delivered in accordance with agreed budgets and time frames.

Appointment of the Project Team

Diabetes WA was responsible for the day-to-day operation of the project. A project management team was appointed. The team, led by an experienced project manager, included Credentialled (accredited) Diabetes Educators (CDEs); content, marketing, and evaluation analysts; and a digital systems consultant. The team met weekly and was responsible for all aspects of developing, managing, and monitoring the project work plan and for securing appropriate resources to ensure the successful completion of the project.

Engagement With Stakeholders

A working group comprising representatives of the state and territory Agents of the NDSS, Diabetes Australia, and the Department was also established. The purpose of the group was to foster communication, collaboration, and input into planning for the adaptation, eventual pilot-testing, and implementation of MyDESMOND at the operational level. The group provided advice and support to the project management team to ensure that the program complemented and integrated with existing education and support services and materials provided through the NDSS, met appropriate standards for content, and included appropriate processes for approval and review.

A plan to guide stakeholder engagement and communication was developed and approved by the Agent working group and steering committee to ensure optimal stakeholder awareness, engagement, motivation, and satisfaction. The development of the plan involved four key processes: (1) identification of internal and external stakeholders likely to be affected by project processes or outcomes; (2) analysis of the level of interest, influence, and involvement in the project; (3) consideration of how to best manage stakeholder expectations during the project to maximize support and minimize the potential for conflict; and (4) planning to regularly update and review stakeholder expectations and building strategies to maintain stakeholder engagement and support throughout the life of the project. The plan included a matrix identifying key project stakeholders in addition to the type of information to be communicated, appropriate channels for communication, and the planned frequency and timing of communication activities. The matrix can be viewed in Multimedia Appendix 1.

Content Adaptation

Overview

The content of MyDESMOND was adapted to ensure that the program was (1) culturally and contextually appropriate for the Australian setting, (2) consistent with the content of the face-to-face version of DESMOND already being delivered in Australia, and (3) aligned with Australian clinical guidelines for the management of T2DM. The content was adapted by a CDE who was also accredited to facilitate the face-to-face adaptation of DESMOND in Australia and by an experienced content analyst. The process initially involved a detailed review of the UK version of MyDESMOND to identify program elements requiring adaptation. The program content was also compared with the Australian *DESMOND Newly Diagnosed and Foundation: Educator Manual and Curriculum 2015* [28] previously adapted for face-to-face delivery of DESMOND in the Australian context. As a result of this work, adaptations were made to program text, quizzes, videos, fact sheets, and the decision maker and health tracker (ie, biomedical) parameters. This process resulted in (1) clinical adaptations, (2) inclusion of local resources, (3) practical adaptations, and (4) adaptations for readability, followed by (5) a 2-step process of independent assessment.

Clinical Adaptation

The clinical guidelines for the management of T2DM cited throughout MyDESMOND were amended in accordance with Australian recommendations and guidelines. Targeted HbA_{1c} levels were amended from 6.5% to 7% (48-53 mmol/mol) [29]. Blood glucose targets were amended from UK standards (ie, 4-7 before meals and <8.5 two hours after) to 6 to 8 mmol/L fasting and 8 to 10 mmol/L after meals [29]. References to medications were updated to reflect those available in Australia [29,30]. Smoking was included in the list of known risk factors for the development of T2DM [29]. Content describing nutritional information was also adapted. For example, information about the fat content of packaged food included on food labels was amended to reflect local standards [31]. Australian guidelines for physical activity and sedentary behavior were referenced [32] in addition to Exercise and Sport Science Australia's position statement on exercise prescription for patients with T2DM and prediabetes [33].

Inclusion of Local Resources

Links and references to UK-focused consumer resources were replaced with links and references to Australian resources. For example, references to services and resources provided through the National Health Service (NHS) in the United Kingdom were replaced with references and links to the services and resources provided through the NDSS in Australia. These included references to the NDSS website landing page and information about blood glucose monitoring; management of hypoglycemia; sick days; medications; food labels; physical activity; sexual health and diabetes; the diabetes annual cycle of care; and advice on driving, travel, healthy cooking, and eating out [34]. Other local resources included information from the Heart Foundation [35], Nutrition Australia [36], and the Australian Government dietary guidelines [31] and physical activity guidelines [32].

Practical Adaptation

A variety of additional practical adaptations were also made to MyDESMOND. These included the conversion of all units of measurement to be consistent with Australian standards. For

example, calories were changed to kilojoules, and ounces were changed to grams. Spelling and grammar were also edited to reflect locally accepted conventions (eg, *programme* was changed to *program*). Branding throughout the program was amended from that of the NHS and the Leicester Diabetes Centre to that of the Australian Government Department of Health and the NDSS. References accrediting the Leicester Diabetes Centre for the original program design were retained. Embedded videos were refilmed to reflect the changes to program content and local language accents and included the stories of Australian people living with T2DM. Images were adapted to suit the Australian context. To ensure consistent formatting throughout the program, the images were edited by the Leicester Diabetes Centre. Processes for registration were also updated to be contextually relevant. This included data to be collected in relation to the participants' postcode information (ie, Australian postcodes), gender (ie, male, female, and X—*other*), ethnicity (ie, including response options for Aboriginal and Torres Strait Islanders), and registrant identification (ie, NDSS registration number rather than NHS number).

Ensuring Readability

Given Australia's multicultural diversity, including migrants of non–English-speaking background, it was important to ensure that MyDESMOND was appropriate for people with lower levels of English literacy. Each learning session and booster session throughout the program was reviewed against the NDSS readability checklist [37]. The purpose of the checklist is to ensure that the content of NDSS materials is suitable for people living with diabetes with low levels of English proficiency and health literacy. The checklist includes items describing appropriate use of language (eg, *Choose words that are familiar and culturally appropriate for your readers*), sentence structure (eg, *Use one idea per sentence*), type of information to be included (eg, *Tell your readers how to get more help or information*), design features (eg, *Choose a colour scheme that is not distracting*), and pictures (eg, *use pictures, logos or photographs to add meaning to the text*).

The readability of MyDESMOND was also assessed using the Flesch–Kincaid Scale [38], a scale that has commonly been applied to assess the readability of health information materials [39]. The score was calculated using the proofing feature in Word for Microsoft 365 [40]. The targeted reading level was equivalent to grade 7-9 (approximate age 13-15 years) as specified in the NDSS readability standards [37]. Specific content assessed at a level higher than this was revised, ensuring that the language, length, and complexity of the sentences were appropriate for the targeted level of readability. Furthermore, to enhance the accessibility of MyDESMOND, we presented content in a variety of formats, including text, videos, video transcripts, and simple pictures.

Independent Assessment

Upon completion of the amendments to the program content, MyDESMOND was subjected to a 2-stage process of independent assessment to confirm that the content was appropriate for the Australian context and complied with all relevant standards and guidelines. First, a review was conducted by 2 external CDEs (ie, these CDEs were not affiliated with Diabetes WA or otherwise involved in the MyDESMOND project). The program was then sent to the Medical Education and Scientific Advisory Council for review. The Medical Education and Scientific Advisory Council is a body maintained by the Australian Diabetes Educators Association and the Australian Diabetes Society and provides strategic advice to the NDSS on medical, educational, and scientific matters. The results of these 2 stages of review confirmed that the content of MyDESMOND was consistent with relevant standards and guidelines and, therefore, likely to be suitable for Australian consumers.

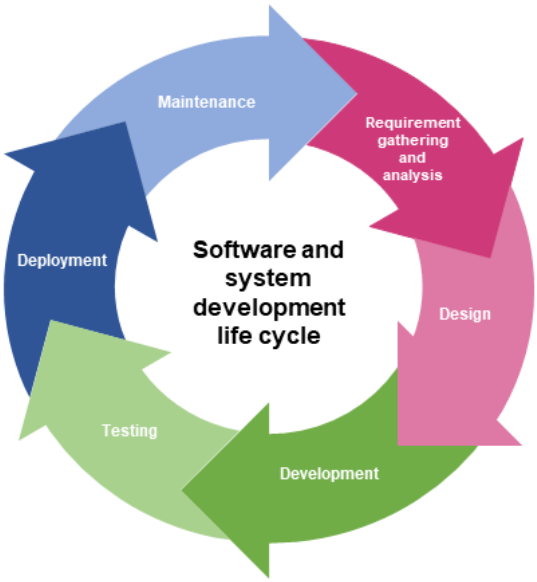
Digital Adaptation

Overview

The digital specifications of MyDESMOND were modified to suit the Australian setting. This part of the project was overseen by a member of the project management team who was an experienced project systems consultant. Coding and development work outside of the content management system was undertaken by the UK-based developers of the original MyDESMOND platform. The modifications were guided by a software and system development life cycle approach. Such approaches offer a stage-based conceptualization of the life of a system or software from inception to maintenance [41]. The format of life cycle models and the terms applied to describe the stages of the life cycle vary between models depending on the context and intended application [41]. As shown in Figure 4, stages may include requirement gathering and analysis, design, development, testing, deployment, and maintenance.

The requirement gathering and analysis phase of a life cycle typically involves identification of stakeholder needs and building understanding of how users will interact with the system and how the system should function. Given that MyDESMOND had been developed and offered to people with T2DM in the United Kingdom, this phase of the project involved ensuring that the program met local standards and legislative requirements for information and data management. The design phase of the life cycle usually involves the determination of a high-level system design, whereas the development phase involves the configuration of the infrastructure as well as database and system coding. In the case of MyDESMOND, this work had been undertaken by a UK-based web developer when the program was originally developed. The next phase of the life cycle involved testing to verify that the system worked as expected under conditions simulating *real life*. Deployment occurs when system functionality is verified and involves the release of the software or system to end users. When a system is deployed and users start to interact with it, issues or defects may be identified, at which point the life cycle continues to evolve.

Figure 4. Software development life cycle.



Adaptations Based on Relevant Standards and Local Legislative Requirements

MyDESMOND was adapted to meet relevant standards and local legislative requirements for best-practice information and data management. Relevant standards and legislation informing local adaptations included (1) the National Institute of Standards and Technology [42] information system security plan template; (2) the Australian Government protective security policy framework [43], including policies related to sensitive and classified information, access to information, safeguarding information from cyber threats, and robust information and communication technology systems; (3) the Australian *Privacy Act 1988* [44] and Privacy Principles [45], including standards, rights, and obligations around the collection, use, and disclosure of personal information, the organization's governance and accountability, the integrity and collection of personal information, and the rights of individuals to access their personal information; and (4) infrastructure and software security controls aligned with the Australian Government information security manual [43] for the requisite information classification.

The review of relevant standards and local legislation resulted in changes to the program privacy policy and terms and conditions and the addition of a feature to enable users to opt in or out to receive future direct marketing campaigns. More specifically, the amendments included (1) strengthening passphrase (password) management practices for user accounts, including the adoption of protocols for the frequency of required changes to passphrases and limitations on the frequency of user-instigated daily password changes, the reuse of passphrases, and the use of sequential passphrases; (2) implementation of multifactor authentication; (3) recording of relevant information for logged events, including a description of the event, the date and time it occurred, and the users, processes, and equipment involved; and (4) implementation of the Secure Hash Algorithm 2 family of Secure Hash Algorithms [46] for enhanced data security.

To ensure data sovereignty, whereby MyDESMOND data were stored and accessible only from within Australia [47], the application was installed on Australian servers in accordance with the Secure Shell Protocol [48]. The Secure Shell Protocol ensures secure remote logging and network services over insecure networks through processes of authentication and encryption.

Testing

The Australian MyDESMOND platform was tested to validate that the program functioned as specified and, thus, could be considered *production-ready*. A test plan outlining the strategies for testing and execution and the processes for test management was developed. The plan was reviewed by key stakeholders before its execution to foster commitment to resourcing and time frames.

Testing focused on the functionality of the platform (ie, the application functioned as intended), capacity for backup and recovery (ie, verifying that the system had appropriate methods for data replication and that data could be effectively restored if required), and penetration (ie, identification of vulnerabilities in the infrastructure and software that might be exploited to compromise the confidentiality, integrity, or availability of the system). Additional nonfunctional testing typically conducted at this stage of the development life cycle (eg, availability, reliability, and performance) was not required as the MyDESMOND Australia installation had been configured to replicate the UK environment, which had been operating nationally in the United Kingdom since 2018 with >15,000 users (well in excess of anticipated user numbers for the Australian pilot).

The project team developed test cases for each function and feature of MyDESMOND to identify inputs and expected outputs, allow for remediation of issues where the actual output varied from anticipated outputs, and identify opportunities to improve user experience. The cases were manually executed by professional test engineers using a range of common web browsers (eg, Google Chrome, Safari, Internet Explorer, and

Firefox) and devices (eg, desktop and laptop computers and iOS and Android phones and tablets). Project team members participated in testing (eg, project managers and coordinators, CDEs, marketing coordinators, and evaluation coordinators). Testing was repeated when system shortfalls were identified and rectified. In total, 736 test cases were executed on web browsers, and 566 were executed on devices. More than 100 issues were reported and categorized according to severity (eg, low, medium, and high). No issues were categorized as *high*. Examples of issues included duplication of words, poor rendering of display, inaccurate or unhelpful error messages, incorrect navigation, inappropriate or missing verification and validation, and problems connecting with third-party services. Upon completion of testing, MyDESMOND was profiled as *low-risk* and determined as functionally fit for purpose in the Australian setting. The disaster recovery plan was successfully tested, with the system fully recovered and restored. Specialist internet technology security consultants conducted penetration testing to simulate a malicious user both with and without credentials. This test demonstrated that MyDESMOND was not susceptible to cyberattacks and would not expose users to unacceptable risk. Upon completion of functional, disaster recovery, and penetration testing, MyDESMOND was recommended for deployment to the production environment in preparation for consumer pilot-testing.

Discussion

Principal Findings

The steps outlined in this paper facilitated the cultural adaptation of MyDESMOND, the first evidence-based web-based DSMES program for Australian people living with T2DM. By documenting the systematic approach, including the appointment of a multidisciplinary project management team, active stakeholder engagement, and the specific processes undertaken to adapt program content and the digital platform, this paper provides a road map for others to follow when adapting digital health programs for delivery in new contexts.

Adaptation of existing evidence-based health programs to new settings has the potential to introduce services into new environments in a cost-effective and timely manner with reasonable prospects of successful outcomes [49]. However, health interventions interact with features of the context in which they are embedded; it cannot be assumed that an intervention that is effective in one setting will be effective in another setting [50]. Adaptations to suit the characteristics of the new setting can enhance the likelihood that the intervention will achieve similar outcomes to those realized in the original setting [50]. The adaptation of MyDESMOND optimized the program for delivery in the Australian setting. The approach undertaken—including the selection of an existing suitable and effective program; review of original program materials; identification of localized objectives, core components, and mismatches between the original program and new context; and adaptation of original materials—was consistent with established frameworks for the adaptation of traditional health programs for use in new contexts [49].

The incorporation of stakeholder input into the design of behavioral interventions to improve public health is critical, as is the development of strategies for optimal and sustained stakeholder engagement [51]. The engagement of a diverse group of stakeholders in the contextual adaptation of MyDESMOND facilitated input and *buy-in* from funders, service providers, health professionals, and others who stood to be affected by the implementation of the program in the short and long term. Inclusion of these parties from project instigation to piloting, evaluation, and implementation improves the likelihood of ongoing *buy in* and support from those who have influence over program delivery, thereby enhancing the potential sustainability of the program in the future.

Strengths and Limitations

A strength of this work is the adoption of an existing evidence-based program grounded in behavioral theory and designed using a systematic intervention-mapping approach [26]. It is recommended that health intervention design be informed by theoretical frameworks [27], and evidence shows that theory-based interventions are effective in improving HbA_{1c} levels, self-efficacy, diabetes knowledge, and self-care behaviors in people with T2DM [52]. Thus, when choosing an appropriate digital DSMES program to support Australians with T2DM, MyDESMOND was an ideal candidate. MyDESMOND Australia is evidence-based, underpinned by behavioral theory, and optimized for delivery in the Australian setting, thereby increasing the likelihood that it will be effective in improving outcomes in Australians with T2DM.

Robust security and privacy protections for the rights of consumers are key to ensure that digital health interventions are effective, safe, and trusted by users; a proactive approach that considers potential and emerging threats, relevant standards and legislative requirements, and evolving ethical issues is essential [53]. Accordingly, the identification of relevant standards and laws for digital information, combined with extensive localized testing of the MyDESMOND platform, is a strength of this work. Moreover, the appropriateness of the adaptations to program content and the digital platform included processes for external review and system testing, respectively. The incorporation of these processes provided checks and balances to confirm that MyDESMOND was likely to be contextually appropriate and suitable for use among Australians with T2DM.

We followed a systematic approach to adapt MyDESMOND to the Australian setting. However, there are limitations that need to be acknowledged. When working with project stakeholders, it is critical to collaboratively establish and agree on detailed timelines, systematic work plans, and goal-setting processes with continual revision of standards set and group norms [27]. Despite a priori planning of strategies for optimal stakeholder engagement, the high level of project governance by senior stakeholders from a range of external organizations with competing agendas led to delays in the establishment of the initial MyDESMOND project plan. In addition, although the 2-stage review process of program content provided some assurance of the cultural appropriateness of the adaptations, the time taken by external assessors to review and provide feedback

was underestimated, leading to further project delays. However, these delays did not affect the quality or overall timeline of the project.

Delays were also experienced in adapting the MyDESMOND digital platform owing to a lack of system administrator training or hand over and lack of architectural diagrams and administrator guides or description of access levels. This resulted in a process of trial and error as the local project team learned how to navigate the system. To mitigate this unforeseen issue, we engaged a content editor experienced in technology and digital communications to ensure that our team developed and followed robust system user guides. Involving the UK developers of the original MyDESMOND platform in the project from the outset also provided an additional source of knowledge and expertise to overcome this issue. Consequently, we recommend involving original program developers as stakeholders when adapting digital programs to new settings, where possible.

Although the adaptation of MyDESMOND Australia involved a collaborative approach and extensive stakeholder engagement, program users (ie, NDSS registrants) were not directly involved in the adaptation at this stage. However, follow-up pilot-testing involves an initial exploratory phase to seek preliminary feedback about program acceptability from users before advancing to more comprehensive phases of pilot-testing.

Key Recommendations When Adapting Digital Programs

We provide a list of recommendations to consider for future adaptation work in digital self-management programs (Textbox 1). These recommendations can be applied to other long-term health conditions and cultural settings.

Textbox 1. Key recommendations to consider for the adaptation of future digital self-management programs.

Key recommendations for the adaptation of digital self-management programs

- Involve individuals with multidisciplinary skills and expertise in the project management team. This may include those experienced in project management, program or clinical content, and digital systems and processes in addition to the original program developers.
- Involve stakeholders across multiple levels of influence. This might include funding or commissioning bodies, senior executives and operational-level employees of service-providing organizations, health professionals and clinicians, original program developers, and others who may have an interest in or be affected by program implementation.
- When working with independent advisors and bodies, it is critical to establish workable time frames and deadlines and seek agreement to those time frames by all concerned.
- Create a clear plan for how stakeholder engagement will be sustained throughout the life of the project. Consider the priorities and agendas of each of the stakeholders.
- Use information and resources from a range of credible local sources to inform content adaptations. These may include government guidelines and recommendations as well as information and resources from peak industry bodies, advisory councils, and professional societies.
- When adapting an existing digital program to a new setting, it is important to ensure that the relevant team members have been provided with adequate system administration training and that appropriate technical user guides and resources are supplied.
- Incorporate a process for external or independent review of program content to ensure that it is culturally appropriate and relevant to the target population.
- Adopt a systematic approach to digital adaptations to ensure optimal performance in the local environment. Consider all relevant standards, local legislative requirements, and matters surrounding data sovereignty.
- Conduct rigorous testing of the digital program to ensure that it performs as intended in the local setting.

Future Directions

Further work is needed to ensure the feasibility of delivering MyDESMOND in the Australian setting in addition to establishing the acceptability and potential effectiveness of the program among Australian people with T2DM. Consumer pilot-testing was undertaken upon completion of the adaptation project described in this paper and will be reported in detail at a later date. Briefly, the aims of the 2-phase pilot were to (1) determine if MyDESMOND provided a useful, engaging, and relevant learning experience for people in Australia living with T2DM; (2) obtain consumer feedback on the functional and technical aspects of the program; (3) determine the potential effectiveness of MyDESMOND in increasing diabetes empowerment and reducing diabetes-related distress among Australians with T2DM; and (4) further assess consumer satisfaction with MyDESMOND after further adaptations were made based on the findings of phase 1 of pilot-testing. Should consumer pilot-testing indicate that MyDESMOND is acceptable, feasible, and likely to be effective, the program will be recommended for inclusion in the suite of programs and services provided to people with T2DM throughout Australia via the NDSS.

Conclusions

This paper provides a road map for the adaptation of digital health programs for delivery in new contexts, including recommendations for engaging the project management team and stakeholders and processes for adapting content and the digital platform. We believe that the systematic approach that we adopted to adapt MyDESMOND to the Australian setting would be applicable to other development studies that aim to contextually adapt an existing digital program.

Acknowledgments

The original web-based version of the Diabetes Education and Self-Management for Ongoing and Newly Diagnosed program was developed by a multidisciplinary team at the Leicester Diabetes Centre, a leading applied health research unit committed to improving the lives and care of people with diabetes and other long-term conditions. The Leicester Diabetes Centre is a collaboration between the University of Leicester and the University Hospitals of Leicester National Health Service Trust. The authors would like to thank the multidisciplinary team at the Leicester Diabetes Centre for their ongoing support in this project. They also thank the funders of the web-based Diabetes Education and Self-Management for Ongoing and Newly Diagnosed Australia project and the Australian Government Department of Health. Stakeholder engagement was a strength of the work outlined in this paper. The authors would like to thank all stakeholders, including members of the Australian Government Department of Health, National Diabetes Services Scheme, steering committee, project management team, Agent working group, and Medical Education and Scientific Advisory Council, as well as the contributing employees from Diabetes Western Australia and the health professionals involved in the independent review processes. KK, MH, and MJD are supported by the National Institute for Health Research Applied Research Collaboration East Midlands and the National Institute for Health Research Leicester Biomedical Research Centre.

Authors' Contributions

MH led on the underpinning theories and behavior change included in the original web-based version of the Diabetes Education and Self-Management for Ongoing and Newly Diagnosed (MyDESMOND) program. MJD and KK were involved in the development, testing, and evaluation of the Diabetes Education and Self-Management for Ongoing and Newly Diagnosed program and supported the work of MH and The Leicester Diabetes Center in the adaptation and development of MyDESMOND. SC was the project manager of the MyDESMOND project in Australia, CL led the local adaptation of the program content, KW led the local adaptation of the digital platform, and DS was a member of the Agent working group. JO led the development of the manuscript. MH, SC, CL, KW, VM, and NW revised the content and structure of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

Diabetes Western Australia (WA) was funded by the Australian Government Department of Health to adapt, pilot, evaluate, and implement the web-based version of the Diabetes Education and Self-Management for Ongoing and Newly Diagnosed program in Australia. The Australian Government Department of Health reviewed a draft of this paper before its submission to the journal; however, the authors retained editorial control over the content of the manuscript. Diabetes WA receives funding from the National Diabetes Services Scheme to deliver diabetes self-management education and support, including the Diabetes Education and Self-Management for Ongoing and Newly Diagnosed program, to people with diabetes in WA and to evaluate education and support programs facilitated nationally through the National Diabetes Services Scheme. The Leicester Diabetes Centre and the Diabetes Research Centre at the University of Leicester receive licensing fees from Diabetes WA to support the implementation of Diabetes Education and Self-Management for Ongoing and Newly Diagnosed programs throughout Australia. DS and CL are employees of Diabetes WA. JO, SC, NW, KW, and VM were employees at Diabetes WA when the work was undertaken. MH is employed by the Diabetes Research Centre, University of Leicester, Leicester, United Kingdom.

Multimedia Appendix 1
Stakeholder Engagement Matrix.
[DOCX File , 19 KB - jmir_v24i4e26339_app1.docx]

References

1. Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WHO/IDF consultation. World Health Organization. 2006. URL: https://apps.who.int/iris/handle/10665/43588 [accessed 2021-01-03]
2. Zhou B, Lu Y, Hajifathalian K, Bentham J, Di Cesare M, Danaei G, et al. Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. The Lancet 2016;387(10027):1513-1530. [doi: 10.1016/S0140-6736(16)00618-8]
3. Sainsbury E, Shi Y, Flack J, Coagiuri S. Burden of diabetes in Australia: it's time for more action. The University of Sydney, Sydney. 2018 Jul. URL: https://www.sydney.edu.au/content/dam/corporate/documents/faculty-of-medicine-and-health/research/centres-institutes-groups/burden-of-diabetes-its-time-for-more-action-report.pdf [accessed 2021-02-01]
4. Diabetes. World Health Organization. 2020. URL: https://www.who.int/health-topics/diabetes#tab=tab_1 [accessed 2021-01-02]
5. Diabetes: web report. Australian Institute of Health and Welfare. 2020. URL: https://www.aihw.gov.au/reports/diabetes/diabetes-snapshot/contents/how-many-australians-have-diabetes/type-2-diabetes [accessed 2021-01-03]
6. Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat Rev Endocrinol 2018 Feb 8;14(2):88-98. [doi: 10.1038/nrendo.2017.151] [Medline: 29219149]

7. American Association of Diabetes Educators. An effective model of diabetes care and education: revising the AADE7 self-care behaviors. Diabetes Educ 2020 Apr;46(2):139-160. [doi: 10.1177/0145721719894903] [Medline: 31928334]
8. Hermanns N, Ehrmann D, Finke-Groene K, Kulzer B. Trends in diabetes self-management education: where are we coming from and where are we going? A narrative review. Diabet Med 2020 Mar 04;37(3):436-447. [doi: 10.1111/dme.14256] [Medline: 32017188]
9. Chrvala CA, Sherr D, Lipman RD. Diabetes self-management education for adults with type 2 diabetes mellitus: a systematic review of the effect on glycemic control. Patient Educ Couns 2016 Jun;99(6):926-943 [FREE Full text] [doi: 10.1016/j.pec.2015.11.003] [Medline: 26658704]
10. He X, Li J, Wang B, Yao Q, Li L, Song R, et al. Diabetes self-management education reduces risk of all-cause mortality in type 2 diabetes patients: a systematic review and meta-analysis. Endocrine 2017 Mar;55(3):712-731. [doi: 10.1007/s12020-016-1168-2] [Medline: 27837440]
11. Siegel K, Ali M, Zhou X, Ng B, Jawanda S, Proia K, et al. Cost-effectiveness of interventions to manage diabetes: has the evidence changed since 2008? Diabetes Care 2020 Jul;43(7):1557-1592. [doi: 10.2337/dci20-0017] [Medline: 33534729]
12. Skinner TC, Carey ME, Cradock S, Daly H, Davies MJ, Doherty Y, DESMOND Collaborative. Diabetes Education and Self-Management for Ongoing and Newly Diagnosed (DESMOND): process modelling of pilot study. Patient Educ Couns 2006 Dec;64(1-3):369-377. [doi: 10.1016/j.pec.2006.04.007] [Medline: 17011154]
13. Gillett M, Dallosso HM, Dixon S, Brennan A, Carey ME, Campbell MJ, et al. Delivering the diabetes education and self management for ongoing and newly diagnosed (DESMOND) programme for people with newly diagnosed type 2 diabetes: cost effectiveness analysis. Br Med J 2010 Aug 20;341:c4093 [FREE Full text] [doi: 10.1136/bmj.c4093] [Medline: 20729270]
14. Miller VM, Davies MJ, Etherton-Beer C, McGough S, Schofield D, Jensen JF, et al. Increasing patient activation through diabetes self-management education: outcomes of DESMOND in regional Western Australia. Patient Educ Couns 2020 Apr;103(4):848-853. [doi: 10.1016/j.pec.2019.10.013] [Medline: 31676100]
15. Olson J, White B, Mitchell H, Halliday J, Skinner T, Schofield D, et al. The design of an evaluation framework for diabetes self-management education and support programs delivered nationally. BMC Health Serv Res 2022 Jan 09;22(1):46 [FREE Full text] [doi: 10.1186/s12913-021-07374-4] [Medline: 35000599]
16. Quinn LM, Davies MJ, Northern A, Brough C, Schreder S, Stribling B, et al. Use of MyDesmond digital education programme to support self-management in people with type 2 diabetes during the COVID-19 pandemic. Diabet Med 2021 Mar 23;38(3):e14469 [FREE Full text] [doi: 10.1111/dme.14469] [Medline: 33274489]
17. Hadjiconstantinou M, Barker MM, Brough C, Schreder S, Northern A, Stribling B, et al. Improved diabetes-related distress and self-efficacy outcomes in a self-management digital programme for people with type 2 diabetes, myDESMOND. Diabet Med 2021 May 15;38(5):e14551. [doi: 10.1111/dme.14551] [Medline: 33675068]
18. Sarkar U, Piette J, Gonzales R, Lessler D, Chew L, Reilly B, et al. Preferences for self-management support: findings from a survey of diabetes patients in safety-net health systems. Patient Educ Couns 2008 Jan;70(1):102-110 [FREE Full text] [doi: 10.1016/j.pec.2007.09.008] [Medline: 17997264]
19. Hertroijs D, Elissen A, Brouwers M, Hiligsmann M, Schaper N, Ruwaard D. Preferences of people with Type 2 diabetes for diabetes care: a discrete choice experiment. Diabet Med 2020 Nov;37(11):1807-1815 [FREE Full text] [doi: 10.1111/dme.13969] [Medline: 31001855]
20. Area of Australia - States and Territories. Geoscience Australia. 2021. URL: https://www.ga.gov.au/scientific-topics/national-location-information/dimensions/area-of-australia-states-and-territories [accessed 2021-05-06]
21. Rural and remote health. Australian Institute of Health and Welfare. 2019. URL: https://www.aihw.gov.au/reports/rural-remote-australians/rural-remote-health/contents/profile-of-rural-and-remote-australians [accessed 2021-06-06]
22. Remoteness in Australia - Graphs. PHIDU Torrens University Australia, Sydney. 2020. URL: https://phidu.torrens.edu.au/social-health-atlases/graphs [accessed 2021-02-03]
23. Greenwood DA, Gee PM, Fatkin KJ, Peeples M. A systematic review of reviews evaluating technology-enabled diabetes self-management education and support. J Diabetes Sci Technol 2017 May 31;11(5):1015-1027. [doi: 10.1177/1932296817713506] [Medline: 28560898]
24. Pereira K, Phillips B, Johnson C, Vorderstrasse A. Internet delivered diabetes self-management education: a review. Diabetes Technol Ther 2015 Jan;17(1):55-63. [doi: 10.1089/dia.2014.0155] [Medline: 25238257]
25. Hadjiconstantinou M, Byrne J, Bodicoat DH, Robertson N, Eborall H, Khunti K, et al. Do web-based interventions improve well-being in type 2 diabetes? A systematic review and meta-analysis. J Med Internet Res 2016 Oct 21;18(10):e270 [FREE Full text] [doi: 10.2196/jmir.5991] [Medline: 27769955]
26. Hadjiconstantinou M, Schreder S, Brough C, Northern A, Stribling B, Khunti K, et al. Using intervention mapping to develop a digital self-management program for people with type 2 diabetes: tutorial on MyDESMOND. J Med Internet Res 2020 May 11;22(5):e17316 [FREE Full text] [doi: 10.2196/17316] [Medline: 32391797]
27. Eldredge LK, Markham CM, Ruiter RA, Fernandez ME, Kok G, Parcel GS. Planning Health Promotion Programs: An Intervention Mapping Approach. San Francisco: John Wiley & Sons; 2016.
28. Newly diagnosed and foundation: educator manual and curriculum [training manual]. In: DESMOND Australia. Perth: Diabetes WA; 2015.

29. General practice management of type 2 diabetes. Royal Australian College of General Practitioners. 2016. URL: https://www.racgp.org.au/FSDEDEV/media/documents/Clinical%20Resources/Guidelines/Diabetes/General-practice-management-of-type-2-diabetes_1.pdf [accessed 2021-06-02]
30. Medications for type 2 diabetes fact sheet. National Diabetes Services Scheme. 2020. URL: https://www.ndss.com.au/about-diabetes/resources/find-a-resource/medications-for-type-2-diabetes-fact-sheet/ [accessed 2021-02-03]
31. Australian dietary guidelines. Department of Health, Australian Government. 2015. URL: https://www.eatforhealth.gov.au/ [accessed 2021-06-05]
32. Australia's physical activity and sedentary behaviour guidelines and the Australian 24-hour movement guidelines. Department of Health, Australian Government. 2019. URL: https://www1.health.gov.au/internet/main/publishing.nsf/Content/health-pubhlth-strateg-phys-act-guidelines [accessed 2021-05-18]
33. Hordern MD, Dunstan DW, Prins JB, Baker MK, Singh MA, Coombes JS. Exercise prescription for patients with type 2 diabetes and pre-diabetes: a position statement from Exercise and Sport Science Australia. J Sci Med Sport 2012 Jan;15(1):25-31. [doi: 10.1016/j.jsams.2011.04.005] [Medline: 21621458]
34. National Diabetes Services Scheme. Diabetes Australia. 2020. URL: https://www.ndss.com.au/ [accessed 2021-06-01]
35. Heart Foundation. National Heart Foundation of Australia. 2019. URL: https://www.heartfoundation.org.au/ [accessed 2021-07-15]
36. Nutrition Austalia. The Australian Nutrition Foundation. 2020. URL: https://nutritionaustralia.org/ [accessed 2021-04-12]
37. National Diabetes Services Scheme. Readability Checklist. Canberra: Diabetes Australia; 2015.
38. Kincaid J, Fishburne JR, Rogers R, Chissom B. Derivation of new readability formulas (automated readability index, Fog count and Flesch reading ease formula) for navy enlisted personnel. Naval Technical Training Command Millington TN Research Branch. 1975. URL: https://apps.dtic.mil/sti/citations/ADA006655 [accessed 2021-02-03]
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. Office support: get you document's readability and level statistics. Microsoft. 2020. URL: https://support.microsoft.com/en-us/office/get-your-document-s-readability-and-level-statistics-85b4969e-e80a-4777-8dd3-f7fc3c8b3fd2 [accessed 2021-06-06]
41. Ruparelia NB. Software development lifecycle models. SIGSOFT Softw Eng Notes 2010 May 11;35(3):8-13. [doi: 10.1145/1764810.1764814]
42. Swanson M, Hash J, Bowen P. Guide for developing security plans for federal information systems. National Institute of Standards and Technology. 2006. URL: https://csrc.nist.gov/publications/detail/sp/800-18/rev-1/final [accessed 2021-12-01]
43. Protective security policy framework. Attorney-General's Department, Australian Government. 2020. URL: https://www.protectivesecurity.gov.au [accessed 2021-04-25]
44. Privacy Act. Attorney-General's Department, Australian Government. 1988. URL: https://www.ag.gov.au/rights-and-protections/privacy [accessed 2021-05-30]
45. Australian privacy principles. Office of the Australian Information Commissioner. 2020. URL: https://www.oaic.gov.au/privacy/australian-privacy-principles [accessed 2021-06-01]
46. Handschuh H. SHA-0, SHA-1, SHA-2 (Secure Hash Algorithm). In: Encyclopedia of Cryptography and Security. Boston, MA: Springer US; 2011:1190-1193.
47. Irion K. Government cloud computing and national data sovereignty. Policy Internet 2013 Jan 23;4(3-4):40-71 [FREE Full text] [doi: 10.1002/poi3.10]
48. Ylonen T, Lonvick C. The Secure Shell (SSH) protocol architecture. Cisco Systems, Inc. 2006. URL: https://www.hjp.at/doc/rfc/rfc4251.html [accessed 2021-11-10]
49. Card JJ, Solomon J, Cunningham SD. How to adapt effective programs for use in new contexts. Health Promot Pract 2011 Jan;12(1):25-35. [doi: 10.1177/1524839909348592] [Medline: 19858321]
50. Movsisyan A, Arnold L, Evans R, Hallingberg B, Moore G, O'Cathain A, et al. Adapting evidence-informed complex population health interventions for new contexts: a systematic review of guidance. Implement Sci 2019 Dec 17;14(1):105 [FREE Full text] [doi: 10.1186/s13012-019-0956-5] [Medline: 31847920]
51. Byrne M, McSharry J, Meade O, Lavoie KL, Bacon SL. An international, Delphi consensus study to identify priorities for methodological research in behavioral trials in health research. Trials 2020 Mar 23;21(1):292 [FREE Full text] [doi: 10.1186/s13063-020-04235-z] [Medline: 32293510]
52. Zhao F, Suhonen R, Koskinen S, Leino-Kilpi H. Theory-based self-management educational interventions on patients with type 2 diabetes: a systematic review and meta-analysis of randomized controlled trials. J Adv Nurs 2017 Apr;73(4):812-833 [FREE Full text] [doi: 10.1111/jan.13163] [Medline: 27681948]
53. Arigo D, Jake-Schoffman DE, Wolin K, Beckjord E, Hekler EB, Pagoto SL. The history and future of digital health in the field of behavioral medicine. J Behav Med 2019 Feb;42(1):67-83. [doi: 10.1007/s10865-018-9966-z] [Medline: 30825090]

Abbreviations

CDE: Credentialled Diabetes Educator
DESMOND: Diabetes Education and Self-Management for Ongoing and Newly Diagnosed
DSMES: diabetes self-management education and support
HbA_{1c}: glycated hemoglobin A1c
MyDESMOND: web-based Diabetes Education and Self-Management for Ongoing and Newly Diagnosed
NDSS: National Diabetes Services Scheme
NHS: National Health Service
T2DM: type 2 diabetes mellitus
WA: Western Australia

Edited by G Eysenbach; submitted 08.12.20; peer-reviewed by A Nguyen, S Baptista; comments to author 11.03.21; revised version received 24.03.21; accepted 27.12.21; published 20.04.22.

Please cite as:
Olson J, Hadjiconstantinou M, Luff C, Watts K, Watson N, Miller V, Schofield D, Khunti K, Davies MJ, Calginari S
From the United Kingdom to Australia—Adapting a Web-Based Self-management Education Program to Support the Management of Type 2 Diabetes: Tutorial
J Med Internet Res 2022;24(4):e26339
URL: https://www.jmir.org/2022/4/e26339
doi:10.2196/26339
PMID:35442198

Original Paper

Effectiveness of an Internet-Based and Telephone-Assisted Training for Parents of 4-Year-Old Children With Disruptive Behavior: Implementation Research

Andre Sourander[1,2,3*], MD, PhD; Terja Ristkari[1,2*], RN, MSN; Marjo Kurki[1,2,4], PhD; Sonja Gilbert[1,2], PhD; Susanna Hinkka-Yli-Salomäki[1,2], LicPh; Malin Kinnunen[1,2], RN; Laura Pulkki-Råback[1,2,5], PhD; Patrick J McGrath[6,7,8,9], OC, FRSC, FCAHS, PhD

[1]Department of Child Psychiatry, University of Turku, Turku, Finland
[2]Finland INVEST Research Flagship, University of Turku, Turku, Finland
[3]Department of Child Psychiatry, Turku University Hospital, Turku, Finland
[4]ITLA Children's Foundation, Helsinki, Finland
[5]Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland
[6]Centre for Research in Family Health, IWK Health Centre, Halifax, NS, Canada
[7]Faculty of Science, Dalhousie University, Halifax, NS, Canada
[8]Strongest Families Institute, Halifax, NS, Canada
[9]Department of Psychiatry, Dalhousie University, Halifax, NS, Canada
[*]these authors contributed equally

Corresponding Author:
Andre Sourander, MD, PhD
Department of Child Psychiatry
University of Turku
Lemminkäisenkatu 3, Teutori 3rd floor
Turku, 20014
Finland
Phone: 358 40 5571633
Email: andre.sourander@utu.fi

Abstract

Background: There is a lack of effectiveness studies when digital parent training programs are implemented in real-world practice. The efficacy of the internet-based and telephone-assisted Finnish Strongest Families Smart Website (SFSW) parent training intervention on the disruptive behavior of 4-year-old children was studied in a randomized controlled trial setting in Southwest Finland between 2011 and 2013. After that, the intervention was implemented nationwide in child health clinics from 2015 onwards.

Objective: The main aim of this study was to compare the treatment characteristics and effectiveness of the SFSW parent training intervention between the families who received the intervention when it was implemented as a normal practice in child health clinics and the families who received the same intervention during the randomized controlled trial.

Methods: The implementation group comprised 600 families who were recruited in the SFSW intervention between January 2015 and May 2017 in real-world implementation. The RCT intervention group comprised 232 families who were recruited between October 2011 and November 2013. The same demographic and child and parent measures were collected from both study groups and were compared using linear mixed-effect models for repeated measurements. The child psychopathology and functioning level were measured using the Child Behavior Checklist (CBCL) version 1.5-5 for preschool children, the Inventory of Callous-Unemotional Traits (ICU), and a modified version of the Barkley Home Situations Questionnaire. Parenting skills were measured using the 31-item Parenting Scale and the shorter 21-item Depression, Anxiety and Stress Scale (DASS-21). The estimated child and parent outcomes were adjusted for CBCL externalizing scores at baseline, maternal education, duration of the behavior problems, and paternal age. The baseline measurements of each outcome were used as covariates.

Results: The implementation group was more likely to complete the intervention than the RCT intervention group (514/600, 85.7% vs 176/232, 75.9%, respectively; P<.001). There were no significant differences between the implementation and RCT

intervention groups with regard to child measures, including CBCL externalizing score (–0.2, 95% CI –1.3 to 1.6; *P*=.83), total score (–0.7, 95% CI –3.0 to 4.5; *P*=.70), internalizing score (–0.3, 95% CI –1.0 to 1.6; *P*=.64), and ICU total score (–0.4, 95% Cl –1.9 to 1.2; *P*=.64). No significant difference was detected in the Parenting Scale total score (0.0, 95% Cl –0.1 to 0.1; *P*=.50), while DASS-21 total score differed nearly significantly (2.5, 95% Cl 0.0-5.1; *P*=.05), indicating better improvement in the implementation group.

Conclusions: The internet-based and telephone-assisted SFSW parent training intervention was effectively implemented in real-world settings. These findings have implications for addressing the unmet needs of children with disruptive behavior problems. Our initiative could also provide a quick socially distanced solution for the considerable mental health impact of the COVID-19 pandemic.

Trial Registration: ClinicalTrials.gov NCT01750996; https://clinicaltrials.gov/ct2/show/NCT01750996

International Registered Report Identifier (IRRID): RR2-10.1186/1471-2458-13-985

(J Med Internet Res 2022;24(4):e27900) doi:10.2196/27900

KEYWORDS

parent training; early intervention; implementation; disruptive behavior; behavior problems; preschool children; internet-assisted; child mental health; mental health; behavior; intervention; children; parents

Introduction

Background

There is mounting evidence from randomized controlled trials (RCTs) that parents can be trained to tackle and reduce children's disruptive behavior and improve their parenting skills [1-3]. These findings are of upmost importance to public health professionals because children who exhibit disruptive behavior face increased risks of adult psychiatric disorders, substance use, crime, suicide, and other adversities [4,5]. Sufficiently strong evidence has been published on the efficacy of parent training to suggest that psychosocial services for children should include evidence-based parent training programs [6,7]. The need for services to tackle childhood disruptive behavior is enormous, but only a minority of families receive them [8]. There are challenges to implementing traditional face-to-face group-based parent training programs in real-world settings. One issue is the large number of barriers such as high cost, poor access, inconvenience, and low fidelity [3,9]. Another is keeping the content of the intervention consistent with the original evidence-based treatment [2].

Digitally assisted interventions are becoming more common, as they can overcome the barriers associated with conventional programs [3,9]. They are also likely to become increasingly popular, as child mental health services struggle to deal with the considerable increase in demand for their services as a result of the COVID-19 pandemic. This unprecedented global health emergency is expected to have major ongoing effects on child mental health owing to factors such as quarantine measures, social distancing, and school closures [10]. The pandemic started at a time when resources were already under pressure, and these are expected to be further affected by manpower shortages and a global recession that puts even greater pressure on health budgets. Digitally assisted interventions are cost-effective solutions that require fewer personnel and can reach geographically remote areas that would otherwise be outside of the reach of specialist services.

RCT studies have shown that remote and digitally assisted parent training programs have worked well in clinical settings [11,12]. We previously reported 12-month and 24-month follow-up studies of the first RCT on the Strongest Families Smart Website (SFSW). This RCT used a population-based sample and provided an internet-based parent training intervention with weekly telephone coaching [13-15]. The development of the SFSW intervention was based on the social learning and cognitive behavioral theories as well as positive parenting practices [16-18]. The target population was 4-year-old children who displayed high levels of disruptive behavior when they were screened during annual health checkups at child health clinics across Southwest Finland. The RCT showed that the children and parents who received the SFSW parent training program derived significant benefits from the initiative. The children displayed significant reductions in their disruptive behavior and other psychiatric symptom domains at their 24-month follow-up assessments. They also demonstrated the same improvements when they were compared with an education control group. The education control group received access to a static website that provided parents with information on how to tackle behavior problems and 1 phone call with a coach. Improved parenting skills were maintained in the intervention group at the 24-month follow-up assessment [14].

There has been growing interest in implementation research during the past 2 decades. Dissemination refers to how knowledge of new practices is actively and passively extended, and implementation refers to how new practices are incorporated into real-world environments. The term *implementation gap* is used to refer to the difference between our knowledge of *what works* and *how it works* [19,20]. Unfortunately, the strong effects that are observed in controlled RCT settings can weaken or become ambiguous when they are implemented in real-world settings [9]. Meta-analyses have shown that effective implementation has been associated with better outcomes, and the magnitude of the mean effect sizes was considerably higher when programs were carefully implemented and when fidelity was confirmed [21]. Successfully converting psychosocial interventions from experimental environments to real-world practice requires a solid framework and a structured implementation plan [22]. Research on evidence-based parent training programs after the RCT stage has often focused on

examining the characteristics of an optimal implementation environment rather than maintaining the effectiveness of the intervention. We are not aware of any previous reports on the effectiveness of implementing digital interventions for disruptive behavior so that they can form part of the routine care that children below school age can receive.

Objectives

This was the first study to report the effectiveness of the SFSW internet-based and telephone-assisted parent training program for preschool children when it was implemented in real-world settings. The intervention was put into practice after the population-based screening was used to identify children with disruptive behavior problems during routine visits to Finnish child health clinics at the age of 4 years. The primary aim was to report the changes in the children's psychopathology and functioning level and any improvement in their families' parenting skills. The children and their parents were followed up 6 months after the SFSW intervention was nationally implemented in Finnish primary care child health clinics. We compared the treatment characteristics and effectiveness between the families who received the SFSW intervention in these real-world settings from January 2015 to May 2017 and the families who received the intervention during the RCT from October 2011 to November 2013. Finally, we verified the findings by carrying out the following additional analyses. The first analysis excluded families who did not complete the parent training program. The second analysis excluded the Turku study site from the implementation study group because it was the only site that participated in both the RCT intervention and the implementation phases. In the third analysis, we compared the implementation and the RCT education control group. Our hypothesis was that the effectiveness of the SFSW intervention would be maintained if the protocol used in our previous RCT and the structured implementation plan were strictly adhered to.

Methods

Study Design

This study was a longitudinal comparison of 2 parallel groups. The implementation group comprised 600 families who received the SFSW internet-based and telephone-assisted parent training program in the real-world setting between January 2015 and May 2017. The implementation phase covered 95 child health clinics in 12 administrative regions across Finland. The RCT intervention group comprised 232 families who had been recruited by 42 child health clinics in 7 administrative regions in Southwest Finland between October 2011 and November 2013. The administrative regions in both the RCT and implementation studies contained both urban and rural areas. Turku was the only region that participated in both studies.

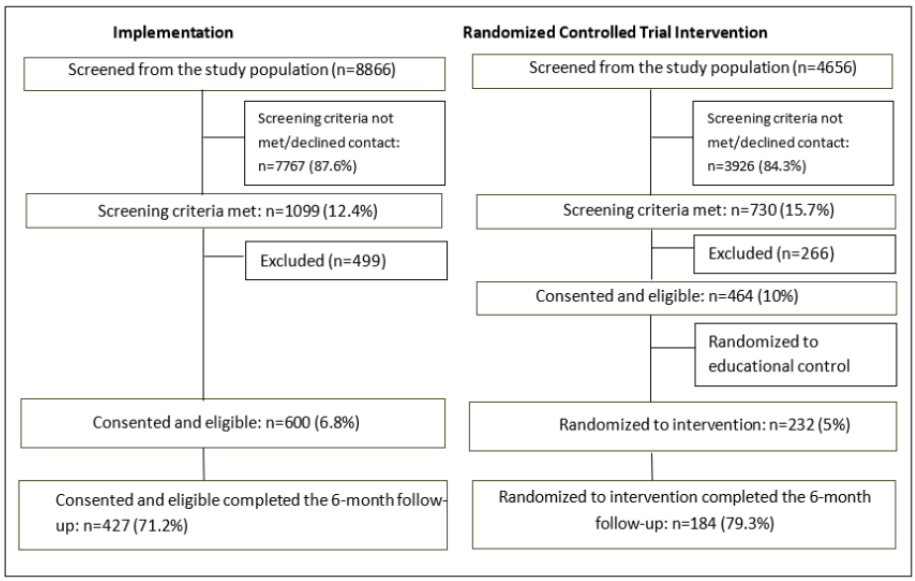
There were both differences and similarities between the implementation and the RCT intervention studies. First, the implementation group received the intervention when it was integrated as a normal practice of the child health clinics, and therefore, all families who met the inclusion criteria were eligible to enter. In the implementation phase, both participants and the health care workers received information that the SFSW parent training intervention has been evaluated as an intervention with strong documented effects by the Finnish national evaluation and classification system for evidence-based interventions [23]. This evaluation was partly based on the results of our previous RCT study [13,14]. In contrast, in the RCT, the intervention was not integrated as a normal practice of the child health clinics. Only those families who were randomized to the intervention group received the intervention. Second, in the implementation phase, an implementation plan, including decision supporting and administration component, was followed. This was important because the implementation phase included increasing number of communities in the whole Finland while the RCT was conducted in a predetermined area of Southwest Finland. Third, the most important similarity was that the content of the SFSW intervention was maintained in the implementation group as identical as possible with that of the original RCT intervention. In both groups, the same psychopathology and parenting measures were collected at baseline and 6-month follow-up. Data on children's daily activities were collected only for the implementation group immediately after the intervention and at the 6-month follow-up. The timeline of the RCT and implementation studies is shown in Figure S1 of Multimedia Appendix 1. The study protocol of the RCT has previously been published [24] and registered at ClinicalTrials.gov (NCT01750996).

Participants

This study focused on the 6-month follow-up assessments of children who displayed a high level of disruptive behavior when they were screened at 4 years of age during routine child health clinic visits. The screening procedure in the implementation study followed the same principles that were used in the RCT study. It was integrated into the standard 4-year-old child health checkups carried out by the child health clinics in the participating administrative regions [13]. All children living in Finland are invited to annual health assessments before they start school at 7 years of age, and attendance rates are just under 100% [25].

In the implementation group, the first 600 eligible parents who agreed to take part in the program received the SFSW parent training intervention. Initially, 8866 children were screened for highly disruptive behavior and 1099 (12.4%) met the screening criteria. The implementation group equated to 6.8% (600/8866) of the initial population-based sample and 54.6% (600/1099) of those who were eligible to take part. The reference group consisted of 232 families who were randomized to receive the intervention during the previous RCT study [13,14]. Information was obtained from 427 (71.2%) of the 600 families in the implementation group at the 6-month follow-up assessments compared to 184 (79.3%) of the 232 families in the RCT intervention group. Figure 1 shows the flowchart of the implementation and RCT intervention groups. The families were typically recruited within 1 month of the child's fourth birthday. They received a study information pack and were asked to bring the completed health questionnaire to the clinic.

Figure 1. Flowchart of the families in the implementation and randomized controlled trial intervention groups.



Inclusion and Exclusion Criteria

The screening measures and enrollment criteria were identical for the implementation and RCT studies [13]. Population-based screening for behavior problems was conducted for all children at the age of 4 years by using the conduct scale of the Strengths and Difficulties Questionnaire [26-28]. The parents were asked if their child had mild, moderate, or severe problems through a single question: "Overall, do you think that your child has difficulties in one or more of the following areas: emotions, behavior, or being able to get on with other people?" About 16.5% of the children (16.7% [1477/8866] and 16.3% [758/4656] in the implementation and RCT groups, respectively) who were screened, scored 5 or more out of 10 corresponding to the 80th percentile cutoff point and reported that the child had difficulties. This indicated a high level of behavior problems. The other inclusion criteria were that the parents perceived that child had at least minor difficulties in emotions, behavior, or social interactions. To participate in the study, the family had to live in an administrative region participating in the study, at least one parent had to speak native Finnish or Swedish, and they needed access to a telephone, computer, and internet connection. We excluded children who were unable to speak in full sentences, had hearing or vision impairments, or were receiving or had received behavior treatment. The exclusion criteria also included children who had been diagnosed with autism, Down syndrome, fetal alcohol syndrome, an intellectual disability, a severe mental disorder such as psychosis or depression, or who had a genetic diagnosis of mental retardation. We also excluded parents whose children did not live with them because they were subject to child protection services owing to child custody, abuse, or neglect issues. Details of the inclusion and exclusion criteria have previously been reported [24].

Procedure

The participants in the implementation and RCT intervention study groups received the SFSW parent training program, which combines an interactive website with weekly telephone coaching [15,29]. One parent was identified for each child and they filled in the web-based questionnaire. However, they were also encouraged to get the child's other parent involved in the program as much as possible. The program was guided by coaches who were professionals of health care and social services, that is, public health nurses, public nurses, or social workers from the child services. Of note, in the RCT study, there were 6 coaches, and during the implementation study, there were 10 coaches, 6 of whom had not participated in the RCT. The coaches had weekly phone calls with the parents, which were sometimes organized using texts or emails, and they monitored their progress on the website. The intervention consisted of 11 weekly themes that were explored during the interactive web-based program and the associated telephone coaching sessions. After the baseline survey, the coaches called the parents and they agreed to personalize goals tailored to individual behavior problems demonstrated by the child. The program aimed to reduce the problems identified by the parents by teaching them positive and practical parenting skills. During the first 7 weeks, the parent learned positive and practical problem-solving skills and were encouraged to develop an understanding of their child's emotional development. The primary aim was to reorient the parent so that they noticed the child's positive, not negative, behavior and reacted with a positive response. The second aim was to apply the skills in everyday situations, to plan daily activities in advance, and to use the methods they were taught to reinforce positive behavior. The final weekly themes focused on reinforcing their new skills and developing sustained positive parenting. The parents practiced the acquired skills with their child, independent of the coach's support, and learned how to sustain the skills once the program had finished. The content and the conceptual framework of the weekly themes are depicted in Table 1. Each internet-based session comprised an introduction to the weekly theme, session content, video exercises, troubleshooting tips, and a review of what the parent had learnt. Instructional videos and audio clips illustrated the practical applications of the parents' new skills. The coaches gave the parents feedback about their progress in applying the new skills and encouraged them throughout the program. They only proceeded to the next

weekly theme when the parents had mastered the skill-related questions in the current one. This typically took 1 or 2 weeks. The children did not have access to the website or take part in the coaching calls. We are not aware of any potential adverse effects of the parent training in this study or in previous studies [30].

Table 1. The content and the conceptual framework of the skill training process of the Strongest Families Smart Website internet-based and telephone-assisted parent training intervention.

Session	Training components	Key training elements	Parental goals	Coaching elements	Parental action
Introduction to the program	Telephone coaching	Set up the parents for success	Reorient the parents to "How to break the negative circle"	Working alliance Identifying behavior problems Goal setting Present the first weekly theme	Actively start to notice the good
Notice the good	Web-based material (text, videos, audio clips) Telephone coaching	Positive and active parenting	Boost self-esteem of the child and parents and change the parents' views of the child	Working alliance Evaluate the goal setting by modeling, practice such as role play, feedback, support	Notice good behavior often Positive verbal interaction and body language
Spread attention around	Web-based material (text, videos, audio clips) Telephone coaching	Positive, impartial parenting	Strengthen child's empathy skills	Same as above	Learn to spread attention actively Praise the child for interacting positively with others
Ignore whining and complaining	Web-based material (text, videos, audio clips) Telephone coaching	Positive, self-controlled parenting	Teaches parents self-regulation	Same as above	Use positive thinking to stay calm and in control of the situations
Prepare for changes	Web-based material (text, videos, audio clips) Telephone coaching	Positive, proactive parenting	Reinforce good daily routines	Same as above	Warn that behavior must change Use positive "when you do this, then this will happen" statements
Plan ahead at home	Web-based material (text, videos, audio clips) Telephone coaching	Positive, proactive parenting	Reinforce child's active role and involve them in planning	Same as above	Listens to the child's ideas, plans daily situations at home
Reinforce by rewarding	Web-based material (text, videos) Telephone coaching	Positive, active parenting	Involve the child in planning and reinforce good daily routines	Same as above	Understand realistic goal setting and how to use praises and rewards
Plan ahead outside the home	Web-based material (text, videos) Telephone coaching	Positive, proactive parenting	Reinforce child's active role and involve them in planning	Same as above	Listen to the child's ideas Plan situations outside the home
Cooperate with day care	Web-based material (text, videos) Telephone coaching	Positive cooperation and communication between parent and day care	Help child to manage and succeed	Same as above	Set realistic goals and rewards Cooperate
Plan how to use time-out	Web-based material (text, videos) Telephone coaching	Positive, self-controlled parenting	Teach self-regulation and consistency	Reassure and use positive skills How to use time-out	Learn to be consequent Plan how to manage difficult situations
Revise: Problem-solving and future application of skills	Web-based material (text, videos) Telephone coaching	Positive daily parenting in future	Remind parents of positive proactive parenting skills	Ensure that parent is using all the skills and stays on track	Understand how using skills helps to prevent setbacks

Quality Assurance and Implementation Plan

To ensure the integrity of the intervention and the accuracy of the data, several quality assurance measures were in effect during the implementation phase. These were similar to the quality assurance measures during the RCT study [13,14]. The implementation plan is summarized below and has been previously described in detail [15]. The implementation plan

was driven by 3 core components [15,19]. First was recruitment, staff selection, and training. Once the coaches were recruited, they received intensive training on the SFSW program and were supervised and regularly monitored to make sure they adhered to the protocol. Together with supervision and staff performance evaluation, this provided systematic quality assurance [15]. The second core component was ongoing supervision and staff performance evaluation. The coaches took part in systematic weekly supervision meetings and group case conferences, where they reviewed and discussed the families they were coaching. Coaches with previous experience of the SFSW program acted as supervisors. After each telephone call, the coaches assessed their own performance on a scale of 4-10. The supervisor received a message from the digital platform about self-assessments that scored 6 or more and discussed the content of the call with the coach. To ensure the fidelity of the data, about 10% of the phone calls was audited by the coach supervisors with the parent's permission and evaluated for competency. Additional training and monitoring of future calls were provided, if indicated. The coaches were required to report any adverse effects such as safety issues, abuse, or neglect to the supervisors, and the case was reported to the child protective services. Of note, 3 cases were reported during the implementation study and none during the RCT. The third core component was the decision supporting and administration. The development, delivery, and implementation process of the digital SFSW parent training intervention were centralized at the Research Center for Child Psychiatry at the University of Turku. The research group and the assisting staff of the Research Center introduced the SFSW intervention and the implementation process to the directors of child and family health services of the primary health care of each administrative region. A jointly funded research contract was signed by both parties. The research group maintained contact with the directors across the study region by organizing regular meetings and providing them with user-friendly monthly progress reports, which included the number of families who had been screened and enrolled. Training was offered to the team leaders of the child health clinics and public health nurses in order to integrate the intervention into primary health care. Moreover, local and national media were involved to increase public awareness of the SFSW intervention.

Measures

Child Measures

The outcome measures were the same in the implementation and RCT studies [13,14]. The main measurement tool used to measure disruptive behavior was the 24-item Child Behavior Checklist 1.5-5 (CBCL/1.5-5) version for preschool children [31]. The CBCL/1.5-5 asks parents to rate emotional, behavioral, and social problems and has an additional section where they can provide extra information. It yields total scores and syndrome scales for the following items: emotionally reactive, anxious/depressed, somatic complaints, withdrawn, sleep problems, attention problems, and aggressive behavior. The first 4 syndromes yield the internalizing score, while the last 2 yield the externalizing score. The CBCL/1.5-5 also includes 5 subscores from the Diagnostic and Statistical Manual of Mental Disorders, fifth edition: affective, anxiety, pervasive developmental problems, attention-deficit/hyperactivity disorder, and oppositional disorder [32]. A large cross-cultural study from 24 countries, including Finland, reported good psychometric properties and good internal consistency for the CBCL preschool version (Cronbach alphas for total, externalizing, and internalizing scores: .94, .88, and .84, respectively) [33,34]. We used the Inventory of Callous-Unemotional Traits (ICU) to measure child psychopathy traits. The instrument consists of 24 items and has been reported to have good psychometric properties for 4-year-old children [35,36]. Cronbach alphas of .93, .81, .88, and .86 have previously been reported for total score, callousness, uncaring, and unemotional scores, respectively, for 4-year-old children [35].

Daily activities were only assessed for the implementation study. Parents were asked to rate the impact of the child´s behavior during daily transitions, including getting dressed, getting ready for day care, during the evening meal, and getting ready for bed. It also covered social interactions, including playing with siblings and other children during a car or bicycle ride and in public places such as the supermarket. A Cronbach alpha of .64 was calculated using our implementation data. The questionnaire was adapted from the Barkley Home Situations Questionnaire, which asks the parent to rate whether the child's behavior causes problems during specified daily routines [37].

Parent Measures

The Parenting Scale, which is a 30-item questionnaire, was used to measure parenting skills [38,39]. Cronbach alphas of .78, .66, .68, and .50 were calculated for total score, laxness, overreactivity, and hostility, respectively, by using our implementation data. We evaluated the parents' stress, anxiety, and depression symptoms with the shorter 21-item Depression, Anxiety and Stress Scale (DASS-21) [40]. The internal consistency of DASS-21 has been reported as 0.93, 0.88, 0.82, and 0.90 for total scale and DASS-21, respectively, in a large study that represented a nonclinical sample [41].

Statistical Analyses

The analyses compared the 600 families in the real-world implementation group to the 232 families in the RCT intervention group. Categorical demographic variables, including the child, parent, and family characteristics, are presented as numbers and percentages. Continuous demographic variables, including the parents' age and duration of child's behavioral problems, are presented as means and standard deviations. We explored any differences at baseline between the 2 groups by using Pearson chi-square test or Fisher exact test for the categorical variables and the two-tailed Student *t* test for the continuous variables. The primary and secondary outcome variables were analyzed with a linear mixed-effect model for repeated measurements. The within factor was time, namely, baseline and 6-months follow-up, and the RCT intervention group and the implementation group provided the between factor. The covariates in the statistical models were CBCL externalizing scores at baseline, maternal education, duration of behavior problems, and the baseline measurement of each outcome. The statistical model used to analyze the CBCL externalizing score consisted of the group and time main effects, the group-by-time interaction effect, and the following

covariates: the CBCL externalizing score at baseline, maternal education, and duration of behavior problems. Meanwhile, the statistical model used to analyze all the secondary outcome variables, namely, the CBCL total and other CBCL subscores, ICU, the Parenting Scale, and DASS-21 consisted of the group and time main effects and the group-by-time interaction effect. It also included the following covariates: the specific secondary variable to be analyzed at baseline, the CBCL externalizing score at baseline, maternal education, and the duration of behavior problems.

The sensitivity analyses comprised the families who had completed the parent training program as well as the treatment comparisons. Turku was excluded from analysis, as it was the only site that had taken part in both the implementation and RCT intervention studies. As the study subjects in the implementation group were recruited from January 2015 to May 2017 and in the RCT intervention group from October 2011 to November 2013, we also tested the effect of the recruitment year on the CBCL externalizing score at baseline. The model included the effects of recruitment year, maternal education, and duration of behavior problems. The effect of the recruitment year was insignificant (P=.17). An additional analysis also compared the implementation to the RCT education control group. The model included the CBCL externalizing score at baseline, maternal education, duration of behavior problems, paternal age, and the baseline measurements of each outcome as covariates. A P value <.05 was considered to be statistically significant. The statistical analyses were performed using SAS 9.4 (SAS Institute).

Ethics Approval

Ethical approval for the implementation study was received from the University of Turku (approval number: 18/2018). The parents provided written informed consent for both the implementation and the RCT studies.

Results

The number of families who discontinued the program was 86 (14.3%) of the 600 families in the implementation group compared to 56 (24.1%) of the 232 families in the RCT intervention group. This meant that the odds ratio was 1.9 with a 95% CI of 1.3 to 2.8 (P<.001), as seen in Figure 2. The 6-month follow-up assessment was completed by 71.2% (427/600) of the parents in the implementation group and 79.3% (184/232) of the parents in the RCT intervention group (P<.001), as seen in Figure 1. Table 2 shows that there were no differences between the implementation group and the RCT intervention group when it came to the parent, family, and child characteristics and the factors related to parent training program. However, the mothers in the implementation group had higher educational levels than the mothers in the RCT intervention group (P=.046) and the children experienced a longer duration of behavior problems (P=.004). The mean duration of the telephone coaching calls was 37 minutes in both the implementation and the RCT intervention groups. The total duration of telephone coaching plus the average time spent on the program website was 13.8 hours in the implementation group and 14.1 hours in the RCT intervention group (P=.49).

Figure 2. Kaplan-Meier curves of families completing the program in the implementation and the randomized controlled trial intervention groups. RCT: randomized controlled trial, fixed axes according to editor comments.

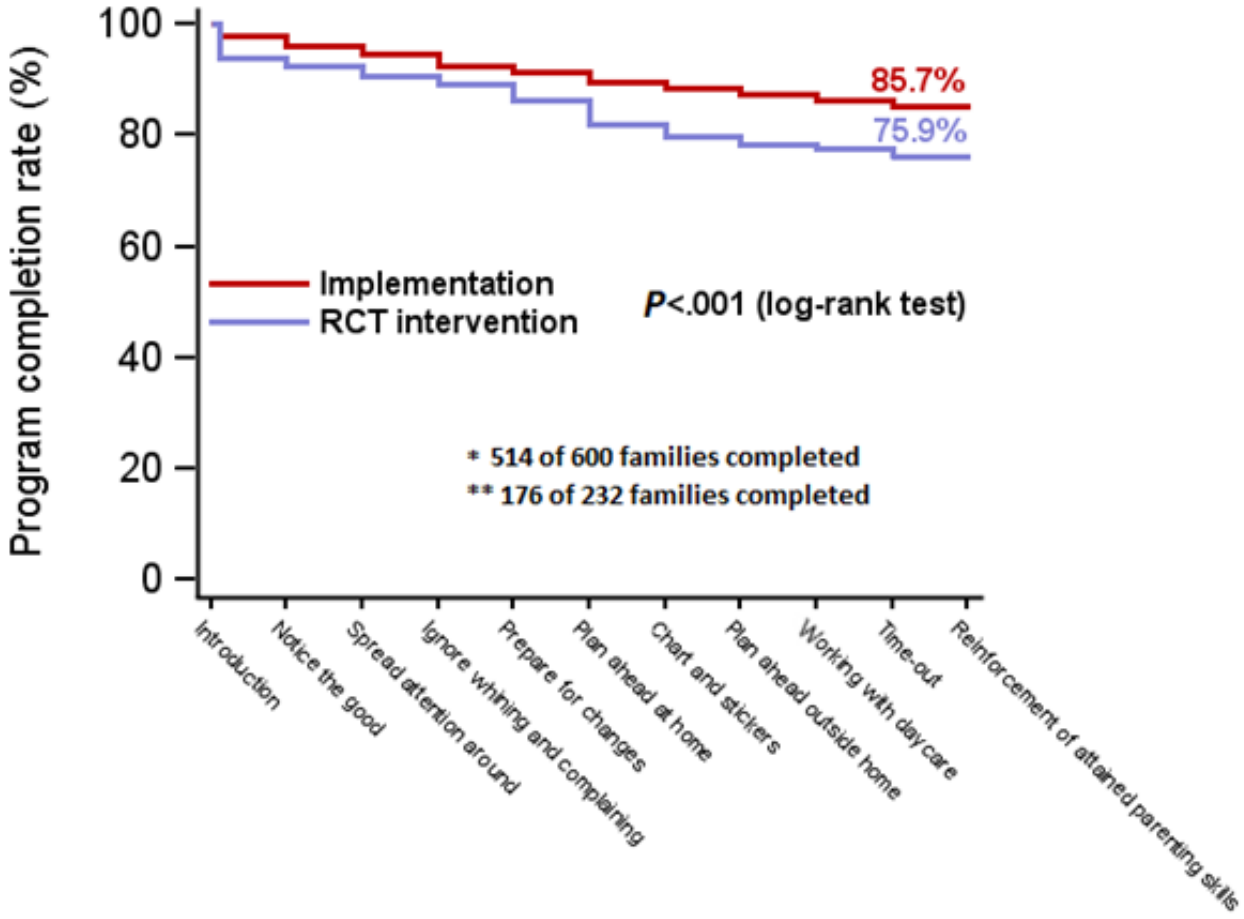


Table 2. Demographic characteristics of the families and treatment factors in the implementation and the randomized controlled trial intervention groups.

Demographics	Implementation group (n=600)	Randomized controlled trial intervention (n=232)	*P* value
Parent and family characteristics			
Family structure[a], n (%)			.54
Two biological parents	489 (81.6)	191 (83.5)	
Single biological parent	82 (13.7)	24 (10.4)	
Biological parent and foster parent	19 (3.2)	9 (3.9)	
Other	9 (1.5)	5 (2.2)	
Age (years), mean (SD)			
Maternal	30.3 (4.8)	30.5 (5.4)	.68
Paternal	32.7 (5.7)	33.2 (5.9)	.28
Maternal education[b], n (%)			.046
Elementary school or less	15 (2.5)	13 (5.7)	
Secondary education	204 (34.2)	85 (37)	
College or university degree	378 (63.3)	132 (57.4)	
Paternal education[c], n (%)			.28
Elementary school or less	27 (4.8)	16 (7.4)	
Secondary education	280 (50.1)	99 (45.8)	
College or university degree	252 (45.1)	101 (46.8)	
Child characteristics, n (%)			
Sex			.82
Female	238 (39.7)	90 (38.8)	
Male	362 (60.3)	142 (61.2)	
Day care outside home[d]			.29
Yes	476 (79.9)	192 (83.1)	
No	120 (20.1)	39 (16.9)	
Behavioral problems			.18
Minor	301 (50.2)	129 (55.6)	
Definite	252 (42)	92 (39.7)	
Severe	47 (7.8)	11 (4.7)	
Duration of problems[e]			.004
<6 months	193 (33)	102 (45.1)	
6-12 months	155 (26.5)	44 (19.5)	
>12 months	237 (40.5)	80 (35.4)	
Program characteristics, mean (SD)			
Total number of calls	10.4 (2.5)	10.1 (3.3)	.20
Duration of calls for the 11 themes (min)	37.3 (11.0)	37.3 (13.5)	.96
Duration of website access per theme (min)	45.3 (19.3)	47.8 (19.9)	.12
Total duration of calls (h)	6.5 (2.4)	6.4 (3.3)	.65
Total duration of website access (h)	7.3 (2.8)	7.5 (3.2)	.56
Total duration of program (h)	13.8 (4.3)	14.1 (5.4)	.49

[a]Missing observations: implementation group (n=1); randomized controlled trial group (n=2).

[b]Missing observations: implementation group (n=3); randomized controlled trial group (n=2). Pairwise comparisons: elementary school or less versus secondary education (*P*=.06); elementary school or less versus college or university degree (*P*=.02); secondary education versus college or university degree (*P*=.28).

[c]Missing observations: implementation group (n=41); randomized controlled trial group (n=1).

[d]Missing observations: implementation group (n=4); randomized controlled trial group (n=1).

[e]Missing observations: implementation group (n=15); randomized controlled trial group (n=6). Pairwise comparisons: <6 months versus 6-12 months (*P*=.003); <6 months versus >12 months (*P*=.01); 6-12 months versus >12 months (*P*=.42).

In the implementation group, there were significant improvements from the baseline to the 6-month follow-up assessment in the primary outcome, which was the CBCL externalizing score. The same was true for the secondary outcomes: CBCL total and internalizing scores and the total scores of the ICU, Parenting Scale, and DASS-21 (Table 3). The sensitivity analysis, which included the participants who completed the whole program (Table S1 of Multimedia Appendix 2), yielded similar estimates of the improvements in all the outcomes. Table 4 shows the mean scores of the primary outcome, CBCL externalizing score, and the secondary outcomes at baseline and 6 months in the implementation and the RCT intervention groups. There were no significant differences between the 2 groups in the CBCL externalizing, total, or internalizing scores. In addition, no significant differences were seen in the total scores of the Parenting Scale or ICU. The estimated difference of 2.5 (95% CI 0.0-5.1) points in DASS-21 nearly reached statistical significance (*P*=.05), indicating better improvement in the implementation group when it was compared to that of the RCT intervention group. Of note, the improvement in DASS-21 showed significantly better improvement in the implementation group (estimated difference 1.1, 95% Cl 0.1-2.2; *P*=.04). When only the participants who completed the whole parent training program in the implementation group were compared to those in the RCT intervention group, the results remained similar (Table S2 of Multimedia Appendix 3).

The additional analyses compared the changes in primary and secondary outcomes between the implementation and the RCT education control groups, as shown in Table S3 of Multimedia Appendix 4. There were significant differences between the groups in CBCL externalizing, total, and internalizing scores, as well as the total scores of the Parenting Scale and DASS-21. However, the total ICU score did not reach statistical significance (*P*=.27). As the city of Turku participated in both the implementation study and the RCT study, we repeated the analyses by excluding the participants living in Turku from the implementation group. This did not show any significant differences in any of the symptom scores between the study groups (Table S4 of Multimedia Appendix 5). Changes in daily activities from the baseline assessment to posttreatment and the 6-month follow-up assessment are shown in the Table S5 of Multimedia Appendix 6. This information was only obtained from the implementation group; therefore, comparisons with the RCT intervention group could not be made. There were significant improvements in all measurements for social interactions and daily transitions from baseline to posttreatment and to the 6-month follow-up. The data for daily activities were obtained from 83% (498/600) of the participants in posttreatment and 66.5% (399/600) of the participants in the follow-up.

Table 3. Change from baseline to 6 months in child psychopathology, parenting skills, and parents' stress in the implementation group.

Variable	Baseline (n=600), mean[a] (SE)	After 6 months (n=600), mean[a] (SE)	Mean change[b] (SE)	95% CI	*P*[c] value
Child measures					
Primary outcome					
Child Behavior Checklist externalizing score	21.1 (0.5)	14.8 (0.5)	6.2 (0.4)	5.5 to 7.0	<.001
Secondary outcomes					
Child Behavior Checklist Total score	48.8 (1.2)	33.6 (1.3)	15.2 (1.0)	13.3 to 17.2	<.001
Child Behavior Checklist Internalizing score	12.1 (0.4)	8.5 (0.5)	3.6 (0.4)	2.9 to 4.3	<.001
Symptom domains					
Aggression	18.0 (0.4)	12.5 (0.4)	5.5 (0.3)	4.9 to 6.1	<.001
Attention	3.1 (0.1)	2.4 (0.1)	0.7 (0.1)	0.6 to 1.0	<.001
Sleep	4.0 (0.2)	2.5 (0.2)	1.5 (0.1)	1.2 to 1.7	<.001
Withdrawn	2.4 (0.1)	1.6 (0.1)	0.8 (0.1)	0.6 to 1.0	<.001
Somatic	2.9 (0.1)	2.0 (0.2)	0.8 (0.1)	0.6 to 1.1	<.001
Anxious	2.9 (0.1)	2.0 (0.1)	0.8 (0.1)	0.6 to 1.0	<.001
Emotional	3.9 (0.2)	2.8 (0.2)	1.2 (0.1)	0.9 to 1.4	<.001
Diagnostic and Statistical Manual of Mental Disorders, fifth edition subscores					
Affective problems	3.3 (0.1)	2.0 (0.2)	1.3 (0.1)	1.1 to 1.5	<.001
Anxiety problems	4.2 (0.2)	2.9 (0.2)	1.4 (0.1)	1.1 to 1.6	<.001
PDD[d] problems	4.7 (0.2)	3.3 (0.2)	1.4 (0.2)	1.1 to 1.7	<.001
ADHD[e] problems	6.0 (0.2)	4.5 (0.2)	1.6 (0.1)	1.3 to 1.8	<.001
ODD[f] problems	6.5 (0.2)	4.6 (0.2)	1.9 (0.1)	1.6 to 2.1	<.001
Inventory of Callous-Unemotional Traits					
Total	24.6 (0.5)	20.6 (0.5)	4.0 (0.4)	3.2 to 4.7	<.001
Callousness	8.3 (0.2)	6.2 (0.2)	2.2 (0.2)	1.8 to 2.5	<.001
Uncaring	13.2 (0.2)	11.6 (0.3)	1.6 (0.2)	1.3 to 2.0	<.001
Unemotional	3.1 (0.1)	2.9 (0.1)	0.2 (0.1)	–0.1 to 0.4	.30
Parent measures					
Parenting scale					
Total	3.2 (0.0)	2.7 (0.0)	0.6 (0.0)	0.5 to 0.6	<.001
Laxness	2.7 (0.0)	2.2 (0.0)	0.4 (0.1)	0.4 to 0.5	<.001
Overreactivity	3.9 (0.1)	3.1 (0.1)	0.8 (0.0)	0.7 to 0.9	<.001
Hostility	1.9 (0.0)	1.6 (0.1)	0.3 (0.1)	0.3 to 0.4	<.001
21-item Depression, Anxiety and Stress Scale short form					
Total	18.5 (1.1)	12.1 (1.1)	6.4 (0.8)	4.9 to 7.9	<.001
Depression	5.2 (0.5)	3.1 (0.5)	2.1 (0.3)	1.4 to 2.7	<.001
Anxiety	2.4 (0.3)	1.4 (0.3)	1.1 (0.2)	0.7 to 1.4	<.001
Stress	11.0 (0.5)	7.7 (0.5)	3.3 (0.4)	2.6 to 4.0	<.001

[a]Least-squares means.

[b]Change from baseline to 6 months after providing informed consent.

[c]Adjusted with maternal education and duration of problems.

[d]PDD: pervasive developmental disorder.

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

[f]ODD: oppositional defiant disorder.

Table 4. Mean changes from baseline to 6 months in child psychopathology, parenting skills, and parents' stress in the implementation and randomized controlled trial intervention groups.

Variable	Mean (SE) change from baseline to 6 months		Implementation versus RCT intervention, mean (95% CI)	P^c value
	Implementation group (n=600), mean[a] (SE)	RCT[b] intervention (n=232), mean[a] (SE)		
Child measures				
Primary outcome				
Child Behavior Checklist externalizing score	6.3 (0.4)	6.1 (0.6)	–0.2 (–1.3 to 1.6)	.83
Secondary outcomes				
Child Behavior Checklist total score	15.3 (1.0)	14.6 (1.6)	–0.7 (–3.0 to 4.5)	.70
Child Behavior Checklist internalizing score	3.7 (0.4)	3.4 (0.6)	–0.3 (–1.0 to 1.6)	.64
Symptom domains				
Aggression	5.5 (0.3)	5.5 (0.5)	–0.0 (–1.2 to 1.3)	.95
Attention	0.7 (0.1)	0.6 (0.1)	–0.1 (–0.2 to 0.4)	.53
Sleep	1.5 (0.1)	1.5 (0.2)	–0.0 (–0.5 to 0.5)	1.0
Withdrawn	0.8 (0.1)	0.5 (0.2)	–0.3 (–0.0 to 0.7)	.08
Somatic	0.8 (0.1)	0.6 (0.2)	–0.2 (–0.2 to 0.7)	.29
Anxious	0.9 (0.1)	1.0 (0.2)	–0.1 (0.5 to 0.3)	.62
Emotional	1.2 (0.1)	1.3 (0.2)	–0.1 (–0.7 to 0.4)	.58
Diagnostic and Statistical Manual of Mental Disorders, fifth edition subscores				
Affective problems	1.3 (0.1)	1.3 (0.2)	0.0 (–0.4 to 0.5)	.95
Anxiety problems	1.4 (0.1)	1.5 (0.2)	–0.1 (–0.6 to 0.4)	.69
PDD[d] problems	1.4 (0.2)	1.2 (0.3)	0.2 (–0.3 to 0.8)	.41
ADHD[e] problems	1.6 (0.1)	1.2 (0.2)	3.5 (–0.2 to 0.9)	.17
ODD[f] problems	1.9 (0.1)	2.2 (0.2)	–0.3 (–0.7– 0.2)	.26
Inventory of Callous-Unemotional Traits				
Total	4.0 (0.4)	4.3 (0.7)	–0.4 (–1.9 to 1.2)	.64
Callousness	2.0 (0.2)	2.1 (0.3)	–0.1 (–0.7– 0.8)	.83
Uncaring	1.6 (0.2)	1.9 (0.3)	–0.2 (–1.0 to 0.5)	.53
Unemotional	0.2 (0.1)	0.3 (0.2)	–0.1 (–0.6 to 0.3)	.44
Parent measures				
Parenting scale				
Total	0.6 (0.0)	0.5 (0.0)	0.0 (–0.1 to 0.1)	.50
Laxness	0.4 (0.0)	0.4 (0.1)	0.0 (–0.1 to 0.2)	.79
Overreactivity	0.8 (0.1)	0.6 (0.1)	0.2 (–0.0 to 0.4)	.07
Hostility	0.3 (0.0)	0.3 (0.1)	0.0 (–0.1 to 0.2)	.85
21-item Depression, Anxiety and Stress Scale short form				
Total	6.4 (0.7)	3.9 (1.1)	2.5 (0.0 to 5.1)	.05
Depression	2.1 (0.3)	1.0 (0.5)	1.1 (0.1 to 2.2)	.036
Anxiety	1.0 (0.2)	0.8 (0.3)	0.3 (–0.4 to 0.1)	.44
Stress	3.3. (0.4)	2.2 (0.6)	1.1 (–0.2 to 2.4)	.09

[a]Least-squares means.

[b]RCT: randomized controlled trial.

[c]Adjusted with maternal education and duration of problems.

[d]PDD: pervasive developmental disorder.

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

[f]ODD: oppositional defiant disorder.

Discussion

This was the first population-based study to evaluate the effectiveness of an internet-based and telephone-assisted parent training intervention for children with behavior problems when it was implemented in real-world practice. The children's psychiatric problems improved, including externalizing and internalizing problems and callousness. The findings were remarkable from the perspective of the children's social development, as the program had significant effects on daily transitions and activities such as getting dressed, dining behavior, activities outside the home, and interactions with other people. Parents reported that their parenting skills had improved and they demonstrated less distress in dealing with their children at the 6-month follow-up. Most importantly, this study shows that the improvements that had been achieved were similar to those reported for the intervention group in the RCT. There was no difference in the changes in the children's psychiatric problems or parenting skills when the implementation and RCT groups were compared. Furthermore, when changes between the implementation and RCT education control groups were compared, the implementation group showed significantly better improvements in the children's externalizing and internalizing problems as well as in parenting skills and parents' distress. In addition to the effectiveness of the treatment, the ability to engage and retain parents in the program is one of the keys to successful parent training interventions [42-44]. Previously, we reported high parental satisfaction levels in both the RCT and implementation groups [15]. High satisfaction levels and the quality of relationships between parents and professionals have been associated with greater improvements in the effectiveness of interventions [45,46]. The dropout rate in our RCT study was 24%, while previous studies on digital parenting interventions report usually 30%-50% dropout rates [12,47-50]. In general, high dropout rates in digital interventions have been especially associated with nonguided interventions [43,51-54]. The reasons for the exceptionally low dropout rate in the implementation phase (14%) are likely to be multiple. One possible explanation is that in the implementation phase, the SFSW intervention that was offered had gained research-based evidence and the benefits of it were known and communicated to the professionals in the primary health care, especially in child health clinics, and to parents and largely in the media. Thus, the public and the professionals were aware of the intervention and its benefits. It is very likely that this convinced both health nurses at the child health clinics who motivate the parents in engaging in the program and the concerned parents tackling with their child's challenges.

In order to successfully implement interventions, we need to know whether they work and *why* they work [19]. Success can be related to how appropriate the background theory is, the context where the intervention takes place, practical issues such as how easy it is to attend sessions, and specific intervention practices such as practicing specific parenting skills [55]. Our SFSW intervention fulfilled these criteria well. It was based on the social learning theory and the cognitive behavioral theory as well as principles of positive parenting, which provided a sound theoretical framework for the intervention. The context of the program was well-defined, including a clear definition of the population that the program was aimed at, and there were clear inclusion and exclusion criteria. The program also had a clear structure, including a description of the core components, which was practiced through modeling, practice, feedback, and support. It has previously been emphasized that a solid framework and a structured implementation plan are needed to successfully make the transition from evidence-based psychosocial interventions to *real-world* clinical practice [22]. We systematically followed a structured plan during the implementation process [15]. The SFSW program contained the core implementation drivers that facilitated the process when intervention was implemented in the primary health care. The same quality assurance measures were in place during the RCT and implementation phase. These were based on the centralized delivery of the intervention, which used a digital platform and ongoing training, monitoring, and supervision of the program coaches. It is important to note that the primary health care staff were also provided with ongoing training. In addition, the program was effectively administered by including regular meetings with the directors of the child and family services and providing them with user-friendly reports. Media coverage raised awareness, and this made it easier to recruit families and increased the perceived value of the program [15].

Several practical features of the program may have paved the way for positive outcomes during the real-world implementation. First, the program was much easier for the parents than face-to-face interventions because they did not need to leave home or work or make childcare arrangements. Second, the telephone coaching provided immediate problem-solving, which may have been more rewarding for the parents than communicating using emails or text messages. A recent meta-analysis showed that digital interventions that included support and guidance, such as telephone calls, had larger effect sizes on mental health outcomes than smartphone interventions without any personal support [56]. Third, the coaches were well-trained and formed good relationships with the parents [15], which is central to the success of any intervention [57].

There were some limitations in our study. First, although the parental and child outcomes were measured using well-validated questionnaires, they were rated by the same person, namely, the parent. One parent was identified for each child, but they were also encouraged to get the child's other parent involved in the program as much as possible. Further details on the level of parental involvement could have added to the richness of the data, but there were practical limitations to collecting this. To

reduce the possibility of the common rater variance, observations by other informants such as day care personnel could have validated our findings. Second, we have discussed mechanisms that could have been responsible for the positive outcomes. However, there is very little empirical evidence on whether the effects of the intervention resulted from the internet sessions, the personal telephone coaching, parental motivation, or a combination of those factors. Further studies need to examine factors that explain these positive outcomes. Personalized medicine is increasingly being used to move away from one-size-fits-all interventions to those that are more tailored to individual needs. This approach could yield useful information on the mechanisms underlying interventions and enable more accurate targeting.

The target group, content, and effectiveness of the intervention were maintained when the implementation group results were compared with the findings of the RCT intervention. Internet-based telephone-assisted parent training interventions may have advantages over traditional group-based treatment approaches when the goal is to identify children at risk in the community at an early stage. This new approach can provide effective parent training for a large number of families, including many who would not normally participate in clinic-based services. Referring families who need parent training to clinical services often results in substantial delays and they need other support while they are waiting. Digitally delivered interventions move child mental health treatment outside traditional clinics and into people's homes and schools, increasing access and reducing stigma. In addition, they can be increased to help more families, and parents are more likely to stay with the program until the end. There is a global shortage of skilled staff who can address child mental health problems in low- and high-income countries and even in countries with public health care [58,59]. This could become an even greater issue when demand inevitably increases because of the impact of the COVID-19 pandemic on children and the effects of the expected global recession on health care budgets. Our study highlights the positive findings that were demonstrated when our internet-based training and phone coaching initiative provided support for the parents of children with behavior problems, who were identified using population-based screening at primary health care. This initiative made the successful transition from an RCT to real-world settings, and our findings may have potential global implications for addressing the unmet needs of children with mental health issues if the findings are repeated in other sociocultural contexts.

Acknowledgments

We wish to thank the family coaches and Project Engineer Atte Sinokki from the Department of Child Psychiatry, University of Turku, as well as the public health directors and health nurses at the participating municipal child health clinics who made this study possible. This research was funded by the Academy of Finland Flagship Program Inequalities, Interventions, and a New Welfare State (decision 320162), the Academy of Finland Health from Cohorts and Biobanks Program (decision 308552), and the European Research Council under the European Union's Horizon 2020 research and innovation program (grant 101020767) (author AS).

Conflicts of Interest

AS is the founder and director of Digifamilies, which provides evidence-based treatments to Finnish public health services. The Strongest Families Institute (SFI) is a not-for-profit organization that delivers services to Canadian families. PJM is the cofounder and Chair of SFI Board of Directors. The other authors have no conflicts of interest to declare.

Multimedia Appendix 1
Timeline of the randomized controlled trial and implementation studies.
[DOCX File , 13 KB - jmir_v24i4e27900_app1.docx]

Multimedia Appendix 2
Change from baseline to 6 months in child psychopathology, parenting skills and parents' stress in the Implementation group for participants who completed the program.
[DOCX File , 31 KB - jmir_v24i4e27900_app2.docx]

Multimedia Appendix 3
Mean changes from baseline to 6 months in child psychopathology, parenting skills, and parents' stress in the implementation and the randomized controlled trial intervention groups for participants who completed the program.
[DOCX File , 23 KB - jmir_v24i4e27900_app3.docx]

Multimedia Appendix 4
Mean changes from baseline to 6 months in child psychopathology, parenting skills, and parents' stress in the implementation and randomized controlled trial educational control groups.
[DOCX File , 23 KB - jmir_v24i4e27900_app4.docx]

Multimedia Appendix 5
Change from baseline to 6 months in child psychopathology, parenting skills, and parents' stress in the implementation and randomized controlled trial intervention groups. The city of Turku is excluded from the implementation data.
[DOCX File , 25 KB - jmir_v24i4e27900_app5.docx]

Multimedia Appendix 6
Change from baseline to posttreatment and 6 months in daily activities and social interactions in the implementation group (n=600).
[DOCX File , 26 KB - jmir_v24i4e27900_app6.docx]

References

1. Dretzke J, Davenport C, Frew E, Barlow J, Stewart-Brown S, Bayliss S, et al. The clinical effectiveness of different parenting programmes for children with conduct problems: a systematic review of randomised controlled trials. Child Adolesc Psychiatry Ment Health 2009 Mar 04;3(1):7 [FREE Full text] [doi: 10.1186/1753-2000-3-7] [Medline: 19261188]
2. Michelson D, Davenport C, Dretzke J, Barlow J, Day C. Do evidence-based interventions work when tested in the "real world?" A systematic review and meta-analysis of parent management training for the treatment of child disruptive behavior. Clin Child Fam Psychol Rev 2013 Mar;16(1):18-34. [doi: 10.1007/s10567-013-0128-0] [Medline: 23420407]
3. Baumel A, Pawar A, Kane JM, Correll CU. Digital Parent Training for Children with Disruptive Behaviors: Systematic Review and Meta-Analysis of Randomized Trials. J Child Adolesc Psychopharmacol 2016 Oct;26(8):740-749. [doi: 10.1089/cap.2016.0048] [Medline: 27286325]
4. Caspi A, Moffitt TE, Newman DL, Silva PA. Behavioral observations at age 3 years predict adult psychiatric disorders. Longitudinal evidence from a birth cohort. Arch Gen Psychiatry 1996 Nov;53(11):1033-1039. [doi: 10.1001/archpsyc.1996.01830110071009] [Medline: 8911226]
5. Sourander A, Jensen P, Davies M, Niemelä S, Elonheimo H, Ristkari T, et al. Who is at greatest risk of adverse long-term outcomes? The Finnish From a Boy to a Man study. J Am Acad Child Adolesc Psychiatry 2007 Sep;46(9):1148-1161. [doi: 10.1097/chi.0b013e31809861e9] [Medline: 17712238]
6. Perrin EC, Leslie LK, Boat T. Parenting as Primary Prevention. JAMA Pediatr 2016 Jul 01;170(7):637-638. [doi: 10.1001/jamapediatrics.2016.0225] [Medline: 27182902]
7. Weber L, Kamp-Becker I, Christiansen H, Mingebach T. Treatment of child externalizing behavior problems: a comprehensive review and meta-meta-analysis on effects of parent-based interventions on parental characteristics. Eur Child Adolesc Psychiatry 2019 Aug;28(8):1025-1036. [doi: 10.1007/s00787-018-1175-3] [Medline: 29948228]
8. Jensen PS, Goldman E, Offord D, Costello EJ, Friedman R, Huff B, et al. Overlooked and underserved: "action signs" for identifying children with unmet mental health needs. Pediatrics 2011 Nov;128(5):970-979 [FREE Full text] [doi: 10.1542/peds.2009-0367] [Medline: 22025589]
9. Breitenstein SM, Gross D, Christophersen R. Digital delivery methods of parenting training interventions: a systematic review. Worldviews Evid Based Nurs 2014 Jun;11(3):168-176. [doi: 10.1111/wvn.12040] [Medline: 24842341]
10. Mastnak W. Psychopathological problems related to the COVID-19 pandemic and possible prevention with music therapy. Acta Paediatr 2020 Aug;109(8):1516-1518 [FREE Full text] [doi: 10.1111/apa.15346] [Medline: 32395840]
11. McGrath PJ, Lingley-Pottie P, Thurston C, MacLean C, Cunningham C, Waschbusch DA, et al. Telephone-based mental health interventions for child disruptive behavior or anxiety disorders: randomized trials and overall analysis. J Am Acad Child Adolesc Psychiatry 2011 Nov;50(11):1162-1172. [doi: 10.1016/j.jaac.2011.07.013] [Medline: 22024004]
12. Sanders MR, Baker S, Turner KMT. A randomized controlled trial evaluating the efficacy of Triple P Online with parents of children with early-onset conduct problems. Behav Res Ther 2012 Nov;50(11):675-684. [doi: 10.1016/j.brat.2012.07.004] [Medline: 22982082]
13. Sourander A, McGrath PJ, Ristkari T, Cunningham C, Huttunen J, Lingley-Pottie P, et al. Internet-Assisted Parent Training Intervention for Disruptive Behavior in 4-Year-Old Children: A Randomized Clinical Trial. JAMA Psychiatry 2016 Apr;73(4):378-387. [doi: 10.1001/jamapsychiatry.2015.3411] [Medline: 26913614]
14. Sourander A, McGrath PJ, Ristkari T, Cunningham C, Huttunen J, Hinkka-Yli-Salomäki S, et al. Two-Year Follow-Up of Internet and Telephone Assisted Parent Training for Disruptive Behavior at Age 4. J Am Acad Child Adolesc Psychiatry 2018 Sep;57(9):658-668.e1. [doi: 10.1016/j.jaac.2018.07.001] [Medline: 30196869]
15. Ristkari T, Kurki M, Suominen A, Gilbert S, Sinokki A, Kinnunen M, et al. Web-Based Parent Training Intervention With Telephone Coaching for Disruptive Behavior in 4-Year-Old Children in Real-World Practice: Implementation Study. J Med Internet Res 2019 Apr 11;21(4):e11446 [FREE Full text] [doi: 10.2196/11446] [Medline: 30973337]
16. Bandura A. Social Learning Theory, 1st edition. Morristown, New York: General Learning Press; 1971.
17. Reinecke MA, Dattilio FM, Freeman A, editors. Cognitive Therapy with Children and Adolescents: A Casebook for Clinical Practice, 2nd ed. New York City, New York: Guilford Press; 2003.

18. Piquero AR, Jennings WG, Diamond B, Farrington DP, Tremblay RE, Welsh BC, et al. A meta-analysis update on the effects of early family/parent training programs on antisocial behavior and delinquency. J Exp Criminol 2016 Apr 13;12(2):229-248. [doi: 10.1007/s11292-016-9256-0]
19. Fixsen DL, Blase KA, Naoom SF, Wallace F. Core Implementation Components. Research on Social Work Practice 2009 May 27;19(5):531-540. [doi: 10.1177/1049731509335549]
20. May CR, Johnson M, Finch T. Implementation, context and complexity. Implement Sci 2016 Oct 19;11(1):141 [FREE Full text] [doi: 10.1186/s13012-016-0506-3] [Medline: 27756414]
21. Durlak JA, DuPre EP. Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. Am J Community Psychol 2008 Jun;41(3-4):327-350. [doi: 10.1007/s10464-008-9165-0] [Medline: 18322790]
22. Vis C, Mol M, Kleiboer A, Bührmann L, Finch T, Smit J, et al. Improving Implementation of eMental Health for Mood Disorders in Routine Practice: Systematic Review of Barriers and Facilitating Factors. JMIR Ment Health 2018 Mar 16;5(1):e20 [FREE Full text] [doi: 10.2196/mental.9769] [Medline: 29549072]
23. Kasvun tuki (early interventions) resource. ITLA Children's Foundation. 2016 Jun 23. URL: https://kasvuntuki.fi/en/interventions/voimaperheet/ [accessed 2021-02-15]
24. McGrath PJ, Sourander A, Lingley-Pottie P, Ristkari T, Cunningham C, Huttunen J, et al. Remote population-based intervention for disruptive behavior at age four: study protocol for a randomized trial of Internet-assisted parent training (Strongest Families Finland-Canada). BMC Public Health 2013 Oct 21;13:985 [FREE Full text] [doi: 10.1186/1471-2458-13-985] [Medline: 24139323]
25. Child welfare clinic. National Institute for Health and Welfare. URL: https://thl.fi/fi/web/lapset-nuoret-ja-perheet/peruspalvelut/aitiys_ja_lastenneuvola/lastenneuvola [accessed 2021-02-15]
26. Goodman R. The Strengths and Difficulties Questionnaire: a research note. J Child Psychol Psychiatry 1997 Jul;38(5):581-586. [doi: 10.1111/j.1469-7610.1997.tb01545.x] [Medline: 9255702]
27. Koskelainen M, Sourander A, Kaljonen A. The Strengths and Difficulties Questionnaire among Finnish school-aged children and adolescents. Eur Child Adolesc Psychiatry 2000 Dec;9(4):277-284. [doi: 10.1007/s007870070031] [Medline: 11202103]
28. Borg A, Kaukonen P, Salmelin R, Joukamaa M, Tamminen T. Reliability of the strengths and difficulties questionnaire among Finnish 4-9-year-old children. Nord J Psychiatry 2012 Dec;66(6):403-413. [doi: 10.3109/08039488.2012.660706] [Medline: 22397524]
29. Ristkari T, Mishina K, Lehtola M, Sourander A, Kurki M. Public health nurses' experiences of assessing disruptive behaviour in children and supporting the use of an Internet-based parent training programme. Scand J Caring Sci 2020 Jun;34(2):420-427 [FREE Full text] [doi: 10.1111/scs.12744] [Medline: 31487074]
30. Barlow J, Smailagic N, Huband N, Roloff V, Bennett C. Group-based parent training programmes for improving parental psychosocial health. Cochrane Database Syst Rev 2012 Jun 13(6):CD002020. [doi: 10.1002/14651858.CD002020.pub3] [Medline: 22696327]
31. Achenbach TM, Rescorla LA. Manual for the ASEBA Preschool Forms & Profiles. Burlington, VT: AESBA; 2000.
32. American Psychiatric Association. Anxiety disorders. In: Diagnostic and Statistical Manual of Mental Disorders, 5th ed. Washington DC: American Psychiatric Publishing; May 22, 2013.
33. Sourander A. Emotional and behavioural problems in a sample of Finnish three-year-olds. Eur Child Adolesc Psychiatry 2001 Jun;10(2):98-104. [doi: 10.1007/s007870170032] [Medline: 11469289]
34. Rescorla LA, Achenbach TM, Ivanova MY, Harder VS, Otten L, Bilenberg N, et al. International comparisons of behavioral and emotional problems in preschool children: parents' reports from 24 societies. J Clin Child Adolesc Psychol 2011;40(3):456-467 [FREE Full text] [doi: 10.1080/15374416.2011.563472] [Medline: 21534056]
35. Kimonis ER, Fanti KA, Anastassiou-Hadjicharalambous X, Mertan B, Goulter N, Katsimicha E. Can Callous-Unemotional Traits be Reliably Measured in Preschoolers? J Abnorm Child Psychol 2016 May 7;44(4):625-638 [FREE Full text] [doi: 10.1007/s10802-015-0075-y] [Medline: 26344015]
36. Ezpeleta L, de la Osa N, Granero R, Penelo E, Domènech JM. Inventory of callous-unemotional traits in a community sample of preschoolers. J Clin Child Adolesc Psychol 2013;42(1):91-105. [doi: 10.1080/15374416.2012.734221] [Medline: 23095075]
37. Barkley RA, Murphy KR. Home situations questionnaire. In: Attention-Deficit Hyperactivity Disorder: A Clinical Workbook, 2nd Edition. New York City, New York: Guilford Press; 1998.
38. Arnold DS, O'Leary SG, Wolff LS, Acker MM. The Parenting Scale: A measure of dysfunctional parenting in discipline situations. Psychological Assessment 1993;5(2):137-144. [doi: 10.1037/1040-3590.5.2.137]
39. Rhoades KA, O'Leary SG. Factor structure and validity of the parenting scale. J Clin Child Adolesc Psychol 2007;36(2):137-146. [doi: 10.1080/15374410701274157] [Medline: 17484687]
40. Lovibond SH, Lovibond PF. Manual for the Depression Anxiety Stress Scales, 2nd ed. Sydney, NSW: Psychology Foundation of Australia; 1995.
41. Henry JD, Crawford JR. The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. Br J Clin Psychol 2005 Jun;44(Pt 2):227-239. [doi: 10.1348/014466505X29657] [Medline: 16004657]

42. Chacko A, Jensen SA, Lowry LS, Cornwell M, Chimklis A, Chan E, et al. Engagement in Behavioral Parent Training: Review of the Literature and Implications for Practice. Clin Child Fam Psychol Rev 2016 Sep;19(3):204-215. [doi: 10.1007/s10567-016-0205-2] [Medline: 27311693]
43. Dadds MR, Sicouri G, Piotrowska PJ, Collins DA, Hawes DJ, Moul C, et al. Keeping Parents Involved: Predicting Attrition in a Self-Directed, Online Program for Childhood Conduct Problems. J Clin Child Adolesc Psychol 2019;48(6):881-893. [doi: 10.1080/15374416.2018.1485109] [Medline: 30067388]
44. Ozbek A, Gencer O, Mustan AT. Which parents dropout from an evidence-based parenting programme (Triple-P) at CAMHS? Comparison of programme-completing and dropout parents. Clin Child Psychol Psychiatry 2019 Jan;24(1):144-157. [doi: 10.1177/1359104518792294] [Medline: 30101599]
45. Kazdin AE, Whitley MK. Pretreatment social relations, therapeutic alliance, and improvements in parenting practices in parent management training. J Consult Clin Psychol 2006 Apr;74(2):346-355. [doi: 10.1037/0022-006X.74.2.346] [Medline: 16649879]
46. Schmidt F, Chomycz S, Houlding C, Kruse A, Franks J. The Association Between Therapeutic Alliance and Treatment Outcomes in a Group Triple P Intervention. J Child Fam Stud 2013 Jul 3;23(8):1337-1350. [doi: 10.1007/s10826-013-9792-4]
47. van der Zanden RAP, Speetjens P, Arntz KS, Onrust SA. Online group course for parents with mental illness: development and pilot study. J Med Internet Res 2010 Dec 19;12(5):e50 [FREE Full text] [doi: 10.2196/jmir.1394] [Medline: 21169178]
48. Enebrink P, Högström J, Forster M, Ghaderi A. Internet-based parent management training: a randomized controlled study. Behav Res Ther 2012 Apr;50(4):240-249. [doi: 10.1016/j.brat.2012.01.006] [Medline: 22398153]
49. Kable JA, Coles CD, Strickland D, Taddeo E. Comparing the Effectiveness of On-Line versus In-Person Caregiver Education and Training for Behavioral Regulation in Families of Children with FASD. Int J Ment Health Addict 2012 Dec;10(6):791-803 [FREE Full text] [doi: 10.1007/s11469-012-9376-3] [Medline: 31024223]
50. Hall CM, Bierman KL. Technology-assisted Interventions for Parents of Young Children: Emerging Practices, Current Research, and Future Directions. Early Child Res Q 2015;33:21-32 [FREE Full text] [doi: 10.1016/j.ecresq.2015.05.003] [Medline: 27773964]
51. Baumel A, Muench F, Edan S, Kane JM. Objective User Engagement With Mental Health Apps: Systematic Search and Panel-Based Usage Analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
52. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: A systematic review and meta-analysis. J Affect Disord 2020 Feb 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
53. Karyotaki E, Efthimiou O, Miguel C, Bermpohl FMG, Furukawa TA, Cuijpers P, Individual Patient Data Meta-Analyses for Depression (IPDMA-DE) Collaboration, et al. Internet-Based Cognitive Behavioral Therapy for Depression: A Systematic Review and Individual Patient Data Network Meta-analysis. JAMA Psychiatry 2021 Apr 01;78(4):361-371 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.4364] [Medline: 33471111]
54. 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]
55. Blasé K, Fixsen DL. Core intervention components: identifying and operationalizing what makes programs work. ASPE: Office of the Assistant Secretary for Planning and Evaluation. 2013 Jan 13. URL: https://aspe.hhs.gov/report/core-intervention-components-identifying-and-operationalizing-what-makes-programs-work [accessed 2021-02-15]
56. 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;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
57. Flückiger C, Del Re AC, Wampold BE, Horvath AO. The alliance in adult psychotherapy: A meta-analytic synthesis. Psychotherapy (Chic) 2018 Dec;55(4):316-340. [doi: 10.1037/pst0000172] [Medline: 29792475]
58. Juengsiragulwit D. Opportunities and obstacles in child and adolescent mental health services in low- and middle-income countries: a review of the literature. WHO South East Asia J Public Health 2015;4(2):110-122 [FREE Full text] [doi: 10.4103/2224-3151.206680] [Medline: 28607309]
59. Measuring progress: Commitments to support and expand the mental health workforce in England. British Medical Association. URL: https://www.bma.org.uk/collective-voice/policy-and-research/education-training-and-workforce/mental-health-workforce-report [accessed 2021-02-15]

Abbreviations

CBCL: Child Behavior Checklist
DASS-21: 21-item Depression, Anxiety and Stress Scale
ICU: Inventory of Callous-Unemotional Traits
RCT: randomized controlled trial
SFSW: Strongest Families Smart Website

Edited by A Mavragani; submitted 03.03.21; peer-reviewed by Y Perry, A Woolard; comments to author 20.04.21; revised version received 31.08.21; accepted 15.12.21; published 04.04.22.

Please cite as:
Sourander A, Ristkari T, Kurki M, Gilbert S, Hinkka-Yli-Salomäki S, Kinnunen M, Pulkki-Råback L, McGrath PJ
Effectiveness of an Internet-Based and Telephone-Assisted Training for Parents of 4-Year-Old Children With Disruptive Behavior: Implementation Research
J Med Internet Res 2022;24(4):e27900
URL: https://www.jmir.org/2022/4/e27900
doi:10.2196/27900
PMID:35377332

Original Paper

Association Between Mobile App Use and Caregivers' Support System, Time Spent on Caregiving, and Perceived Well-being: Survey Study From a Large Employer

Pelin Ozluk[1], PhD; Rebecca Cobb[1], MA; Alyson Hoots[2], BA; Malgorzata Sylwestrzak[1], MA

[1]HealthCore Inc, Wilmington, DE, United States

[2]Anthem Inc, Indianapolis, IN, United States

Corresponding Author:
Pelin Ozluk, PhD
HealthCore Inc
123 Justison Street Suite 200
Wilmington, DE, 19801
United States
Phone: 1 302 230 2023
Email: pozluk@healthcore.com

Abstract

Background: Mobile technology to address caregiver needs has been on the rise. There is limited evidence of effectiveness of such technologies on caregiver experiences.

Objective: This study evaluates the effectiveness of ianacare, a mobile app, among employees of a large employer. ianacare mobilizes personal social circles to help with everyday tasks. Through the use of ianacare, we evaluate the associations between coordinating caregiving tasks among a caregiver's personal support network and outcomes related to the caregiver's support system, time use, perceived productivity, and perceived health and well-being. Caregiver tasks include tasks such as meal preparation, respite care, pet care, and transportation. Time use is the measure of a caregiver's time spent on caregiving tasks and how much time they had to take off from work to attend planned or unplanned caregiving tasks.

Methods: We conducted 2 surveys to assess within-participant changes in outcomes for the unpaid, employed, caregivers after 6 weeks of using the mobile app (n=176) between March 30, 2020, and May 11, 2020. The surveys contained questions in three domains: the caregiver's support system, time use and perceived productivity, and perceived health and well-being. The results of the linear probability models are presented below.

Results: App use was significantly associated with decreasing the probability of doing most caregiving tasks alone by 9.1% points (SE 0.04; *P*=.01) and increasing the probability of at least one person helping the primary caregiver by 8.0% points (SE 0.035; *P*=.02). App use was also associated with improving the time use of the primary caregiver who took significantly less time off work to attend to caregiving duties by 12.5% points (SE 0.04; *P*=.003) and decreased the probability of spending more than 30 hours weekly on caregiving by 9.1% points (SE 0.04; *P*=.02). Additional findings on the positive impact of the app included a decrease in the probability of reporting feeling overwhelmed by caregiving tasks by 12.5% points (SE 0.04; *P*=.003) and a decrease in the probability of reporting negative health effects by 6.8% points (SE 0.04; *P*=.07) because of caregiving. Although subjects reported that COVID-19 increased their stress attributed to caregiving and prevented them from requesting help for some caregiving tasks, using the app was still associated with improvements in receiving help and lessening of the negative effects of caregiving on the caregivers.

Conclusions: App use was associated with improvements in 7 of 11 caregiver outcomes across three main categories: their support system, time spent on caregiving, and perceived health and well-being. These findings provide encouraging evidence that the mobile app can significantly reduce caregiver burden by leveraging a caregiver's support network despite the additional challenges brought by COVID-19 on caregivers.

(J Med Internet Res 2022;24(4):e28504) doi:10.2196/28504

KEYWORDS

caregiving; mobile app; mobile phone

Introduction

Background

In the United States, caregiver burden is a rising problem with a rapidly growing senior population. The number of Americans providing unpaid care has increased from 43.5 million to 53 million over the past 5 years [1]. Although people of all demographics (eg, young children and spouses with health problems) may need caregiving, older adults constitute a large proportion of care recipients. A growing population of older adults with longer life expectancies means that the number of people needing caregiving in the future is likely to continue to increase [1]. Caregiving encompasses a broad range of activities that often depend on the severity of the care recipient's diagnosis and condition [2]. In addition, the relationship between family caregivers and care recipients can determine the types of activities required. For example, taking care of a parent with mobility issues will be different from taking care of a spouse with cancer. Caregiving is associated with several negative health outcomes such as mental distress, poor self-care, sleep deprivation, and caregiver burden, which are often overlooked by clinicians [1,3]. In the remaining part of the paper, a caregiver is defined as any unpaid family caregiver who may be taking care of a relative or a friend. Caregiver burden may also negatively affect work-life balance. Approximately 10% of caregivers had to give up work entirely or retire early [1].

The literature finds that whether the burden associated with caregiving is subjective (ie, self-reported outcomes of happiness, health status, and quality of life) or objective (ie, outcomes of time spent, expenses, and taking care of daily tasks) depends on the care recipient's condition and caregiver characteristics [4-6]. For example, caring for a person living with severe physical disabilities or schizophrenia was found to be associated with subjective burden on the caregiver, although the link between these conditions and the objective burden was not clear [5,6]. Furthermore, decreased social activity and feelings of isolation as one takes the role of caregiver can exacerbate the burden of caregiving, leading to poorer physical and mental health [3]. Research also shows caregiver characteristics, such as being female or living with the care recipient, may also contribute to caregiver burden [7,8]. Another study found that perceived social support may be more consistently related to subjective burden than the actual received social support [9].

Asking for help may be especially difficult for primary caregivers. Often, they end up taking on the entire burden of caregiving, even though other resources such as community services or other family members or friends are available and willing to help [1,3]. There are many reasons why caregivers may not ask for help: financial concerns, fear of losing their privacy, or feelings of shame regarding care needs [3]. One hypothesized link between social connections and health is that the social support people receive from their network of friends and loved ones may *buffer* against the detrimental physical consequences of psychosocial stress. Increased social support is found to be associated with a decrease in caregiver burden [10].

Although most of the literature focuses on caregiving burden, research has also pointed out that there are positive aspects attributed to caregiving [7,11-13]. Specifically, caregiving can bring opportunities such as being able to give back, discovering personal strength, becoming closer to the care recipient, and gaining a sense of accomplishment and competence.

Owing to the growing population of caregivers, there is an increasing interest in technological innovations to ease caregiving burden. Web-based interventions among caregivers appear to be focused on certain diagnosis groups (ie, dementia and chronic conditions) [14-20]. Most of the web-based interventions evaluated provide information, education, and peer or professional support. A review of web-based interventions to improve caregiver health and general caregiving outcomes found significant reductions in stress or distress because of technological interventions. However, results from the evaluations of such interventions were mixed owing to small sample sizes or weak study designs [15].

Although many studies have examined the efficacy of web-based interventions, our knowledge on the role of mobile apps designed for caregiver burden is relatively limited [21-23]. Ghahramani and Wang [21] investigated predictors of caregivers' willingness to adopt caregiving-related mobile apps. They found that caregivers' capabilities and skills in using an app, an app's effectiveness in responding to the caregivers' needs, and the degree of control caregivers had over their responsibilities were factors that affected the willingness of caregivers to adopt caregiving-related mobile apps. As a care recipient's health was perceived to be more severe, caregivers reported being more likely to use an app. In addition, as the threat of unexpected health changes became more likely, caregivers reported perceiving an app as a more efficient tool [21].

Our study aims to evaluate the effectiveness of a mobile app in decreasing caregiver burden and increasing perceived support and well-being among employees of a large employer. The authors of the paper were employed by the large employer during the study period and dissemination of study results. The mobile app works by connecting the unpaid, primary caregiver to his or her personal social support group (ie, friends, family members, and community services) in a convenient way, which helps to lower the barrier in requesting and offering support. In addition, the app provides the caregiver with access to multiple resources and tools. Specifically, this study contributes to the literature in three ways. First, we evaluate the impact of a mobile app specifically designed for unpaid caregivers—a topic that has not been widely studied but hypothesized to reduce burden [21-23]. The app's specific function is to make it easier to coordinate help for the caregiver. Through this focus, we can better understand how the app positively affects the caregiver and decreases the burden. Second, we evaluate outcomes that were not typically reported in previous studies, such as the impact of caregiving on perceived work productivity and time use. Finally, we provide suggestive evidence on how a public health crisis, such as COVID-19, affects caregiver burden and how the app is able to decrease caregiver burden despite unforeseen factors and limitations.

Caregiver App

The ianacare app was selected through a competition among technology start-ups as a part of a large employer's strategy to use technology that aims to improve health and well-being across its employees. Ianacare was 1 of 5 companies selected from a pool of 126 applicants. Applications from start-ups were judged based on the technical implementation feasibility and merit of the proposed solution.

The app was launched in 2019. Being family caregivers themselves, the founders observed that help was often not exchanged between the caregiver and their support groups owing to the feeling of intense burden on behalf of the caregiver. The ianacare mobile app was designed to leverage technology and act as an effective buffer to organize and mobilize social networks around caregivers; the *iana* of ianacare stands for *I Am Not Alone*.

The mobile app allows a support team to be mobilized when the caregiver makes needs known, such as coordinating physician visits, dropping off groceries, and medication delivery. This mobilization allows the burden of care to be distributed among more people and makes it easier to coordinate schedules, while providing a platform for sharing emotional support. Once the app is downloaded, the caregiver is asked to create his or her team of supporters by inviting them to download the app and join the caregiver's support team. When there is help needed for a specific caregiving task, the caregiver can click on the appropriate icon of available options (ie, errands, check-in visits, meals, rest or breaks, pet care, childcare, rides, and other events). After choosing one of the help options, the caregiver can specify the person or persons that he or she wants help from. They also specify the locations, dates, and time on a calendar visible by the caregiver's support group, as well as any other specifications. Once help is requested, a notification is sent, and the caregiver is notified when someone accepts the request. Both caregivers and care recipients can request and accept help, post updates, and invite or remove supporters. Caregivers and support group members can also post updates and pictures keeping each other engaged in the care recipient's situation confidentially on the app. One of the advantages of ianacare is the built-in choice for requesting different types of help that can be sent to a caregiver's team of supporters all at once, eliminating the need to manage one-on-one conversations. Team members can see who helped on which tasks, which can lead to a more even distribution of caregiving tasks among the team. Our hypothesis is that caregivers using the app will be more likely to ask and receive help from their support system, reducing the burden associated with having to tailor each help request as individual conversations whenever help is needed. When requesting help becomes easier, caregivers will be more likely to seek it. This leads to improvements in other domains such as work productivity, well-being, leisure time, and a general feeling of support.

This study attempts to fill the gap in the literature on this topic by evaluating the impact of the app on a sample of unpaid, employed caregivers using a pre- and postsurvey design that asks questions on both objective and subjective outcomes of caregiving. We focus on three primary outcomes for the primary caregiver: (1) support system, as defined by the availability of helpers to the caregiver, ease of assistance with caregiving tasks, and feelings of being supported as a caregiver; (2) time use and productivity outcomes, defined as hours spent on caregiving tasks, caregiving tasks that affect work hours, perceived productivity at work, and perceived impact of caregiving on caregiver's work; and (3) well-being, defined as the frequency of feeling overwhelmed by caregiving tasks and the impact of caregiving on the caregiver's perceived health [4,24].

Methods

Survey Implementation

We used a pre- and postsurvey design to measure the effectiveness of the app. The app was offered from March 30, 2020, to April 13, 2020, to employees of a large national employer. The employer selected ianacare based on their interest in helping alleviate caregiver burden among their employees. The survey was developed by a study team including the authors of this paper as well as internal company survey data collection experts.

Employees were encouraged to download and use the app for a 6-week period on the announcement of the employer's intranet. Inclusion criteria included: being an unpaid primary caregiver for at least one person, having a smartphone or tablet, and being with the same employer during the 6 weeks between the first and second surveys. Eligible participants completed a web-based 10-minute survey at the time of enrollment. If they were the primary caregiver for more than one person, we asked participants to answer questions only for the person for whom they spent the most time providing care. After 6 weeks of app use, participants were invited by email to take the second web-based survey, which took approximately 10 minutes to complete. A gift card of US $25 was provided upon the successful completion of both surveys. Results from both surveys remained anonymous and were aggregated without individual names. Survey recruitment materials and the informed consent section emphasized the voluntary nature of participation and the option to end participation at any time without penalty or loss of benefits. Participants wishing to withdraw from the study were able to communicate their requests via email. After withdrawal, the participant would no longer receive any study-related communication and the study team would remove their data. We also excluded survey participants if they were never active on the app.

The pre- and postapp use surveys consisted of questions about participants' demographics (age and gender), socioeconomic status (education and income levels), and caregiving characteristics (person being cared for, where person resides, and types of caregiving tasks). The preapp use survey also provided a baseline assessment of the caregiver's support system and the impact of caregiving on perceived productivity and health. The survey question which the health outcome was generated from does not distinguish between physical and mental health. The postapp use survey included identical questions from the first survey. The postapp use survey also included questions on the mobile app's impact on the caregiving burden of participants, their opinions about existing app features,

and other potential features that could be added to the app in the future. Given that the study timeline overlapped with the emergence of COVID-19 in 2020, the postapp use survey also included questions on the impact of COVID-19 on their caregiving situation.

The surveys included both objective and subjective questions. We labeled questions where the respondent's answer could be determined as right or wrong as objective and those questions where the respondent's answer pertained to their perceptions and feelings as subjective.

Ethical Considerations

An institutional review board review was not needed because data were collected as part of the employer's quality improvement initiative (as opposed to human subject research) in an effort to improve benefits offered to employees. For the full survey questionnaire, please refer to Multimedia Appendix 1.

Survey Questions

Caregiver's Support System

Information pertaining to the caregiver's support system data was based on questions in both the pre- and postapp use surveys about the proportion of caregiving tasks performed by the participant and about the frequency of feeling supported as a caregiver.

Time Spent on Caregiving and Perceived Productivity at Work

All participants were asked identical questions before and after app use on the time spent on caregiving and their perception of how caregiving impacted their productivity at work. Participants were asked to report hours spent per week in the past month on caregiving (eg, food prep, care assistance, coordinating physician visits, and grocery shopping) and hours they needed to take off from work owing to planned caregiving tasks (eg, physician appointments) and unplanned caregiving tasks (eg, medical emergencies). Participants were asked to assess how they felt their caregiving role impacted their productivity or focus at work. In addition, they were asked if they felt caregiving put them at a disadvantage compared with coworkers in terms of work performance and recognition.

Caregiver Well-being

All participants were asked identical questions regarding their overall well-being as caregivers over the past 30 days, including their stress level and perceived caregiver burden, both before and after using the mobile app. Specifically, participants were asked to report the frequency of feeling overwhelmed by caregiving tasks (eg, worrying about the person they were caring for) and how often they felt caregiving negatively impacted their own health in the past 30 days (eg, caregiver missing their own physician appointments or missing their own medicine because they were too busy caring for someone else).

Impact of COVID-19 on Caregiving

As the study period overlapped with the 2020 COVID-19 pandemic, participants were asked to assess the impact of COVID-19 on their caregiving role. A 5-point Likert scale was used to scale responses. Participants were asked questions regarding their productivity in terms of caregiving hours related to COVID-19, including how COVID-19 affected the number of hours spent per week on caregiving, whether COVID-19 increased the amount of time needed to take off work to attend to caregiving tasks, participants' productivity based on whether COVID-19 put them at a greater disadvantage at work owing to being a caregiver, and whether COVID-19 negatively impacted focus at work because of increased worry about caregiving tasks. The overall well-being of the participants was assessed using the question of how COVID-19 impacted stress levels around caregiving tasks.

Outcome Measurement

We evaluated the app's effectiveness in three outcome areas: support system, time use and productivity, and well-being. Outcomes related to a caregiver's support system were assessed with binary indicators: (1) whether the respondent reported doing more than half of the caregiving by himself or herself during the past 30 days, (2) whether the respondent reported having no one supporting him or her in caregiving tasks during the past 30 days, and (3) whether the respondent reported *never* or *almost never* to feeling supported by his or her social network during the past 30 days. Outcomes of caregiving time use and productivity were assessed using binary indicators, including: (1) whether the respondent reported spending more than 30 hours weekly on caregiving tasks in the past 30 days, (2) whether the respondent reported that he or she needed to take any time off from work owing to planned caregiving tasks in the past 30 days, (3) whether the respondent reported he or she unexpectedly needed to take any time off from work owing to caregiving tasks, (4) whether the respondent reported his or her work productivity and focus to be negatively affected by caregiving, and (5) whether the respondent reported caregiving put him or her at a disadvantage compared with coworkers in terms of work performance and recognition. Outcomes of well-being were assessed using binary indicators for the probability of reporting (1) fairly or very often feeling overwhelmed by caregiving tasks in the past 30 days and (2) fairly or very often feeling negative health effects related to caregiving.

Statistical Analyses

To measure the association between app use and caregiving outcomes, we compared pre- and postsurvey responses and reported within-participant changes in outcomes using linear probability models. In addition, we compared the distribution of respondents' full scale of answers without turning them into binary variables. For the distributional analyses, we present boxplots to show the distribution of the answers to the questions assessing outcomes of support system, time use and productivity, and self-reported well-being. We also reported the average within-participant change in the full scale of answers before and after app use. Power analysis was not conducted.

Results

Overview

The final sample included 176 individuals who responded to the pre- and postapp surveys (Figure 1). Our sample was predominantly female (167/176, 94.9%), with an average age of 47 (SD 9.71) years. More than half of the participants were White, college educated, and had at least US $75,000 in household income. Most participants reported taking care of their parents, living with the person they were caring for, and providing daily living and administrative assistance (ie, food prep, housekeeping, managing insurance, coordinating physician visits, and financial management; Table 1).

Figure 1. Sample construction.

Table 1. Characteristics of individuals in the final sample (N=176).

Predictors	Values, mean (SD)	Value, n (%)
Age (years)	46.99 (9.71)	176 (100)
Female	0.95 (0.22)	167 (94.9)
Married	0.50 (0.50)	88 (50)
Race		
White	0.55 (0.50)	96 (54.5)
Black	0.24 (0.43)	42 (23.9)
Hispanic or Latino	0.10 (0.30)	18 (10.2)
Asian	0.04 (0.20)	7 (4)
American Indian or Alaska native	0.01 (0.08)	2 (1.1)
Native Hawaiian or other Pacific islander	0.01 (0.08)	2 (1.1)
Other	0.01 (0.11)	2 (1.1)
Prefer not to respond	0.04 (0.20)	7 (4)
Education		
High school	0.04 (0.20)	7 (4)
Some college	0.20 (0.40)	35 (19.9)
Associate degree	0.22 (0.42)	39 (22.2)
Bachelor's degree	0.23 (0.42)	40 (22.7)
Graduate degree or higher	0.30 (0.46)	53 (30.1)
Prefer not to respond	0.01 (0.08)	2 (1.1)
Categories of annual household income (US $)		
<50,000	0.18 (0.38)	32 (18.2)
50,000-75,000	0.07 (0.25)	12 (6.8)
75,000-100,000	0.17 (0.38)	30 (17)
100,000-150,000	0.24 (0.43)	42 (23.9)
>150,000	0.13 (0.33)	23 (13.1)
Prefer not to say	0.21 (0.40)	37 (21)
Number of people in household	3.29 (1.23)	176 (100)
Care recipient		
Parent	0.48 (0.50)	85 (48.3)
Child under 18	0.32 (0.47)	56 (31.8)
Other[a]	0.20 (0.40)	35 (19.9)
Care recipient lives		
With the care recipient	0.63 (0.49)	111 (63.1)
Alone	0.17 (0.38)	30 (17)
Other	0.20 (0.40)	35 (19.9)
Type of tasks done by caregiver		
Daily living assistance	0.69 (0.46)	121 (68.8)
Care assistance	0.37 (0.48)	65 (36.9)
Administration	0.64 (0.48)	113 (64.2)
Other[b]	0.09 (0.28)	16 (9.1)

[a]Includes spouses, children aged >18 years, grandchildren, grandparents, partners, friends, housemates, coworkers, and neighbors.

[b]Includes web-based school assistance, home schooling, nurturing, assisting with socialization and emotional support.

We observed significant differences in 2 of the support system outcomes. Specifically, after using the app, the probability of caregivers reporting *doing more than half of all the caregiving tasks by themselves* decreased by 9.1% points (SE 0.036; *P*=.01). The probability of caregivers reporting *no one helping them in caregiving tasks* also decreased by 7.9% points (SE 0.035; *P*=.02). There was no evidence that the app had any impact on reducing the caregiving load for caregivers who reported doing all caregiving by themselves. We also did not observe any meaningful difference in perceived support by one's support group (Table 2).

Table 2. Association of app use with caregiving outcomes (N=176).

	Support system				Time use and perceived productivity					Perceived health and well-being	
	Does all caregiving (objective)	Does more than half (objective)	No one helps (objective)	No network support (subjective)	More than 30 hours spent weekly (objective)	No planned time taken (objective)	No unplanned time taken (objective)	Negative effect on productivity (subjective)	Feeling disadvantaged at work (subjective)	Often feeling negative health effects (subjective)	Often overwhelmed (subjective)
App effect percentage point change associated with app use (SE)[a]	−0.0227 (0.0351)	−0.0909 (0.0354)	−0.0795 (0.0347)	0.0227 (0.0369)	−0.0909 (0.0398)	0.125 (0.0408)	0.125 (0.0439)	−0.0284 (0.0374)	0.017 (0.0296)	−0.0682 (0.0375)	−0.125 (0.0416)
Observations	352	352	352	352	352	352	352	352	352	352	352
R^2	0.002	0.036	0.029	0.002	0.029	0.051	0.044	0.003	0.002	0.019	0.049
P value	.52	.01	.02	.54	.02	.003	.005	.45	.57	.07	.003

[a]Robust SE.

In terms of time use and perceived productivity outcomes, we observed several improvements. After 6 weeks of app use, both the probability of reporting *taking no time off from work due to planned caregiving tasks* and the probability of reporting *taking no time off of work due to attend unscheduled caregiving tasks* increased by 12.5% points (SE 0.04; *P*=.003 and *P*=.005, respectively). We also observed that after 6 weeks of app use, there was a significant decrease in the probability of reporting *spending more than 30 hours weekly on caregiving tasks* by 9.1% points (SE 0.04; *P*=.02; Table 2).

The probability of reporting *often feeling negative health effects due to caregivin*g decreased by 6.8% points, although the estimate was only significant at a 10% significance level (SE 0.04; *P*=.07). After 6 weeks of app use, the probability of reporting *often feeling overwhelmed by caregiving tasks* decreased by 12.5% points (SE 0.04; *P*=.003; Table 2)**.**

To summarize, we observed improvements in 7 out of 11 caregiver outcomes. App use was significantly associated with decreasing the amount of caregiving tasks that fell on the primary caregiver and increasing the likelihood of at least one person helping him or her. Use of the app was also associated with improvements in the time management of the primary caregiver. Significantly less time was taken off work to attend to caregiving, and the likelihood of spending <30 hours weekly on caregiving was significantly decreased. There is also suggestive evidence that the app may be associated with decreasing feelings of being overwhelmed and improving the perceived impact of caregiving on caregivers' health.

We observed that the responses to the number of people supporting the primary caregiver improved significantly (*P*=.006; Figure 2) with app use. We observed significant changes in responses regarding the number of hours spent on caregiving tasks per week (*P*=.04) and the number of hours taken off from work to attend to unscheduled caregiving duties (*P*=.03; Figure 3). These significant changes pertain to responses shifting toward decreasing hours of caregiving reported. We observed that responses to work outcomes significantly became worse; we observed declines in perceived work productivity or focus (*P*=.004), and feelings of being disadvantaged at work owing to caregiving increased (*P*=.06; Figure 3). However, there was a significant improvement in the perceived impact of caregiving on participants' health (*P*=.02; Figure4). Finally, we observed a positive statistically significant reduction in participants' feelings of being overwhelmed about caregiving after using the app (*P*=.004; Figure 4).

Figure 2. Distribution of responses to support system questions.

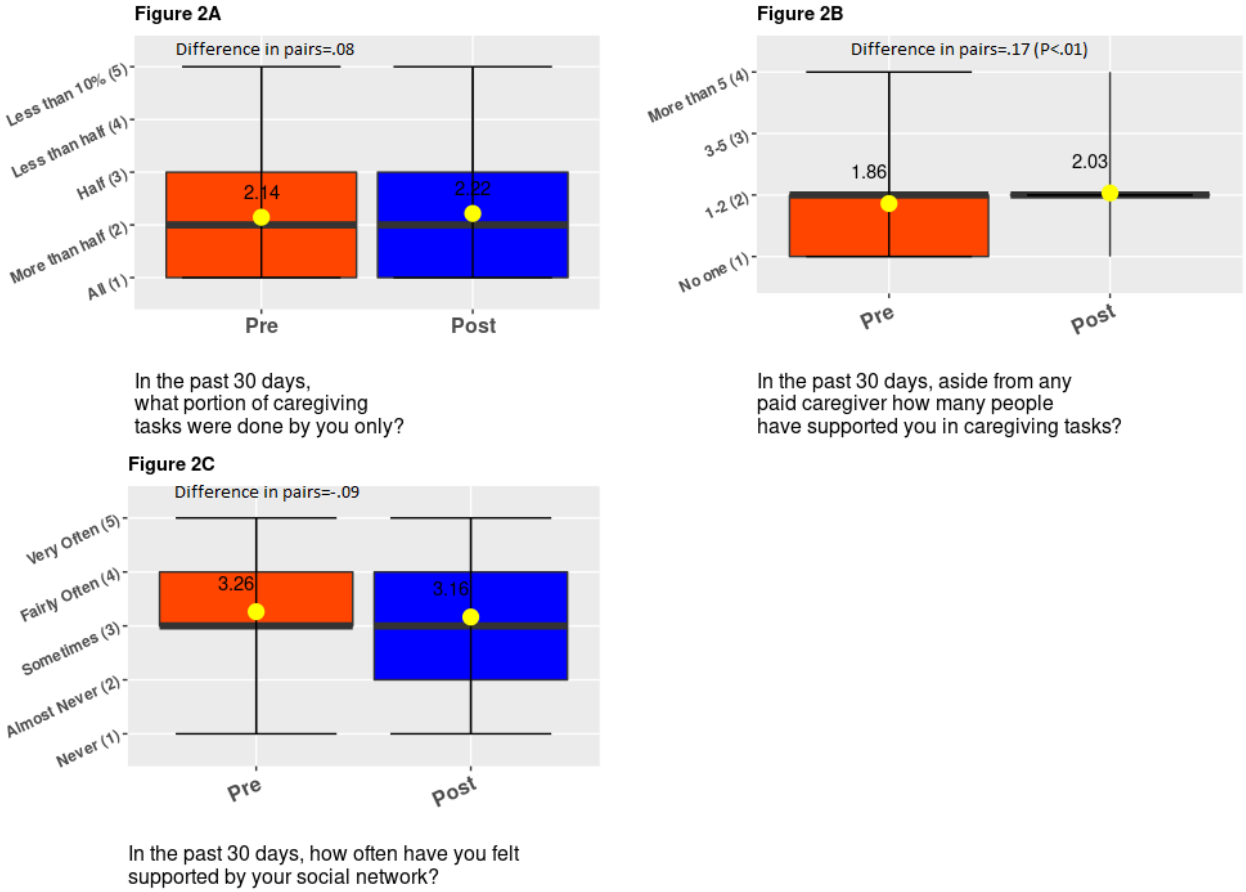


Figure 3. Distribution of responses to time use and perceived productivity questions.

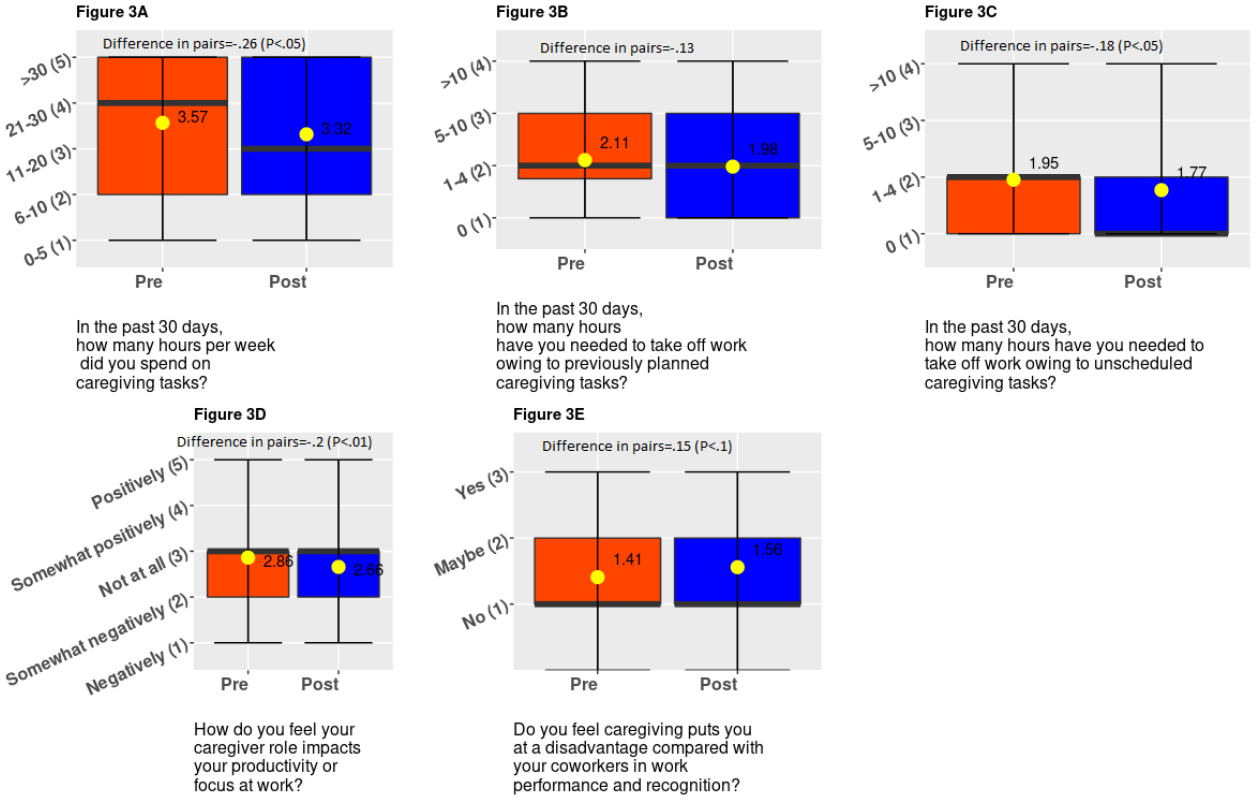
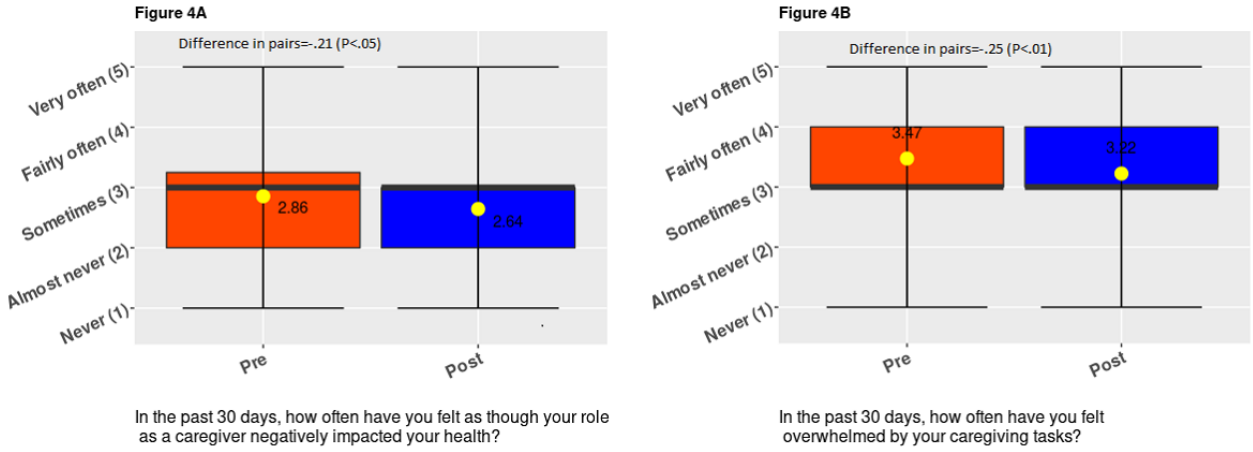


Figure 4. Distribution of responses to perceived health and well-being questions.



Impact of COVID-19 Pandemic

A total of 35.2% (62/176) of the participants said that COVID-19 significantly increased their time spent on caregiving tasks (Table 3). Approximately 67% (118/176) said COVID-19 increased their stress attributed to caregiving, and 40.9% (72/176) reported that COVID-19 negatively impacted their focus at work. We also observed that most participants reported that COVID-19 prevented them from asking others to help, although we observed a significant increase both in the probability of acquiring help from at least one person and the number of people helping in caregiving. We interpret this as suggestive evidence that the app was useful in enabling caregivers to seek more help from others despite COVID-19 making it more difficult. Most participants reported that COVID-19 did not increase the need to take time off from work to attend to caregiving.

Table 3. Responses to questions about COVID-19's impact (N=176).

Factors to assess COVID-19 impact	Significantly increased, n (%)	Somewhat increased, n (%)	No impact, n (%)	Somewhat decreased, n (%)	Significantly decreased, n (%)	Missing, n (%)
Caregiving time spent	62 (35.2)	35 (19.9)	34 (19.3)	10 (5.7)	15 (8.5)	20 (11.4)
Caregiving stress	52 (29.6)	66 (37.5)	29 (16.5)	8 (4.6)	1 (0.6)	20 (11.4)
Increased the need to take time off from work to attend caregiving	21 (11.9)	25 (14.2)	45 (25.6)	19 (10.8)	43 (24.4)	23 (13.1)
Put me at a greater disadvantage as a caregiver	34 (19.3)	31 (17.6)	37 (21)	13 (7.4)	38 (21.6)	23 (13.1)
Negatively impacted my focus at work as a caregiver	24 (13.6)	48 (27.3)	34 (19.3)	15 (8.5)	32 (18.2)	23 (13.1)
Prevented me from asking help from others	67 (38.1)	26 (14.8)	28 (15.9)	6 (3.4)	23 (13.1)	26 (14.8)

Perception of the App

Overall, most respondents had neutral opinions about the apps (Table 4). However, they were more likely to have favorable opinions about decreasing their caregiver burden. For example, 38.1% (67/176) reported that the app made asking others for help with caregiving easier, and 30.1% (53/176) reported that the app enabled them to ask for help for things that they would not have otherwise asked. Of the 176 participants, 74 (42%) reported that the app made them feel more supported in their role as caregivers.

Table 4. Responses to questions about perceived value of the app (N=176).

How the app helped	Strongly agree, n (%)	Somewhat agree, n (%)	Neither agree or disagree, n (%)	Somewhat disagree, n (%)	Strongly disagree, n (%)	Missing, n (%)
Asking help for the things I would not otherwise ask	22 (12.5)	31 (17.6)	46 (26.1.7)	22 (12.5)	29 (16.5)	26 (14.8)
Made asking for help easier	27 (15.3)	40 (22.7)	44 (25)	16 (9.1)	23 (13.1)	26 (14.8)
Supported in my role as a caregiver	37 (21)	37 (21)	40 (22.7)	15 (8.5)	21 (11.9)	26 (14.8)
Made caregiving less stressful	26 (14.8)	30 (17)	50 (28.4)	20 (11.4)	24 (13.6)	26 (14.8)

In the second survey, we also asked respondents how caregiving negatively affected their own health. We found that approximately 17% (30/176) said that it affected their physical health and decreased their quality of life. Approximately 26.1% (46/176) and 23.9% (42/176) reported that caregiving worsened their health by affecting their sleep and work-life balance, and 44.9% (79/176) reported that caregiving increased their stress and anxiety (Table 5).

Table 5. Responses to questions about the caregiving on caregiver's health.

How has caregiving affected your own health	Those who answered "Yes," n (%)
Increased stress or anxiety	79 (44.9)
Affected physical health negatively	29 (16.5)
Affected work-life balance negatively	42 (23.9)
Affected sleep schedule negatively	46 (26.1)
Decreased quality of life	29 (16.5)

Discussion

Principal Findings

Our results suggest that mobile app use was associated with improvements in the objective outcomes of caregiving, such as decreasing the likelihood of doing more than half of the caregiving by themselves and the likelihood of the caregiver having no one to assist them. We note that the app works mainly by improving a caregiver's support system by more conveniently connecting the caregiver to people who can assist with caregiving and making the task of asking for help easier. Therefore, we expected to see relatively more changes in outcomes related to the caregiver's support system than the other outcomes in this study. We were able to detect significant improvements in outcomes that required objective assessment from the participant, such as the portion of caregiving done on their own versus with assistance. However, outcomes that had subjective measures, such as feeling supported by one's support group or perceived work productivity, did not yield any meaningful changes.

We recognize that the COVID-19 pandemic brought substantial changes to people's daily lives and working conditions during the time between the 2 surveys. It is possible that some of the positive and negative outcomes resulted from COVID-19 stay-at-home orders. For example, the increase in the number of people supporting the caregiver could be attributable to more family members working from home or fewer work responsibilities because of a possible job loss, and hence**,** the ability to better support the primary caregiver. Our sample excluded anyone who lost their jobs between surveys; therefore, the results were not affected by a change in the participants' employment status. The decrease in the number of hours taken off from work to attend to caregiving could be attributed to changes in the participants' working conditions, not necessarily to the app use. It is also plausible that because of shelter-in-place restrictions, caregivers may not have been able to use the app at its maximum capacity.

The app required users to recruit help from their social networks. Therefore, caregivers without an existing group of potential helpers would not experience the full benefit of the app. The decline in work-related outcomes could have been associated with the effects of the COVID-19 pandemic because the first survey was conducted before the effects of the COVID-19 pandemic were publicly observed, and the second survey occurred during a peak of the COVID-19 pandemic. It is plausible to attribute the decline in perceived work productivity and the increase in feeling disadvantaged at work to the negative impact of the COVID-19 pandemic.

Limitations

We note that our study sample differs from the samples in many of the studies reviewed in the study by Ploeg et al [15] as well as from most of the other interventions examined in the caregiving domain. First, our study did not restrict the caregiving sample to a particular group of caregivers who provide care to people with specific conditions (ie, heart transplant, patients with cancer, and persons living with dementia) and instead included many types of care recipients. Therefore, we cannot reveal anything to the specific condition of either the care recipient or the caregiver because our sample size was not sufficient for investigating the heterogeneous effects of the app. Second, our sample largely consisted of employees of a large employer, most of whom were at least college educated and had a higher income than the average person. Therefore, our results may not be applicable to the general population of caregivers. It is conceivable that our sample included caregivers with more digital literacy and willingness to use the mobile app compared with the general population, allowing them to reap more benefits. Therefore, our results should be taken with findings from existing literature documenting the link between digital literacy and caregivers' use of mobile interventions to facilitate caregiving [25,26]. Another limitation of this study is that it only focused on the effect of the app on negative outcomes of caregiving because subjects were not asked anything about their positive experiences attributed to their caregiver role. It is plausible that app use may have enhanced the positive impacts of caregiving by reducing isolation and burden associated with caregiving, leading to more positive experiences attributed to caregivers taking care of their loved ones [11]. Finally, as the intervention period coincided with the emergence of the COVID-19 pandemic, resulting in shelter-in-place policies and other changes brought to people's daily lives, we must be cautious about interpreting the pandemic's interference on our findings.

Conclusions and Implications

We found that using a mobile app that facilitated coordination among caregivers resulted in improvements in the time spent on caregiving, perceived health, and perceived well-being of the primary caregiver. This suggests that mobile interventions to facilitate caregiving can be a useful solution for some caregivers. We believe this study adds further support for devoting resources to mobile apps designed to help caregivers. The implications of our findings are especially relevant as the current pandemic not only increased caregiver burden but also allowed a larger proportion of the population to experience the benefits of technology. Future studies are needed to examine the opportunities that mobile technologies can provide to caregivers with no or limited social support groups and how such technologies can be enhanced to obtain the most help for caregivers in the midst of exceeding demand for unpaid caregiving support.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Caregiving surveys.
[DOCX File , 32 KB - jmir_v24i4e28504_app1.docx]

References

1. AARP and National Alliance for Caregiving. Caregiving in the United States 2020. AARP, Washington, DC. 2020. URL: https://www.aarp.org/content/ [accessed 2022-02-11]
2. Kim Y, Schulz R. Family caregivers' strains: comparative analysis of cancer caregiving with dementia, diabetes, and frail elderly caregiving. J Aging Health 2008 Aug;20(5):483-503. [doi: 10.1177/0898264308317533] [Medline: 18420838]
3. Adelman RD, Tmanova LL, Delgado D, Dion S, Lachs MS. Caregiver burden: a clinical review. J Am Med Assoc 2014 Mar 12;311(10):1052-1060. [doi: 10.1001/jama.2014.304] [Medline: 24618967]
4. del-Pino-Casado R, Frías-Osuna A, Palomino-Moral PA, Ruzafa-Martínez M, Ramos-Morcillo AJ. Social support and subjective burden in caregivers of adults and older adults: a meta-analysis. PLoS One 2018;13(1):e0189874 [FREE Full text] [doi: 10.1371/journal.pone.0189874] [Medline: 29293522]
5. Flyckt L, Fatouros-Bergman H, Koernig T. Determinants of subjective and objective burden of informal caregiving of patients with psychotic disorders. Int J Soc Psychiatry 2015 Nov 01;61(7):684-692 [FREE Full text] [doi: 10.1177/0020764015573088] [Medline: 25770207]
6. Fekete C, Tough H, Siegrist J, Brinkhof MW. Health impact of objective burden, subjective burden and positive aspects of caregiving: an observational study among caregivers in Switzerland. BMJ Open 2017 Dec 22;7(12):e017369 [FREE Full text] [doi: 10.1136/bmjopen-2017-017369] [Medline: 29275339]
7. Lin I, Fee HR, Wu HS. Negative and positive caregiving experiences: a closer look at the intersection of gender and relationships. Family Relations 2012 Apr 01;61(2):343-358 [FREE Full text] [doi: 10.1111/j.1741-3729.2011.00692.x] [Medline: 22544989]
8. Conde-Sala JL, Garre-Olmo J, Turró-Garriga O, Vilalta-Franch J, López-Pousa S. Differential features of burden between spouse and adult-child caregivers of patients with Alzheimer's disease: an exploratory comparative design. Int J Nurs Stud 2010 Oct;47(10):1262-1273. [doi: 10.1016/j.ijnurstu.2010.03.001] [Medline: 20374966]
9. Fudge H, Neufeld A, Harrison MJ. Social networks of women caregivers. Public Health Nurs 1997 Feb;14(1):20-27. [doi: 10.1111/j.1525-1446.1997.tb00406.x] [Medline: 9078846]
10. Poulin M, Brown SL, Dillard AJ, Smith DM. Giving to others and the association between stress and mortality. Am J Public Health 2013 Sep 01;103(9):1649-1655 [FREE Full text] [doi: 10.2105/ajph.2012.300876]
11. Peacock S, Forbes D, Markle-Reid M, Hawranik P, Morgan D, Jansen L, et al. The positive aspects of the caregiving journey with dementia: using a strengths-based perspective to reveal opportunities. J Appl Gerontol 2009 Aug 25;29(5):640-659. [doi: 10.1177/0733464809341471]
12. Cohen CA, Colantonio A, Vernich L. Positive aspects of caregiving: rounding out the caregiver experience. Int J Geriatr Psychiatry 2002 Feb;17(2):184-188. [doi: 10.1002/gps.561] [Medline: 11813283]
13. Raschick M, Ingersoll-Dayton B. The costs and rewards of caregiving among aging spouses and adult children*. Fam Rel 2004 Apr;53(3):317-325. [doi: 10.1111/j.0022-2445.2004.0008.x]
14. Leng M, Zhao Y, Xiao H, Li C, Wang Z. Internet-based supportive interventions for family caregivers of people with dementia: systematic review and meta-analysis. J Med Internet Res 2020 Sep 09;22(9):e19468 [FREE Full text] [doi: 10.2196/19468] [Medline: 32902388]
15. Ploeg J, Markle-Reid M, Valaitis R, McAiney C, Duggleby W, Bartholomew A, et al. Web-based interventions to improve mental health, general caregiving outcomes, and general health for informal caregivers of adults with chronic conditions living in the community: rapid evidence review. J Med Internet Res 2017 Jul 28;19(7):e263 [FREE Full text] [doi: 10.2196/jmir.7564] [Medline: 28754652]

16. Sherifali D, Ali MU, Ploeg J, Markle-Reid M, Valaitis R, Bartholomew A, et al. Impact of internet-based interventions on caregiver mental health: systematic review and meta-analysis. J Med Internet Res 2018 Jul 03;20(7):e10668 [FREE Full text] [doi: 10.2196/10668] [Medline: 29970358]
17. Ploeg J, Ali MU, Markle-Reid M, Valaitis R, Bartholomew A, Fitzpatrick-Lewis D, et al. Caregiver-focused, web-based interventions: systematic review and meta-analysis (Part 2). J Med Internet Res 2018 Oct 26;20(10):e11247 [FREE Full text] [doi: 10.2196/11247] [Medline: 30368439]
18. Boots LM, de Vugt ME, van Knippenberg RJ, Kempen GI, Verhey FR. A systematic review of Internet-based supportive interventions for caregivers of patients with dementia. Int J Geriatr Psychiatry 2014 Apr;29(4):331-344. [doi: 10.1002/gps.4016] [Medline: 23963684]
19. Kaltenbaugh DJ, Klem ML, Hu L, Turi E, Haines AJ, Hagerty LJ. Using web-based interventions to support caregivers of patients with cancer: a systematic review. Oncol Nurs Forum 2015 Mar;42(2):156-164. [doi: 10.1188/15.ONF.156-164] [Medline: 25806882]
20. Hu C, Kung S, Rummans TA, Clark MM, Lapid MI. Reducing caregiver stress with internet-based interventions: a systematic review of open-label and randomized controlled trials. J Am Med Inform Assoc 2015 Apr;22(e1):194-209 [FREE Full text] [doi: 10.1136/amiajnl-2014-002817] [Medline: 25125686]
21. Ghahramani F, Wang J. Intention to adopt mhealth apps among informal caregivers: cross-sectional study. JMIR Mhealth Uhealth 2021 Mar 17;9(3):e24755 [FREE Full text] [doi: 10.2196/24755] [Medline: 33729166]
22. Wittenberg E, Xu J, Goldsmith J, Mendoza Y. Caregiver communication about cancer: development of a mhealth resource to support family caregiver communication burden. Psychooncology 2019 Feb;28(2):365-371. [doi: 10.1002/pon.4950] [Medline: 30471236]
23. Grossman MR, Zak DK, Zelinski EM. Mobile apps for caregivers of older adults: quantitative content analysis. JMIR Mhealth Uhealth 2018 Jul 30;6(7):e162 [FREE Full text] [doi: 10.2196/mhealth.9345] [Medline: 30061093]
24. Diener E. Subjective well-being. In: The Science of Well-Being. Dordrecht: Springer; Jan 01, 2009.
25. Chiu TM, Eysenbach G. Theorizing the health service usage behavior of family caregivers: a qualitative study of an internet-based intervention. Int J Med Inform 2011 Nov;80(11):754-764. [doi: 10.1016/j.ijmedinf.2011.08.010] [Medline: 21958550]
26. Shaffer KM, Chow PI, Cohn WF, Ingersoll KS, Ritterband LM. Informal caregivers' use of internet-based health resources: an analysis of the health information national trends survey. JMIR Aging 2018 Dec 18;1(2):e11051 [FREE Full text] [doi: 10.2196/11051] [Medline: 31518244]

Edited by R Kukafka; submitted 04.03.21; peer-reviewed by W Tang, S Peacock; comments to author 13.04.21; revised version received 30.06.21; accepted 07.01.22; published 11.04.22.

Please cite as:
Ozluk P, Cobb R, Hoots A, Sylwestrzak M
Association Between Mobile App Use and Caregivers' Support System, Time Spent on Caregiving, and Perceived Well-being: Survey Study From a Large Employer
J Med Internet Res 2022;24(4):e28504
URL: https://www.jmir.org/2022/4/e28504
doi:10.2196/28504
PMID:35404266

Original Paper

Single-Session, Internet-Based Cognitive Behavioral Therapy to Improve Parenting Skills to Help Children Cope With Anxiety During the COVID-19 Pandemic: Feasibility Study

Tarja Korpilahti-Leino[1,2*], MA; Terhi Luntamo[1,2,3*], MD, PhD; Terja Ristkari[1,2], RN, MNSc; Susanna Hinkka-Yli-Salomäki[1,2], PhLic; Laura Pulkki-Råback[1,2], PhD; Otto Waris[1,2], PhD; Hanna-Maria Matinolli[1,2], PhD; Atte Sinokki[1,2], BEng; Yuko Mori[1,2], MSc; Mami Fukaya[4], MA; Yuko Yamada[5], MA; Andre Sourander[1,2,3], MD, PhD

[1]Research Centre for Child Psychiatry, Department of Clinical Medicine, Faculty of Medicine, University of Turku, Turku, Finland

[2]INVEST Child Psychiatry, INVEST Research Flagship Center, Department of Clinical Medicine, Faculty of Medicine, University of Turku, Turku, Finland

[3]Department of Child Psychiatry, Turku University Hospital, Turku, Finland

[4]Department of Psychology and Human Developmental Sciences, Graduate School of Education and Human Development, Nagoya University, Nagoya, Japan

[5]Section of Medical Care and Consultation, Nagoya City Central Care Center for Disabled Children, Nagoya, Japan

[*]these authors contributed equally

Corresponding Author:
Andre Sourander, MD, PhD
Research Centre for Child Psychiatry
Department of Clinical Medicine, Faculty of Medicine
University of Turku
Lemminkäisenkatu 3
Turku, 20520
Finland
Phone: 358 50 365 3447
Email: andre.sourander@utu.fi

Abstract

Background: The COVID-19 pandemic has had a major impact on families' daily routines and psychosocial well-being, and technology has played a key role in providing socially distanced health care services.

Objective: The first objective of this paper was to describe the content and delivery of a single-session, internet-based cognitive behavioral therapy (iCBT) intervention, which has been developed to help parents cope with children's anxiety and manage daily situations with their children. The second objective was to report user adherence and satisfaction among the first participants who completed the intervention.

Methods: The Let's Cope Together intervention has been developed by our research group. It combines evidence-based CBT elements, such as psychoeducation and skills to manage anxiety, with parent training programs that strengthen how parents interact with their child and handle daily situations. A pre-post design was used to examine user satisfaction and the skills the parents learned. Participants were recruited using advertisements, media activity, day care centers, and schools and asked about background characteristics, emotional symptoms, and parenting practices before they underwent the iCBT. After they completed the 7 themes, they were asked what new parenting skills they had learned from the iCBT and how satisfied they were with the program.

Results: Of the 602 participants who filled in the baseline survey, 196 (32.6%) completed the program's 7 themes, and 189 (31.4%) completed the postintervention survey. Most (138/189, 73.0%) of the participants who completed the postintervention survey were satisfied with the program and had learned skills that eased both their anxiety (141/189, 74.6%) and their children's anxiety (157/189, 83.1%). The majority (157/189, 83.1%) reported that they learned how to organize their daily routines better, and just over one-half (100/189, 53.0%) reported that the program improved how they planned each day with their children.

Conclusions: The single-session iCBT helped parents to face the psychological demands of the COVID-19 pandemic. Future studies should determine how the participation rate and adherence can be optimized in digital, universal interventions. This will help to determine what kinds of programs should be developed, including their content and delivery.

(J Med Internet Res 2022;24(4):e26438) doi:10.2196/26438

KEYWORDS

adolescent; anxiety; child; cognitive behavioral therapy; coping; COVID-19; Internet; mental health; parents; web-based

Introduction

There are considerable concerns about the possible short-term and long-term impact that the COVID-19 pandemic will have on the mental health of children and adolescents [1]. Families have faced substantial concerns and changes to their daily routines and social infrastructures during this global health emergency. These include worrying information and experiences, lockdowns and social distancing, school closures, and employment and economic uncertainties. Evidence suggests that children and adolescents have been experiencing feelings of fear, worry, sadness, anger, and loneliness, in addition to guilt about possibly spreading the virus [2]. Studies have shown that the pandemic has had a substantial impact on the well-being of parents and their children [3], and children and adolescents have reported a high prevalence of anxiety and depression [4]. Deteriorating mental health has been reported by adolescents and young adults in both clinical and community samples [5]. As parents play a significant role in how the crisis affects their children's mental well-being, there is an increased, and urgent, need to provide help for parents. They need this so that they can communicate effectively with their children, ease their children's anxiety, and maintain consistent routines that support the mental health of the whole family [1,6,7].

There is strong evidence that parental factors, such as warmth and modeling, influence children's well-being, by decreasing anxiety [8]. According to a meta-analysis by Yap et al [9], preventive parenting interventions are important tools in reducing child internal problems and may have long-term effects on the child's well-being. The COVID-19 crisis has challenged health care professionals to develop, and deliver, low-threshold interventions that reduce the adverse effects of the pandemic on the mental health of children. The social distancing restrictions that have been used to reduce the spread of the pandemic have made digital mental health interventions more acceptable [10].

Previous studies have shown that digitally delivered parent training, based on cognitive behavioral theory (CBT), has been effective in treating childhood anxiety [11] and disruptive behavior [12-14]. Randomized controlled trials of traditional face-to–face interventions have shown that brief, or single, CBT sessions have been effective in treating childhood anxiety [15] and posttraumatic stress [16]. A single-session, individually tailored, internet-based parenting intervention also improved self-reported parenting factors that are known to influence the development of depression and anxiety in adolescents [17]. These CBT sessions could be a suitable way to deliver basic methods that help parents to cope with anxiety and develop the skills they need during the COVID-19 pandemic. The great advantage of internet sessions is that they are remotely accessible and can be used anytime and anywhere.

There are already numerous online sources that provide psychoeducation for families on how to ease stress and anxiety during the pandemic, including resources provided by the World Health Organization [18] and UNICEF [19]. However, we are not aware of any research on universally delivered internet-based CBT (iCBT) interventions that focus on improving parental skills to help children to cope with anxiety during the COVID-19 pandemic. That is why our research group has developed a single-session iCBT for parents. This new initiative is based on our considerable expertise in developing, evaluating, and implementing evidence-based, digital interventions that provide parental training for children with behavioral problems and reduce children's anxiety [12-14]. The aim of this paper was to describe the content and delivery of the Let's Cope Together intervention and report background characteristics, user adherence, and satisfaction levels of the first participants.

Methods

Description of the Intervention

Let's Cope Together is a universal, single-session iCBT that is intended for all parents with children and adolescents under the age of 18 years in Finland who have access to a digital device with an internet connection. The aim of the intervention is to use this simple, digitally delivered program to improve parental skills, so that they can help their children to cope with any anxiety during the COVID-19 pandemic. The intervention, which is based on CBT and the principles of positive parenting, consists of various online materials (Figure 1).

The intervention was developed in March 2020 and April 2020 by a multidisciplinary team at the University of Turku, Finland. The team included CBT therapists, child psychiatrists, information technology personnel, and experienced family coaches with special training in delivering digital interventions. The content of the program has been specifically tailored to deal with the current COVID-19 pandemic. It includes evidence-based CBT elements that can be used for any kind of anxiety, such as psychoeducation, cognitive restructuring, and relaxation techniques, as well as strategies specifically targeting COVID-19–related anxiety (Figure 2). Parents who go through the program learn how to manage anxiety, by using techniques such as deep breathing. They also learn mental practice techniques that strengthen positive thinking and encourage children to display positive behavior. Additional elements from our existing CBT-based parent training programs have also been included, which strengthen parent-child interactions, manage daily transitions, and help parents plan their day with their child. Table 1 provides an overview of the iCBT content.

The program contains 7 themes that the parents can work through at their own pace. They can complete the intervention in one go or by logging in and out several times. Distinct themes are used to clarify the content, and each theme is comprised of 4 to 5 pages. Our aim is to motivate parents by presenting the

psychoeducation and skill training content of the program in a variety of ways. The total program is comprised of 29 pages of text, 6 videos, 2 audio clips, and 5 manga-style cartoons. The text components are short, simple, and easy to read and are supported by the manga cartoons, which illustrate the key points of each theme (Figure 3). The videos have been included to show how the skills work in practice. There are also links to additional information on each topic.

The intervention is delivered via the internet using a secure TLS-encrypted connection. The website is located on a server maintained by the University of Turku and complies with the university's information security requirements and data protection regulations.

The content of the program was planned by our multidisciplinary team, which previously developed and studied an iCBT for childhood anxiety and a parent training program for behavioral problems. In the midst of the rapidly worsening pandemic in spring 2020, we needed to decide how much information, and what kind of content, a single-session intervention should contain. In addition, we had to take account of the stressful situations that the families who would use the program were facing. The team chose to focus on providing psychoeducation and basic coping skills needed to manage anxiety based on CBT theory and practice. This program also builds on our previous, targeted parent training intervention [13,14]. Our existing experience in this area enabled us to choose the parenting skills that would be most effective in helping *families* to cope with everyday life during the ongoing pandemic. Because the Let's Cope Together program is an unguided digital intervention, we only chose those components that do not require professional guidance. For example, we excluded the reward system featured in our previous internet-based intervention. The core challenge was to make the content as concise as possible. In addition, the content was primarily developed for younger school-aged children, and it was not possible to adjust the content so that it was optimized for all age groups, such as adolescents.

Figure 1. Example of the online material, in which positive mental imagery helps children to manage anxiety [20].

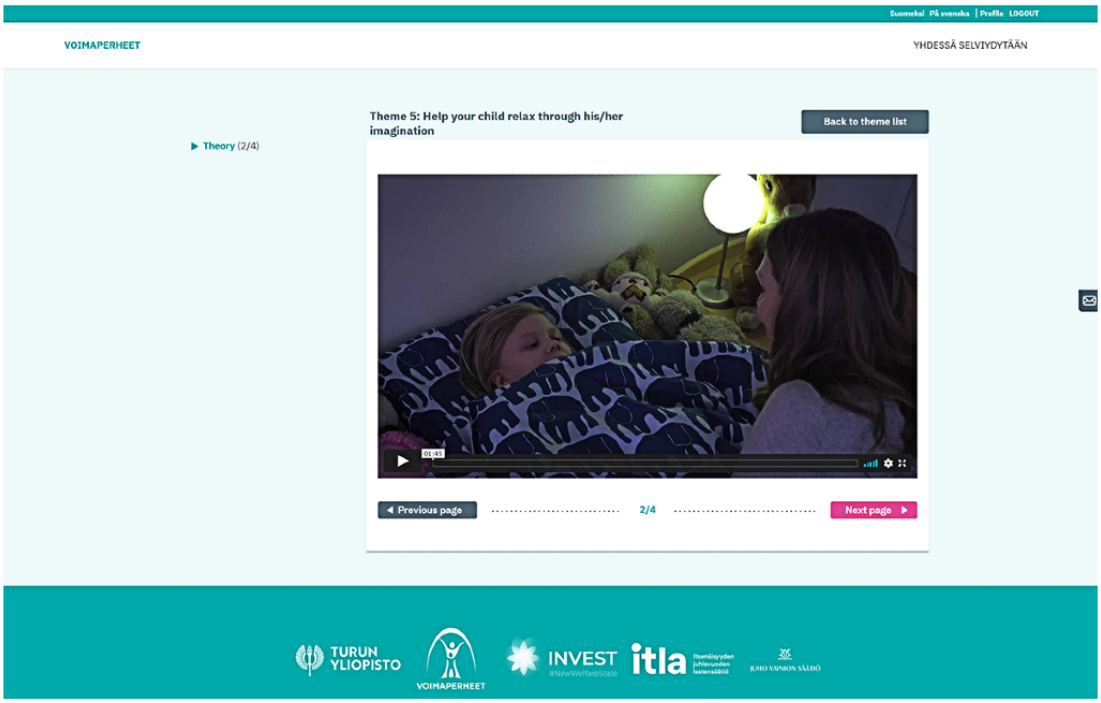
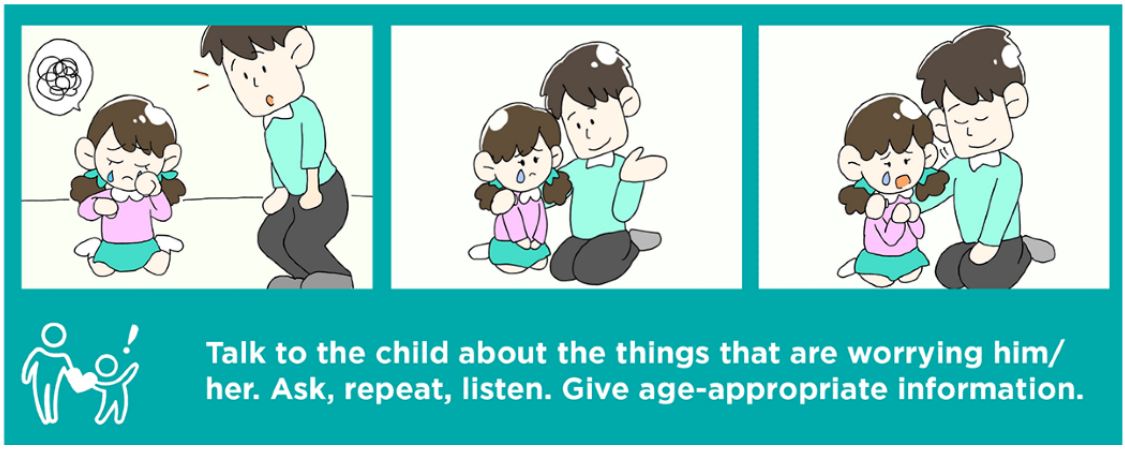


Figure 2. Still from a video that forms part of the program.

Table 1. Content of the Let's Cope Together single-session, internet-based cognitive behavioral therapy (iCBT) intervention

Theme	Content
Theme 1. Introduction	How to use the program
Theme 2. Identifying anxiety	Psychoeducation about anxiety: What are normal worries and fears, and how does anxiety manifest in children´s behavior? Bodily sensations and thinking, anxiety in different age groups and when to seek additional help
Theme 3. Listening to your child	How to talk to children about COVID-19 and their fears about it; how parents can calm themselves down and provide a suitable environment before talking with their child; how to ask open questions, listen, and validate their child's feelings
Theme 4. Relaxing by deep breathing	How parents can learn a deep-breathing technique and teach it to their child, so that they can both learn how to relax
Theme 5. Using mental practice techniques	Instructions on how to use positive mental practice techniques; how parents can teach their child to create a safe, imaginary place by using all of their senses.
Theme 6. Strengthening positive thinking	What are negative thoughts, and how do they affect feelings and behavior? How to recognize them and turn negative thoughts into positive ones
Theme 7. Positive parenting skills	Positive parenting strategies; how to plan daily routines and manage daily transitional situations; how to plan activities with children, motivate them, and give them positive feedback

Figure 3. One of the program's manga-style comic illustrations.



Study Design

There was an urgent need to initiate the study as quickly as possible due to the pandemic. The data were collected between April 17, 2020 and June 1, 2020. The pandemic lockdown started in Finland on March 16, 2020. Schools were closed, hobbies were discontinued, parents worked at home if they could, and schoolwork was carried out remotely. Families were asked not to use day care. A state of emergency was declared, which legalized additional local restrictions.

The participants were recruited through a multimedia advertisement and press release campaign, which included national media, social media—mainly Facebook and Instagram—and digital news. We also asked schools, school nurses, day care providers, and health care personnel to spread the word. There were no specific inclusion and exclusion criteria. Although the program was primarily developed for parents, other adults taking care of children, such as grandparents, other family members, and foster parents, could take part. We asked for the participants' relationship to the child before they entered the program and the study. Professionals working with children were also encouraged to use the content of the program to support their own work. They were able to get a code from the study group so that they could enter the program without completing any questions.

A pre-post design was used. After the participants completed the digital, informed consent form, they were directed to the baseline assessment, which asked them for background information and details of the child's and parents' emotional symptoms and parenting skills. Once they completed this, they could proceed to the intervention. After they had completed the intervention, we asked them about their general satisfaction with the program and whether they felt the program would improve their parenting skills. The participants could contact a member of the research group by phone or email if they wanted further information at any stage of the study.

Measures

The background information we requested before the intervention included the participant's age and level of education, the number of adults who played a parenting role in the child's life, the child's gender, and whether they attended school or day care. We also asked about any emotional symptoms they and their child were experiencing, including sleep problems, somatic symptoms, restlessness and worries, fears, or sadness (Multimedia Appendix 1). The responses ranged from 0 for "not at all" to 5 for "all the time," and the maximum scores were 25 for the child and 20 for the parents. The questions were specifically formulated for this study. They asked the participants about symptoms that are generally included in validated surveys on anxiety and depression, such as the Strengths and Difficulties Questionnaire and the Screen for Child Anxiety Related Disorders [21,22]. It was not possible to use full surveys or their subscales, as we decided to use as few questions as possible. No specific data were collected regarding the reason for the anxiety. The parenting skills included in the pre-intervention survey were related to the content of the intervention and included 12 items on how well the parent could notice the child's anxiety, act as a positive role model, cope with their own anxiety and the child's anxiety, manage daily routines, and plan ahead with the child. The answers ranged from 0 for "strongly disagree" to 6 for "strongly agree," and the maximum score was 72. Cronbach alpha was used to measure the internal consistency of the parenting skills and the child's and the parent's emotional symptoms, and these values were 0.86, 0.81, and 0.79, respectively.

After they finished the program, we asked the parents 3 general questions about how satisfied they were with the program and whether they would recommend it to others. They were also asked 6 questions about whether they had learned new skills about how to identify their child's anxiety, act as a positive role model, cope with their own anxiety and their child's anxiety, manage daily routines, and plan ahead with the child.

Statistical Analyses

The baseline survey was completed by 602 participants (524/602 [87.0%] of them were parents), and 189 (31.4%) also completed the postprogram survey.

We studied the potential differences in baseline measures between those who completed the intervention and those who did not. Pearson chi-square tests or Fisher exact tests were used for categorical variables, and Student *t* tests were used for continuous variables. Statistical significance was based on a 2-sided $P<.05$. All the statistical analyses were performed with SAS version 9.4 (SAS Institute, Cary, NC).

Ethical Issues

The Ethics Committee of the Hospital District of Southwest Finland approved the study (Turku University Hospital Ethical Board Journal number: 14/1801/2020). Participation in the study was voluntary and could be discontinued at any time. The requirements relating to information security were considered carefully when collecting, storing, analyzing, and archiving the data. Research codes were used, and no individuals could be identified from the results.

Results

About 9000 individuals visited the study website, 602 filled in the baseline survey, 468 (468/602, 77.7%) started the intervention, and 196 (196/602, 32.6%) completed all 7 themes. The postintervention survey was completed by 189 (189/602, 31.4%) of the participants who completed the baseline survey (Table 2).

We found that 77 (77/189, 40.7%) participants went through the material by logging in just once, 46 (46/189, 24.3%) by logging in twice, 34 (34/189, 18.0%) by logging in 3 times, and 32 (32/189, 16.9%) by logging in 4 times or more. The mean time that it took to complete the whole program was 73 minutes.

Table 2. Participants who completed each phase of the program and the study.

Sample	Baseline, n (%)	iCBT[a] themes (1-7)[b], n (%)							After iCBT, n (%)
		1	2	3	4	5	6	7	
Number of completers	602 (100)	412 (68.4)	378 (62.8)	340 (56.5)	283 (47.0)	240 (39.9)	214 (35.5)	196 (32.6)	189 (31.4)

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

[b]The participants could only proceed to the next theme or the postintervention survey if all the previous parts of the study were completed.

There were no significant differences between those who finished the study and those who discontinued at some point after filling in the baseline survey, when it came to education level, family structure, emotional symptoms, or self-reported parenting skills (Multimedia Appendix 2). The same was true for the child's gender, child's emotional symptoms, and whether the child attended day care or school. The mean age of the participants was slightly higher for those who finished the study than those who did not (42.2 years vs 40.4 years), and the participants were more likely to be another caregiver, rather than a parent (30/189, 15.9% vs 32/413, 7.8%). In addition, sensitivity analyses showed no significant differences in any of the background factors between those who finished the study and those who discontinued at some point after filling in the baseline survey. This included the analysis that just comprised parents.

Almost three-quarters (138/189, 73.0%) of those who completed the intervention and postintervention survey were satisfied with the program, and even more (153/189, 81.0%) reported that they would recommend the program to others (Table 3).

Most of those who finished the postintervention survey reported that they learned about childhood anxiety (146/189, 77.2%) and the skills they needed to calm themselves (141/189, 74.6%) and their child (157/189, 83.1%). In addition, 48.7% (92/189) of them reported that they had learned how to organize their daily routines better, and 53.0% (100/189) reported that they had learned how to improve the way they planned days with their child. The results remained almost the same when the sensitivity analyses focused on just parents rather than all participants.

Table 3. Participants' satisfaction with the intervention and whether they learned new skills (n=189).

Questionnaire items	Agree, n (%)	Do not agree or disagree, n (%)	Disagree, n (%)	Missing, n (%)
General satisfaction				
Program was easy to use	166 (87.8)	14 (7.4)	2 (1.1)	7 (3.7)
Was satisfied with the program	138 (73.0)	37 (19.6)	7 (3.7)	7 (3.7)
Would recommend the program to others	153 (81.0)	24 (12.7)	5 (2.7)	7 (3.7)
Developed better skills with regard to:				
Knowledge about their child's anxiety	146 (77.2)	23 (12.2)	13 (6.9)	7 (3.7)
How to show a positive example	152 (80.5)	27 (14.3)	3 (1.6)	7 (3.7)
Ways to calm themselves down	141 (74.6)	27 (14.3)	14 (7.4)	7 (3.7)
Ways to calm the child down	157 (83.1)	16 (8.5)	9 (4.7)	7 (3.7)
Organizing daily routines	92 (48.7)	65 (34.4)	25 (13.2)	7 (3.7)
Planning days together with the child	100 (53.0)	64 (33.9)	18 (9.5)	7 (3.7)

Discussion

Principal Findings

To our knowledge, this is the first paper to describe a universal digital parenting intervention that aims to improve parental skills so that they can help their children to cope with anxiety during the COVID-19 pandemic. In general, single-session interventions have been shown to be effective in treating anxiety, conduct problems, and posttraumatic stress disorders in children and adolescents [15,16,23]. Single-session, digitalized parenting interventions have also been effective in improving self-reported parenting practices [17]. Most of the parents who completed both the Let's Cope Together program and postintervention survey were satisfied with the intervention and would recommend it to others. Most parents also reported that they had acquired knowledge about their children's anxiety and had learned skills to calm themselves and their child down. Half of the parents reported that they learned how to organize daily routines and plan ahead better with their child.

The number of participants in this study was rather low, despite our substantial efforts to inform the target population about the study through online resources, including social media and digital news, schools, and day care centers. About 9000 individuals visited the study website, but only 602 registered and completed the baseline questions. Of these, about one-third finished the program and answered the postintervention questions.

There are a number of possible explanations for the low number of participants. Numerous other channels have provided support to help children and parents handle the psychosocial stress of the pandemic. These include digital news, interviews, and blogs that provide advice from experts, such as psychologists or psychotherapists. There are also several chat and telephone services run by third-sector organizations. It is likely that asking participants to provide personal data may have significantly decreased the number of participants, especially as there have been other sources of anonymous help. In addition, some of the services that were available also included human guidance, which was not the case with our program. The pandemic started relatively late in Finland, in comparison to a number of European countries, which meant that the government had more time to prepare. As a result, the number of people who have had the virus, been hospitalized, or died has been among the lowest in the world, as a proportion of the whole population. It is possible that the controlled situation in Finland reduced the need for help and that was reflected in the low participation rates.

It has been generally acknowledged that there are low participation rates in universal interventions. For example, Sampaio et al [24] found that less than 15% of the people who were eligible to take part in a universal program for the parents of preschoolers participated. Another study, by March et al [25], found that about 20% of those that logged on to a website that provided iCBT for youth anxiety did not start the program and about 30% dropped out during the first 2 sessions. It is also common for participants to fail to complete all modules of universal programs, for example due to a lack of time, technical problems, or no need for the support provided by the intervention [26].

There are some important questions that need to be answered, including who participates in programs and who benefits the most from them. It is understandable that families who do not experience the problems that the intervention tackles are not motivated to participate. In the present study, we did not find differences between those who did and did not finish the program, with regards to emotional symptoms, self-reported parenting skills, or sociodemographic characteristics. However, we were only able to study those who filled in the baseline questionnaire. The average baseline scores for emotional symptoms and self-reported parenting skills could be considered quite good already, and it might be that high-functioning families were over-represented. It is likely that parents who are already aware of, and able to think about, their children's feelings may be more motivated to take part in digital, universal programs that require initiative and active participation. At the same time, lower functioning families could need more direct human motivation and support, which was not available in the present study, due to an uncontrollable number of potential participants. Further research is needed on the importance of human guidance with regard to how effective digital interventions are among different groups of participants. In addition, the characteristics of the participating parents might have had an impact on the completion rates in the present study. For example, the iCBT was based on material that was primarily developed by our study group for children under 13 years of age. It may be possible that this younger focus made the parents of adolescents less motivated to complete the program and the study.

At the same time, universal, digital interventions with no human guidance may be cost-effective when they are delivered to those who are willing to participate and able to benefit from the intervention. That is why it is crucial to study how the content and delivery of interventions should be developed, so that they respond to the individual needs of different families. However, the pressure posed by the worsening pandemic, during the spring 2020 development phase, meant that we were not able to study the needs and preferences of the end users, with regard to both the content and delivery methods, before the iCBT was launched. It might have had positive effect on the participation rate if we had been able to do this.

Strengths and Limitations

The limitations of the study included the pre-post test design instead of a randomized controlled trial and the limited number of questions that were specifically developed for the present study. As this was a feasibility study, we only measured general satisfaction and the participants' experience with learning new skills straight after the program. As the pandemic eases, we will report further feedback, including follow-up results on changes in anxiety symptoms. Additionally, we did not have specific inclusion and exclusion criteria. The program was developed for parents, but other adults caring for children could take part. It is possible that participants, particularly those who completed the posttest, were more interested and motivated to learn new skills and more satisfied with the program. These issues limit the interpretation and generalizability of the findings. However, the study design was simple and nonselective and contained a limited number of short questions to encourage participation. The strengths of the study included the considerable previous experience of the multidisciplinary team that developed the intervention. This included CBT therapists, child psychiatrists, information technology personnel, and experienced family coaches with special training in delivering digital interventions.

Conclusion

The internet-based Let's Cope Together single-session parenting intervention helped families to cope with parenting demands during the COVID-19 pandemic. It is likely that we could face similar pandemics or other global crises in the near future. Developing and researching the effectiveness and implementation of universal digital interventions provide important preparation for managing future crises. Enhancing parenting skills is an important goal, as it increases the resilience of families with young children. The impact of current and future crises on mental health is substantial, particularly for those who have been affected most by the current pandemic. These include those with mental health and parenting problems, children and adolescents, immigrants, and frontline staff. Digitally delivered, evidence-based interventions can provide inexpensive, low-threshold help for large populations when face-to-face interventions are unavailable or not feasible.

Acknowledgments

This research was supported by the Inequalities, Interventions and a New Welfare State (INVEST) Research Flagship and Finnish Psychiatric Cohorts (PSYCOHORTS) consortium. This research was funded by the Academy of Finland Flagship Programme (decision number: 320162), the Academy of Finland Health from Cohorts and Biobanks Programme (decision number: 308552), the Academy of Finland special funding for research on COVID-19 epidemic and the mitigation of its effects, the Juho Vainio Foundation, and the Finnish Foundation for Pediatric Research.

Conflicts of Interest

AS is the founder and director of Digifamilies, which provides evidence-based treatments to Finnish public health services.

Multimedia Appendix 1
Questionnaire administered to the parents at baseline.
[PDF File (Adobe PDF File), 171 KB - jmir_v24i4e26438_app1.pdf]

Multimedia Appendix 2
Frequencies and differences in background factors of those who did and did not complete the baseline and postintervention survey.
[DOCX File , 23 KB - jmir_v24i4e26438_app2.docx]

References

1. Golberstein E, Wen H, Miller BF. Coronavirus disease 2019 (COVID-19) and mental health for children and adolescents. JAMA Pediatr 2020 Sep 01;174(9):819-820. [doi: 10.1001/jamapediatrics.2020.1456] [Medline: 32286618]
2. Idoiaga N, Berasategi N, Eiguren A, Picaza M. Exploring children's social and emotional representations of the COVID-19 pandemic. Front Psychol 2020;11:1952 [FREE Full text] [doi: 10.3389/fpsyg.2020.01952] [Medline: 32922334]
3. Patrick SW, Henkhaus LE, Zickafoose JS, Lovell K, Halvorson A, Loch S, et al. Well-being of parents and children during the COVID-19 pandemic: a national survey. Pediatrics 2020 Oct;146(4):1. [doi: 10.1542/peds.2020-016824] [Medline: 32709738]
4. Zhou S, Zhang L, Wang L, Guo Z, Wang J, Chen J, et al. Prevalence and socio-demographic correlates of psychological health problems in Chinese adolescents during the outbreak of COVID-19. Eur Child Adolesc Psychiatry 2020 Jun;29(6):749-758 [FREE Full text] [doi: 10.1007/s00787-020-01541-4] [Medline: 32363492]
5. Hawke LD, Barbic SP, Voineskos A, Szatmari P, Cleverley K, Hayes E, et al. Impacts of COVID-19 on youth mental health, substance use, and well-being: a rapid survey of clinical and community samples. Can J Psychiatry 2020 Oct;65(10):701-709 [FREE Full text] [doi: 10.1177/0706743720940562] [Medline: 32662303]
6. Dalton L, Rapa E, Stein A. Protecting the psychological health of children through effective communication about COVID-19. Lancet Child Adolesc Health 2020 May;4(5):346-347 [FREE Full text] [doi: 10.1016/S2352-4642(20)30097-3] [Medline: 32243784]
7. Fegert JM, Vitiello B, Plener PL, Clemens V. Challenges and burden of the Coronavirus 2019 (COVID-19) pandemic for child and adolescent mental health: a narrative review to highlight clinical and research needs in the acute phase and the long return to normality. Child Adolesc Psychiatry Ment Health 2020;14:20 [FREE Full text] [doi: 10.1186/s13034-020-00329-3] [Medline: 32419840]
8. Yap MBH, Jorm AF. Parental factors associated with childhood anxiety, depression, and internalizing problems: a systematic review and meta-analysis. J Affect Disord 2015 Apr 01;175:424-440. [doi: 10.1016/j.jad.2015.01.050] [Medline: 25679197]
9. Yap MBH, Morgan AJ, Cairns K, Jorm AF, Hetrick SE, Merry S. Parents in prevention: A meta-analysis of randomized controlled trials of parenting interventions to prevent internalizing problems in children from birth to age 18. Clin Psychol Rev 2016 Dec;50:138-158. [doi: 10.1016/j.cpr.2016.10.003] [Medline: 27969003]
10. Wind TR, Rijkeboer M, Andersson G, Riper H. The COVID-19 pandemic: The 'black swan' for mental health care and a turning point for e-health. Internet Interv 2020 Apr;20:100317 [FREE Full text] [doi: 10.1016/j.invent.2020.100317] [Medline: 32289019]
11. Vigerland S, Ljótsson B, Thulin U, Öst LG, Andersson G, Serlachius E. Internet-delivered cognitive behavioural therapy for children with anxiety disorders: A randomised controlled trial. Behav Res Ther 2016 Jan;76:47-56 [FREE Full text] [doi: 10.1016/j.brat.2015.11.006] [Medline: 26649465]
12. Sourander A, McGrath PJ, Ristkari T, Cunningham C, Huttunen J, Lingley-Pottie P, et al. Internet-assisted parent training intervention for disruptive behavior in 4-year-old children: a randomized clinical trial. JAMA Psychiatry 2016 Apr;73(4):378-387. [doi: 10.1001/jamapsychiatry.2015.3411] [Medline: 26913614]

13. Sourander A, McGrath PJ, Ristkari T, Cunningham C, Huttunen J, Hinkka-Yli-Salomäki S, et al. Two-year follow-up of internet and telephone assisted parent training for disruptive behavior at age 4. J Am Acad Child Adolesc Psychiatry 2018 Sep;57(9):658-668.e1. [doi: 10.1016/j.jaac.2018.07.001] [Medline: 30196869]
14. Ristkari T, Kurki M, Suominen A, Gilbert S, Sinokki A, Kinnunen M, et al. Web-based parent training intervention with telephone coaching for disruptive behavior in 4-year-old children in real-world practice: implementation study. J Med Internet Res 2019 Apr 11;21(4):e11446 [FREE Full text] [doi: 10.2196/11446] [Medline: 30973337]
15. Cobham VE, Filus A, Sanders MR. Working with parents to treat anxiety-disordered children: A proof of concept RCT evaluating Fear-less Triple P. Behav Res Ther 2017 Aug;95:128-138. [doi: 10.1016/j.brat.2017.06.004] [Medline: 28641122]
16. Haag A, Landolt MA, Kenardy JA, Schiestl CM, Kimble RM, De Young AC. Preventive intervention for trauma reactions in young injured children: results of a multi-site randomised controlled trial. J Child Psychol Psychiatry 2020 Sep;61(9):988-997. [doi: 10.1111/jcpp.13193] [Medline: 31912485]
17. Cardamone-Breen MC, Jorm AF, Lawrence KA, Rapee RM, Mackinnon AJ, Yap MBH. A single-session, web-based parenting intervention to prevent adolescent depression and anxiety disorders: randomized controlled trial. J Med Internet Res 2018 Apr 26;20(4):e148 [FREE Full text] [doi: 10.2196/jmir.9499] [Medline: 29699964]
18. Coronavirus disease (COVID-19) advice for the public: Advocacy. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/healthy-parenting [accessed 2022-03-08]
19. Coronavirus (COVID-19) parenting tips. UNICEF. URL: https://www.unicef.org/coronavirus/covid-19-parenting-tips [accessed 2022-03-08]
20. Safeplace. Vimeo. 2020 Apr 04. URL: https://vimeo.com/403960791/cd24445443 [accessed 2022-03-08]
21. Birmaher B, Khetarpal S, Brent D, Cully M, Balach L, Kaufman J, et al. The Screen for Child Anxiety Related Emotional Disorders (SCARED): scale construction and psychometric characteristics. J Am Acad Child Adolesc Psychiatry 1997 Apr;36(4):545-553. [doi: 10.1097/00004583-199704000-00018] [Medline: 9100430]
22. Goodman R. The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. J Child Psychol Psychiatry 1999 Jul;40(5):791-799. [Medline: 10433412]
23. 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 Feb;56(2):107-115. [doi: 10.1016/j.jaac.2016.11.007] [Medline: 28117056]
24. Sampaio F, Sarkadi A, Salari R, Zethraeus N, Feldman I. Cost and effects of a universal parenting programme delivered to parents of preschoolers. Eur J Public Health 2015 Dec;25(6):1035-1042. [doi: 10.1093/eurpub/ckv106] [Medline: 26063699]
25. March S, Spence SH, Donovan CL, Kenardy JA. Large-scale dissemination of internet-based cognitive behavioral therapy for youth anxiety: feasibility and acceptability study. J Med Internet Res 2018 Jul 04;20(7):e234 [FREE Full text] [doi: 10.2196/jmir.9211] [Medline: 29973338]
26. Morgan AJ, Rapee RM, Salim A, Goharpey N, Tamir E, McLellan LF, et al. Internet-delivered parenting program for prevention and early intervention of anxiety problems in young children: randomized controlled trial. J Am Acad Child Adolesc Psychiatry 2017 May;56(5):417-425.e1. [doi: 10.1016/j.jaac.2017.02.010] [Medline: 28433091]

Abbreviations

iCBT: internet-based cognitive behavioral therapy

Edited by G Eysenbach; submitted 11.12.20; peer-reviewed by C Păsărelu, M Nord, M Cardamone-Breen; comments to author 29.12.20; revised version received 19.04.21; accepted 27.12.21; published 13.04.22.

Please cite as:
Korpilahti-Leino T, Luntamo T, Ristkari T, Hinkka-Yli-Salomäki S, Pulkki-Råback L, Waris O, Matinolli HM, Sinokki A, Mori Y, Fukaya M, Yamada Y, Sourander A
Single-Session, Internet-Based Cognitive Behavioral Therapy to Improve Parenting Skills to Help Children Cope With Anxiety During the COVID-19 Pandemic: Feasibility Study
J Med Internet Res 2022;24(4):e26438
URL: https://www.jmir.org/2022/4/e26438
doi:10.2196/26438
PMID:35138265

Original Paper

Evaluation of the Immediate Effects of Web-Based Intervention Modules for Goals, Planning, and Coping Planning on Physical Activity: Secondary Analysis of a Randomized Controlled Trial on Weight Loss Maintenance

Elina Mattila[1], PhD; Graham Horgan[2], PhD; António L Palmeira[3], PhD; Ruairi O'Driscoll[4], PhD; R James Stubbs[4], PhD; Berit L Heitmann[5,6], PhD; Marta M Marques[7], PhD

[1]VTT Technical Research Centre of Finland Ltd, Tampere, Finland

[2]Biomathematics and Statistics Scotland, Edinburgh, United Kingdom

[3]Centro de Investigação em Desporto, Educação Física, Exercício e Saúde, Faculdade de Educação Física e Desporto, Universidade Lusófona de Humanidades e Tecnologias, Lisbon, Portugal

[4]School of Psychology, Faculty of Medicine and Health, University of Leeds, Leeds, United Kingdom

[5]Research Unit for Dietary Studies, The Parker Institute, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark

[6]Department of Public Health, Section for General Practice, University of Copenhagen, Copenhagen, Denmark

[7]Comprehensive Health Research Centre, NOVA Medical School, Universidade Nova de Lisboa, Lisbon, Portugal

Corresponding Author:
Elina Mattila, PhD
VTT Technical Research Centre of Finland Ltd
Visiokatu 4
Tampere, 33720
Finland
Phone: 358 407162230
Email: elina.m.mattila@vtt.fi

Abstract

Background: The use of digital interventions can be accurately monitored via log files. However, monitoring engagement with intervention goals or enactment of the actual behaviors targeted by the intervention is more difficult and is usually evaluated based on pre-post measurements in a controlled trial.

Objective: The objective of this paper is to evaluate if engaging with 2 digital intervention modules focusing on (1) physical activity goals and action plans and (2) coping with barriers has immediate effects on the actual physical activity behavior.

Methods: The NoHoW Toolkit (TK), a digital intervention developed to support long-term weight loss maintenance, was evaluated in a 2 x 2 factorial randomized controlled trial. The TK contained various modules based on behavioral self-regulation and motivation theories, as well as contextual emotion regulation approaches, and involved continuous tracking of weight and physical activity through connected commercial devices (Fitbit Aria and Charge 2). Of the 4 trial arms, 2 had access to 2 modules directly targeting physical activity: a module for goal setting and action planning (Goal) and a module for identifying barriers and coping planning (Barriers). Module visits and completion were determined based on TK log files and time spent in the module web page. Seven physical activity metrics (steps; activity; energy expenditure; fairly active, very active and total active minutes; and distance) were compared before and after visiting and completing the modules to examine whether the modules had immediate or sustained effects on physical activity. Immediate effect was determined based on 7-day windows before and after the visit, and sustained effects were evaluated for 1 to 8 weeks after module completion.

Results: Out of the 811 participants, 498 (61.4%) visited the Goal module and 406 (50.1%) visited the Barriers module. The Barriers module had an immediate effect on very active and total active minutes (very active minutes: before median 24.2, IQR 10.4-43.0 vs after median 24.9, IQR 10.0-46.3; *P*=.047; total active minutes: before median 45.1, IQR 22.9-74.9 vs after median 46.9, IQR 22.4-78.4; *P*=.03). The differences were larger when only completed Barriers modules were considered. The Barriers module completion was also associated with sustained effects in fairly active and total active minutes for most of the 8 weeks following module completion and for 3 weeks in very active minutes.

Conclusions: The Barriers module had small, significant, immediate, and sustained effects on active minutes measured by a wrist-worn activity tracker. Future interventions should pay attention to assessing barriers and planning coping mechanisms to overcome them.

Trial Registration: ISRCTN Registry ISRCTN88405328; https://www.isrctn.com/ISRCTN88405328

(J Med Internet Res 2022;24(4):e35614) doi:10.2196/35614

KEYWORDS

digital intervention; Fitbit; weight; weight loss maintenance; physical activity; fitness; exercise; goal setting; action planning; coping planning; control trial; secondary analysis; RCT; randomized controlled trial; long-term effect; short-term effect; immediate effect; sustained effect

Introduction

Digital behavior change interventions (DBCIs) hold the promise of providing personalized and adaptive treatments to improve health. It is already possible to track individual user interactions with DBCI components at a high level of detail and fidelity, such as when and for how long users are accessing them. The evidence linking use to intervention outcomes is mixed, and the relationships are not straightforward. Further, the type of engagement required for effects may also vary across different types of interventions [1,2].

User engagement with the real-world behaviors that a DBCI is attempting to influence, the macro-level engagement [1], is more difficult to determine. Traditionally, behavioral changes are verified using questionnaires or laboratory measurements conducted intermittently, with measurement points several months apart and the risk of reporting bias. At that point, it is impossible to prove causation between specific intervention components and change. Monitoring the immediate effects of DBCI components would enable active process evaluation of interventions and contribute greatly to tailoring effective, adaptive, and personalized interventions. For example, if an intervention component has an immediate effect on the user's behavior, the component could be repeated after a while to amplify the effect. If a component does not lead to an expected effect for a user, it could be switched off and an alternative intervention could be launched to create an adaptive intervention [3].

Research on behavior change techniques has revealed the positive effects of self-regulatory techniques on physical activity behavior. Core self-regulatory techniques are self-monitoring of behavior and feedback; goal setting, that is, identification and formulation of a physical activity goal; action planning, that is, specification of the goal in a detailed plan for the performance of the behavior (context, frequency, duration, and the intensity of the activity); and coping planning, that is, identification of barriers to physical activity and planning ways to overcome them using, for example, an "if, then" approach [4-6].

This study focuses on a web-based toolkit for weight loss maintenance, consisting of various modules based on behavioral self-regulation and motivation theories, as well as contextual emotion regulation. The objective of this study was to investigate whether intervention modules aimed at increasing physical activity through goal setting and action planning, and coping planning had immediate and sustained effects on the physical activity behavior of users. This was a secondary analysis of the data from a European Commission Horizon 2020–funded NoHoW project.

Methods

Ethics Approval

The trial was registered with the ISRCTN registry (ISRCTN88405328). Ethical approval was granted by local institutional ethics committees at the Universities of Leeds (17-0082; 27 February 2017), Lisbon (17/2016; 20 February 2017) and the Capital Region of Denmark (H-16030495; 8 March 2017).

Study Procedures

The NoHoW trial (ISRCTN88405328) was an 18-month, 3-center, 2-by-2 factorial, single-blinded, randomized controlled trial, evaluating a digital weight loss maintenance intervention. The participants were required to be aged ≥18 years, have a verified ≥5% weight loss in the last 12 months with current weight at least 5% below their highest weight, and have had a BMI of ≥25 kg/m^2 before weight loss. A total of 1627 participants were recruited and randomly assigned to one of the following four arms: (1) control or self-monitoring (n=400), (2) *motivation and self-regulation* (n=403), (3) *emotion regulation* (n=416), and (4) *combined* arm (n=408). All participants received activity trackers (Fitbit Charge 2) to be worn throughout the trial, weight scales (Fitbit Aria), and access to the web-based NoHoW Toolkit (TK) tailored to their respective arm. Participants in intervention arms were encouraged to complete 18 intervention modules in the TK during the first 6 months of the trial. The participants received weekly emails during the first 18 weeks as reminders recommending visiting a specific module. A detailed description of the trial is presented by Scott et al [7]. The TK design and content are presented in detail by Marques et al [8].The *motivation and self-regulation* and *combined* arms had 2 modules focusing on physical activity: physical activity goal (Goal) and physical activity barriers (Barriers). The Goal module (see Multimedia Appendix 1 for screenshots) addressed goal setting and action planning, contained information on how to set goals, and had a form to set a goal and detailed plan for either the number of steps per day or other type of physical activity. The estimated duration to complete the Goal module was 10 minutes. The Barriers module (see Multimedia Appendix 2 for screenshots) was introduced later in the intervention and contained a testimonial

on potential barriers of physical activity and an interactive exercise for identifying personal barriers and creating a coping plan to deal with them. The estimated duration to complete the Barriers module was 8 minutes.

Analysis

Visits to the Goal and Barriers modules were identified based on log files of the TK. The duration spent in the modules was calculated based on log events signifying entering and leaving the module. A module visit was considered complete if it lasted at least 33% of the estimated duration of the module (ie, 3.3/10 min for the Goal module and 2.6/8 min for the Barriers module) or if the duration of the visit could not be determined due to a missing end event. The threshold of 33% was determined by intervention designers as the minimum time required to become exposed to the behavior change mechanisms in these modules.

Daily summaries provided by Fitbit were used as the physical activity metrics and included daily steps; activity energy expenditure; active minutes categorized to fairly active, very active, and total active minutes; and distance. Activity metrics were averaged over the 7 days prior to visiting the modules and 7 days after visiting them. Days with less than 1000 steps were considered missing data and were not included. A similar threshold has been used in several previous studies (eg, [9,10]). It was also required that the 7-day periods contained at least 4 days of activity data. The immediate effects are presented both for all visits to the modules and for complete visits. If an immediate effect was found, maintenance of the effects was evaluated for 8 weeks following the module visits, considering only complete module visits. A comparison of overall changes in activity metrics between the first and sixth month of the study is presented in Multimedia Appendix 3.

As most of the physical activity metrics had skewed distributions, nonparametric methods were used. Median and interquartile ranges were calculated, and nonparametric tests (Wilcoxon signed-rank test) were used for comparisons. All analyses were conducted with Matlab R2017a (Mathworks) and SPSS Statistics software (version 26; IBM Corp). Statistical significance was set at $P<.05$.

Results

User Statistics

The modules were available for 811 participants (ie, the participants randomized to the *motivation and self-regulation* and to the *combined* arms). Of the 811 participants, 498 (61.4%) visited the Goal module (252/403 *motivation and self-regulation* and 246/408 *combined*), and 406 (50.1%) participants visited the Barriers module (217/403 *motivation and self-regulation* and 189/408 *combined*). There were 628 visits to the Goal module, of which 309 were complete visits. The Barriers module had 514 visits, with 345 complete. The background characteristics of the visitors and nonvisitors of these modules are presented and compared in Multimedia Appendix 3.

Immediate Effects of Modules

The Goal module was first visited a median of 55 days (IQR 48-64) after the first login to TK, and the Barriers module was first visited a median of 98 days (IQR 91-110) after the first login. Tables 1 and 2 present the median and IQR for the activity metrics before and after visiting the Goal and Barriers modules, respectively. Visiting the Barriers module increased very active and total active minutes.

When only completed module visits were included, the results for the Goal module remained nonsignificant. For the Barriers module, the effects remained and became slightly stronger (for very active minutes: before median 22.4, IQR 10.1-41.0 vs after median 25.0, IQR 10.1-46.1; $P=.007$; and for total active minutes: before median 42.6, IQR 22.4-73.4 vs after median 46.6, IQR 21.8-79.4; $P=.008$).

Table 1. Activity metrics before and after visiting the Goal module.

	Before, median (IQR)	After, median (IQR)	*P* value[a]
Steps	9637 (7383-12,485)	9638 (7354-12,425)	.62
Energy (kcal)	1276 (997-1632)	1274 (987-1640)	.74
Fairly active (min)	18.9 (10.2-33.3)	18.3 (11.0-32.5)	.69
Very active (min)	23.9 (12.2-44.8)	26.4 (12.6-44.8)	.39
Total active (min)	46.6 (25.1-75.4)	48.3 (26.6-76.5)	.44
Distance (km)	6.8 (5.1-8.9)	6.8 (5.1-8.7)	.60

[a]Wilcoxon signed-rank test.

Table 2. Activity metrics before and after visiting the Barriers module.

	Before, median (IQR)	After, median (IQR)	*P* value[a]
Steps	9449 (7240-12,112)	9119 (7202-12,382)	.70
Energy (kcal)	1211 (990-1585)	1260 (943-1611)	.27
Fairly active (min)	16.7 (9.7-31.9)	18.9 (9.6-33.8)	.15
Very active (min)	24.2 (10.4-43.0)	24.9 (10.0-46.3)	.047
Total active (min)	45.1 (22.9-74.9)	46.9 (22.4-78.4)	.03
Distance (km)	6.7 (4.9-8.5)	6.4 (4.9-8.7)	.68

[a]Wilcoxon signed-rank test.

Table 3 presents the 8-week maintenance of the Barriers module effect for the 3 active minute metrics based on completed module visits only. Total active minutes and fairly active minutes were higher than before module completion for most of the 8-week period following the module. Also, very active minutes remained higher for 3 weeks after module completion. Values that significantly differed from the before value were denoted.

Table 3. Median (IQR) values for active minutes categories before and after the Barriers module based on completed modules.

	Fairly active (min), median (IQR)	Very active (min), median (IQR)	Total active (min), median (IQR)
Before	15.9 (9.4-29.4)	22.4 (10.1-41.0)	42.6 (22.4-73.4)
Week 1 after	18.4 (9.2-33.3)	25.0 (10.1-46.1)[a]	46.6 (21.8-79.4)[a]
Week 2 after	19.1 (10.9-31.8)[a]	26.6 (12.0-45.0)[b]	45.6 (26.4-75.7)[b]
Week 3 after	18.6 (10.2-34.6)[b]	24.4 (11.9-45.6)[a]	44.6 (25.3-78.2)[a]
Week 4 after	18.1 (11.6-37.0)[b]	23.1 (12.0-43.2)	47.2 (25.2-76.8)[a]
Week 5 after	19.6 (10.6-30.6)[a]	24.4 (11.9-42.1)	44.7 (25.7-72.3)
Week 6 after	19.6 (11.4-33.3)[b]	23.6 (12.4-41.2)	45.3 (27.1-78.5)[a]
Week 7 after	20.3 (11.2-32.7)[b]	24.0 (12.0-40.9)	44.4 (27.6-74.4)[c]
Week 8 after	20.3 (11.2-31.7)[a]	23.7 (11.2-42.3)	46.1 (26.7-73.4)

[a]$P<.01$, Wilcoxon signed-rank test.
[b]$P<.001$.
[c]$P<.05$.

Discussion

Principal Results

This paper investigates the immediate changes in measured physical activity after visiting 2 web-based intervention modules targeting physical activity goals and action planning (Goal), and coping planning (Barriers). The Barriers module had a significant but small, immediate effect on very active and total active minutes during the week after visiting the module. When only completed module visits were considered, increases were larger. Module completion was also associated with sustained increases in all categories of active minutes, of which fairly active minutes and total active minutes were sustained for most of the 8-week period following module completion, and very active minutes for the first 3 weeks. Coping planning, addressed in the Barriers module, has previously been identified as a core self-regulation technique that can directly impact behavior (eg, [11,12]), and our findings support this effect for physical activity. The Goal module did not show similar effects on physical activity. One explanation for the differences in the results is that the coping planning activity in the Barriers module asked participants to identify barriers and strategies to overcome them in the immediate future (following week), while in the Goal module participants were asked to set a goal to increase their activity moderately in an unspecified timeframe.

Strengths and Limitations

The strengths of this study include the large number of participants and the ability to monitor the exposure to specific intervention techniques as well as the subsequent behavior.

The reliability of the variables may be affected by differences in individual wear times of the trackers. This was mitigated by setting a threshold of 1000 steps to consider the day valid. There is also a known tendency for Fitbit (and other trackers) to overestimate the amount of moderate to vigorous activity [13,14]. Furthermore, we used multiple observations for some volunteers to maximize the sample size. Although the rank-based nonparametric tests could not accommodate a random effect to account for the multiple observations, we checked an analysis,

which averaged the observations per participant and found that the results and conclusions were the same.

The timing of the modules may have impacted the results. Although the order of visiting the modules was not technically restricted, the Goal session appeared earlier in the intervention flow and was visited by more participants than the Barriers session. It is thus possible that participants visiting the Barriers session were more committed to the intervention and, therefore, more likely to adhere to behavioral guidance as well. Further, engagement was only assessed based on log files and not by the quality of the action and coping plans done by the user, or the enactment of these.

Future Work

As far as we know, no prior research to which we can directly compare these results exists. This shows the need for more research to examine the direct and immediate as well as longer-term effects of engaging with DBCI content on the enactment of the target behavior. Specifically, such work constitutes an important step toward identifying which behavior change techniques can have a differential impact on physical activity. This knowledge can, in turn, contribute to optimizing DBCIs in an adaptive and personalized way.

Conclusions

A self-regulation–based intervention module addressing physical activity barriers induced a significant increase in active minutes, and the effect was stronger when the module was completed. Module completion was further associated with sustained increases, especially in fairly active and total active minutes.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement 643309. The material presented and views expressed here are the responsibility of the author(s) only. The European Commission takes no responsibility for any use made of the information set out.

Authors' Contributions

EM and MMM conceived the study. EM conducted the analyses with consultation from GH. RJS was the principal investigator for the NoHoW trial, and BLH was the principal coordinator of the NoHoW project. EM, MMM, ALP and GH drafted the manuscript, and all authors revised the manuscript and approved the final version.

Conflicts of Interest

RJS consults for Slimming World through Consulting Leeds, which is a wholly owned subsidiary of the University of Leeds. Slimming World was a former partner in NoHoW. MMM and GH have previously consulted for Slimming World. Other coauthors have no conflicts of interest to declare.

Multimedia Appendix 1
Screenshots of the Goal module.
[PDF File (Adobe PDF File), 346 KB - jmir_v24i4e35614_app1.pdf]

Multimedia Appendix 2
Screenshots of the Barriers module.
[PDF File (Adobe PDF File), 460 KB - jmir_v24i4e35614_app2.pdf]

Multimedia Appendix 3
Additional analyses on background characteristics and activity metrics.
[DOCX File , 17 KB - jmir_v24i4e35614_app3.docx]

References

1. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
2. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. 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]
3. Almirall D, Nahum-Shani I, Sherwood NE, Murphy SA. Introduction to SMART designs for the development of adaptive interventions: with application to weight loss research. Transl Behav Med 2014 Sep;4(3):260-274 [FREE Full text] [doi: 10.1007/s13142-014-0265-0] [Medline: 25264466]

4. Sniehotta FF, Scholz U, Schwarzer R. Bridging the intention-behaviour gap: planning, self-efficacy, and action control in the adoption and maintenance of physical exercise. Psychol Health 2005 Apr;20(2):143-160. [doi: 10.1080/08870440512331317670]
5. Dombrowski SU, Sniehotta FF, Avenell A, Johnston M, MacLennan G, Araújo-Soares V. Identifying active ingredients in complex behavioural interventions for obese adults with obesity-related co-morbidities or additional risk factors for co-morbidities: a systematic review. Health Psychol Rev 2012 Mar;6(1):7-32. [doi: 10.1080/17437199.2010.513298]
6. Hagger MS, Luszczynska A, de Wit J, Benyamini Y, Burkert S, Chamberland P, et al. Implementation intention and planning interventions in health psychology: recommendations from the synergy expert group for research and practice. Psychol Health 2016 Jul;31(7):814-839. [doi: 10.1080/08870446.2016.1146719] [Medline: 26892502]
7. Scott SE, Duarte C, Encantado J, Evans EH, Harjumaa M, Heitmann BL, et al. The NoHoW protocol: a multicentre 2×2 factorial randomised controlled trial investigating an evidence-based digital toolkit for weight loss maintenance in European adults. BMJ Open 2019 Sep 30;9(9):e029425 [FREE Full text] [doi: 10.1136/bmjopen-2019-029425] [Medline: 31575569]
8. Marques MM, Matos M, Mattila E, Encantado J, Duarte C, Teixeira PJ, et al. A theory- and evidence-based digital intervention tool for weight loss maintenance (NoHoW Toolkit): systematic development and refinement study. J Med Internet Res 2021 Dec 03;23(12):e25305 [FREE Full text] [doi: 10.2196/25305] [Medline: 34870602]
9. Orstad SL, Gerchow L, Patel NR, Reddy M, Hernandez C, Wilson DK, et al. Defining valid activity monitor data: a multimethod analysis of weight-loss intervention participants' barriers to wear and first 100 days of physical activity. Informatics 2021 Jun 06;8(2):39. [doi: 10.3390/informatics8020039]
10. Hardcastle SJ, Jiménez-Castuera R, Maxwell-Smith C, Bulsara MK, Hince D. Fitbit wear-time and patterns of activity in cancer survivors throughout a physical activity intervention and follow-up: exploratory analysis from a randomised controlled trial. PLoS One 2020;15(10):e0240967 [FREE Full text] [doi: 10.1371/journal.pone.0240967] [Medline: 33075100]
11. Kwasnicka D, Presseau J, White M, Sniehotta FF. Does planning how to cope with anticipated barriers facilitate health-related behaviour change? A systematic review. Health Psychol Rev 2013 Feb 19;7(2):129-145. [doi: 10.1080/17437199.2013.766832]
12. Schroé H, Van Dyck D, De Paepe A, Poppe L, Loh WW, Verloigne M, et al. Which behaviour change techniques are effective to promote physical activity and reduce sedentary behaviour in adults: a factorial randomized trial of an e- and m-health intervention. Int J Behav Nutr Phys Act 2020 Oct 07;17(1):127 [FREE Full text] [doi: 10.1186/s12966-020-01001-x] [Medline: 33028335]
13. Tedesco S, Sica M, Ancillao A, Timmons S, Barton J, O'Flynn B. Validity evaluation of the Fitbit Charge2 and the Garmin Vivosmart HR+ in free-living environments in an older adult cohort. JMIR Mhealth Uhealth 2019 Jun 19;7(6):e13084 [FREE Full text] [doi: 10.2196/13084] [Medline: 31219048]
14. Redenius N, Kim Y, Byun W. Concurrent validity of the Fitbit for assessing sedentary behavior and moderate-to-vigorous physical activity. BMC Med Res Methodol 2019 Feb 07;19(1):29 [FREE Full text] [doi: 10.1186/s12874-019-0668-1] [Medline: 30732582]

Abbreviations

DBCI: digital behavior change intervention
TK: NoHoW Toolkit

Edited by A Mavragani; submitted 10.12.21; peer-reviewed by G Crombez, E Neter; comments to author 19.01.22; revised version received 23.02.22; accepted 15.03.22; published 14.04.22.

Please cite as:
Mattila E, Horgan G, Palmeira AL, O'Driscoll R, Stubbs RJ, Heitmann BL, Marques MM
Evaluation of the Immediate Effects of Web-Based Intervention Modules for Goals, Planning, and Coping Planning on Physical Activity: Secondary Analysis of a Randomized Controlled Trial on Weight Loss Maintenance
J Med Internet Res 2022;24(4):e35614
URL: https://www.jmir.org/2022/4/e35614
doi:10.2196/35614
PMID:35436232

Original Paper

Cannabis Use in Adults Who Screen Positive for Attention Deficit/Hyperactivity Disorder: CANreduce 2.0 Randomized Controlled Trial Subgroup Analysis

Joachim Ahlers[1], MA; Christian Baumgartner[1], PhD; Mareike Augsburger[1], PhD; Andreas Wenger[1], MSc; Doris Malischnig[2], PhD; Nikolaos Boumparis[1], MSc; Thomas Berger[3], PhD; Lars Stark[4], MA; David D Ebert[5], PhD; Severin Haug[1], PhD; Michael P Schaub[1], PhD

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

[2]Institute for Addiction Prevention, Office of Addiction and Drug Policy of Vienna, Vienna, Austria

[3]Department of Clinical Psychology and Psychotherapy, University of Bern, Bern, Switzerland

[4]Arud Centre for Addiction Medicine, Zurich, Switzerland

[5]Clinical Department for Sport and Health Sciences, Technical University Munich, Munich, Germany

Corresponding Author:
Joachim Ahlers, MA
Swiss Research Institute for Public Health and Addiction
University of Zurich
Konradstrasse 32
Zurich, 8005
Switzerland
Phone: 41 444481160
Email: joachim.ahlers@wisemed.ch

Abstract

Background: Prevalence rates for lifetime cannabis use and cannabis use disorder are much higher in people with attention deficit/hyperactivity disorder than in those without. CANreduce 2.0 is an intervention that is generally effective at reducing cannabis use in cannabis misusers. This self-guided web-based intervention (6-week duration) consists of modules grounded in motivational interviewing and cognitive behavioral therapy.

Objective: We aimed to evaluate whether the CANreduce 2.0 intervention affects cannabis use patterns and symptom severity in adults who screen positive for attention deficit/hyperactivity disorder more than in those who do not.

Methods: We performed a secondary analysis of data from a previous study with the inclusion criterion of cannabis use at least once weekly over the last 30 days. Adults with and without attention deficit/hyperactivity disorder (based on the Adult Attention deficit/hyperactivity disorder Self-Report screener) who were enrolled to the active intervention arms of CANreduce 2.0 were compared regarding the number of days cannabis was used in the preceding 30 days, the cannabis use disorder identification test score (CUDIT) and the severity of dependence scale score (SDS) at baseline and the 3-month follow-up. Secondary outcomes were Generalized Anxiety Disorder score, Center for Epidemiological Studies Depression scale score, retention, intervention adherence, and safety.

Results: Both adults with (n=94) and without (n=273) positive attention-deficit/hyperactivity disorder screening reported significantly reduced frequency (reduction in consumption days: with: mean 11.53, SD 9.28, P<.001; without: mean 8.53, SD 9.4, P<.001) and severity of cannabis use (SDS: with: mean 3.57, SD 3.65, P<.001; without: mean 2.47, SD 3.39, P<.001; CUDIT: with: mean 6.38, SD 5.96, P<.001; without: mean 5.33, SD 6.05, P<.001), as well as anxiety (with: mean 4.31, SD 4.71, P<.001; without: mean 1.84, SD 4.22, P<.001) and depression (with: mean 10.25, SD 10.54; without: mean 4.39, SD 10.22, P<.001). Those who screened positive for attention deficit/hyperactivity disorder also reported significantly decreased attention deficit/hyperactivity disorder scores (mean 4.65, SD 4.44, P<.001). There were no significant differences in change in use (P=.08), dependence (P=.95), use disorder (P=.85), attention deficit/hyperactivity disorder status (P=.84), depression (P=.84), or anxiety (P=.26) between baseline and final follow-up, dependent on positive attention-deficit/hyperactivity disorder screening. Attention deficit/hyperactivity disorder symptom severity at baseline was not associated with reduced cannabis use frequency or severity but was linked to greater reductions in depression (Spearman ρ=.33) and anxiety (Spearman ρ=.28). Individuals with positive

attention deficit/hyperactivity disorder screening were significantly less likely to fill out the consumption diary (*P*=.02), but the association between continuous attention deficit/hyperactivity disorder symptom severity and retention (Spearman ρ=−0.10, *P*=.13) was nonsignificant. There also was no significant intergroup difference in the number of completed modules (with: mean 2.10, SD 2.33; without: mean 2.36, SD 2.36, *P*=.34), and there was no association with attention deficit/hyperactivity disorder symptom severity (Spearman ρ=−0.09; *P*=.43). The same was true for the rate of adverse effects (*P*=.33).

Conclusions: Cannabis users screening positive for attention deficit/hyperactivity disorder may benefit from CANreduce 2.0 to decrease the frequency and severity of cannabis dependence and attenuate symptoms of depression and attention deficit/hyperactivity disorder-related symptoms. This web-based program's advantages include its accessibility for remote users and a personalized counselling option that may contribute to increased adherence and motivation to change among program users.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN) 11086185; http://www.isrctn.com/ISRCTN11086185

(J Med Internet Res 2022;24(4):e30138) doi:10.2196/30138

KEYWORDS

attention deficit/hyperactivity disorder; ADHD; cannabis; cannabis use disorder; CANreduce; web-based self-help tool; online tool; online health; mental health; digital health; anxiety; depression

Introduction

The worldwide prevalence of attention deficit/hyperactivity disorder is estimated to be 5% in children and up to 4% in adults [1-3]. Lifetime prevalence rates for cannabis use are rising [4], with current rates of 26% in Europe [5], 46% in the United States [6], and 28% in Switzerland (4). Lifetime cannabis use is significantly more common in people with attention deficit/hyperactivity disorder (66.1%) than in those without (46.9%); similarly, cannabis use disorders are significantly more common in people with attention deficit/hyperactivity disorder (23.5%) than in those without (8%) [7].

Cannabis is predominantly seen as a safe drug in Western societies [8]. Life satisfaction and stress do not seem to predict the initiation, cessation, or severity of cannabis use in young adults [9,10]; however, research indicates a higher prevalence of various mental health symptoms in cannabis users, such as depression, anxiety [10-13], and attention deficit/hyperactivity disorder, with increased severity in certain populations such as those in the Czech Republic and France [14-16].

In addition, meta-analyses have documented that attention deficit/hyperactivity disorder alone is linked to high comorbidity rates of mental health symptoms, such as anxiety disorders [17] and depression [18], across all life stages.

Recent research [19] showed a causal genetic link between attention deficit/hyperactivity disorder and lifetime cannabis use and emphasizes the hereditary nature of both entities. Specifically, the heritability of attention deficit/hyperactivity disorder is estimated at 70%-80% and of cannabis use initiation at 40%-48% [19].

Cognitive behavioral therapy is an effective treatment option for adults with attention deficit/hyperactivity disorder [20]. Furthermore, web-based cognitive behavioral therapy has been shown to be more effective for cannabis users with attention deficit/hyperactivity disorder than for those without [21]. In addition, the first study [22] that assessed a web-based intervention for people with attention deficit/hyperactivity disorder found it to be successful alleviating attention deficit/hyperactivity disorder symptoms.

There is evidence that integrated cognitive behavioral therapy—a combination of 2 research-based cognitive behavioral therapy methods (one for substance use disorder and one for attention deficit/hyperactivity disorder)—performs significantly better than regular addiction treatment, such as cognitive behavioral therapy alone, among attention deficit/hyperactivity disorder patients with cannabis use disorder. Patel et al [23] found that, in adolescents with attention deficit/hyperactivity disorder, comorbid cannabis use disorder is associated with a 90% lower likelihood of successful attention deficit/hyperactivity disorder treatment outcomes than that without comorbid cannabis use disorder. A recent randomized controlled trial revealed that, compared to regular substance use disorder cognitive behavioral therapy, integrated cognitive behavioral therapy resulted in significantly greater improvement in attention deficit/hyperactivity disorder symptoms in patients with substance use disorder and attention deficit/hyperactivity disorder [24].

Given the high comorbidity rates of attention deficit/hyperactivity disorder and substance use disorder, findings with respect to individuals with attention deficit/hyperactivity disorder and substance use disorder are promising and should be further evaluated to improve treatment outcomes for individuals with attention deficit/hyperactivity disorder and comorbid cannabis use disorder.

Among studies assessing web-based treatments designed to reduce cannabis use, none has considered whether adults with attention deficit/hyperactivity disorder were included or whether attention deficit/hyperactivity disorder is a potential moderator of treatment effectiveness.

Web-based interventions are known for removing barriers against seeking help for addictions, particularly stigmatization and inadequate access to treatment facilities [25]. Even though there is a lack of comparable studies using web-based programs to guide cannabis users with attention deficit/hyperactivity disorder through their process of reducing or quitting substance

use, the effect of cannabis use on persons with attention deficit/hyperactivity disorder and the characteristics of this subgroup of cannabis users have been investigated. Soler Artigas et al [19] discovered a causal relationship between attention deficit/hyperactivity disorder and cannabis use, based on the identification of specific loci predisposing individuals to these traits, which showed that patients with attention deficit/hyperactivity disorder have a 7.9-fold increase in the odds of using cannabis than persons without this condition. Wallace et al [26] revealed that the symptoms associated with attention deficits—such as low scores on neuropsychological performance tests—were likely attributable to cannabis use itself rather than to attention deficit/hyperactivity disorder. Brandt et al [27] retrospectively investigated clinical parameters that correlated with cannabis use among respondents with and without attention deficit/hyperactivity disorder in the National Epidemiologic Survey on Alcohol and Related Conditions, and although 14.3% of the respondents with attention deficit/hyperactivity disorder used cannabis, only 4.3% of those without attention deficit/hyperactivity disorder used cannabis. There was also a significantly higher prevalence of psychiatric and personality disorders in respondents with attention deficit/hyperactivity disorder who consumed cannabis than in those who did not [27]. Patel et al [23] retrospectively evaluated patient data from adolescents with attention deficit/hyperactivity disorder and comorbid cannabis use disorder and observed this patient group's increased need for acute care. A similar study [28] revealed that a perception exists, among attention deficit/hyperactivity disorder–patient web-based forum participants, that cannabis is beneficial and reduces symptoms associated with the condition.

Several recent meta-analyses [25,29-31] have demonstrated the overall effectiveness and great potential of web-based prevention and treatment interventions for cannabis use reduction but that such studies have typically been plagued by high dropout rates. Thus, further analysis of web-based interventions seeking to reduce cannabis use in adults with comorbid attention deficit/hyperactivity disorder is warranted.

In a recent 3-arm randomized controlled trial (CANreduce 2.0 study) [32], we examined the effects of an enhanced self-guided web-based intervention tool with a social presence in treatment-seeking adults who overuse cannabis. The social presence included an eCoach for supportive accountability and human support to enhance adherence to the eHealth intervention. We found moderate to medium effects in the reduction of cannabis use days and significant effects influencing secondary cannabis related outcomes, as well as reducing general anxiety disorder symptoms, when compared to an internet-as-usual control group [33].

Previously published studies [27,34-36] have demonstrated that individuals with attention deficit/hyperactivity disorder may be particularly at risk of using cannabis, that the severity of cannabis involvement is significantly associated with greater endorsement of attention deficit/hyperactivity disorder symptoms [34], and that diagnosis of any psychiatric disorder is significantly higher among those with attention deficit/hyperactivity disorder and concurrent cannabis use [27]. Furthermore, cannabis use has been shown to interact with and decrease the beneficial effects of commonly prescribed medication for individuals with attention deficit/hyperactivity disorder [35]. In addition, cannabis use severity seems to worsen attention deficit/hyperactivity disorder symptoms [36]. Therefore, we anticipated that individuals screening positive for attention deficit/hyperactivity disorder might particularly benefit from the CANreduce 2.0 intervention compared to individuals without attention deficit/hyperactivity disorder symptoms.

The aim of this study was to gain insights into the impact of attention deficit/hyperactivity disorder symptoms on outcomes when individuals who overuse cannabis participate in a web-based intervention, in terms of the program's efficacy reducing cannabis use, while also examining how severity is affected. In particular, we sought to answer the following question: Does attention deficit/hyperactivity disorder severity correlate with change in cannabis consumption posttreatment? Secondary outcomes were changes in attention deficit/hyperactivity disorder symptom severity, intervention adherence and retention, and how safe the intervention was perceived to be. We hypothesized (1) that the CANreduce 2.0 intervention would reduce cannabis use and associated mental health problems (ie, depression and anxiety) in participants whether they screened positive or negative for attention deficit/hyperactivity disorder; (2) that participants screening positive for attention deficit/hyperactivity disorder would benefit to a greater extent, in terms of cannabis use reduction, than those screening negative for attention deficit/hyperactivity disorder; and (3) that baseline attention deficit/hyperactivity disorder symptom severity would correlate with intervention-related changes in other outcomes.

Methods

Study Design

We conducted secondary analysis of the CANreduce 2.0 data set [33]. The sample was extracted from the 2 active intervention arms (n=367), excluding all individuals in the internet-as-usual group because the study's purpose was not to determine whether either active intervention was effective relative to a control condition but to compare the degree of effect in patients screening positive versus screening negative for attention deficit/hyperactivity disorder. These 367 adults included 94 young adults who screened positive and 273 who screened negative for comorbid attention deficit/hyperactivity disorder. For our analysis, participants with attention deficit/hyperactivity disorder in the 2 active intervention arms were pooled, as were those who screened negative.

In the original 3-arm RCT, 2 active arms with an adherence-focused, guidance enhanced, web-based self-help intervention with and without a mostly automated personal eCoach were compared with a nonactive arm (waiting-list controls with access to internet services as usual). The concept of adherence-focused guidance stems from observations that guided self-help programs are more effective than programs without guidance, based on the supportive-accountability model [37]. Each of the 2 active interventions consisted of 8 modules specifically developed to decrease cannabis use and reduce

symptoms of common mental disorders such as attention deficit/hyperactivity disorder, anxiety, and depression. Module content was based on the strategies of motivational interviewing and cognitive behavioral therapy. Study participants were assessed after the intervention, by comparing their baseline characteristics with reports obtained at the 3-month follow-up assessment. Retention of participants in these subgroups was evaluated weekly until the end of the intervention.

Recruitment

Recruitment took place from August 2016 through December 2018. Potential participants were recruited via the CANreduce websites [38,39] and associated health-related websites linked to the study. Recruitment also was achieved through advertisements in relevant internet forums and newspapers (or web-based versions thereof) and search engine website advertisements. The recruitment process was not attention deficit/hyperactivity disorder–specific—recruitment was neither addressed to persons with attention deficit/hyperactivity disorder, nor performed in attention deficit/hyperactivity disorder-related institutions, websites, or forums. After completing the 3-month follow-up survey, participants were offered either a voucher (30€, approximately US $33.50) or the choice to donate the equivalent amount to charity.

Consent Procedure, Registration, and Randomization

Participants could register on the website and had to provide only minimal personal data (email address; a phone number to be contacted, if follow-up questionnaires were not filled out; and basic demographic data), in accordance with the CANreduce 2.0 research protocol.

Inclusion and Exclusion Criteria

Interested individuals initially were informed about the purpose, background, and structure of the study. They were provided with information on the inclusion and exclusion criteria (Table 1), followed by information on ethical safe-guards (the right to withdraw at any time, confidentiality) and on data protection and safety arrangements. Informed consent consisted of activating several check boxes that restated important study points and clicking a consent submission button. Potential participants who stated they were in any psychosocial or psychiatric treatment for their cannabis use disorder were excluded. There was no question exploring previous face-to-face attention deficit/hyperactivity disorder diagnostic or related medication, since the rate of participants receiving current psychiatric or pharmacological treatment for adult attention deficit/hyperactivity disorder was expected to be low. Those still interested and eligible were then asked to register on the website and complete a baseline assessment, after which they were randomized by a computer algorithm in a 1:1:1 ratio into 1 of the 3 study arms (2 active, 1 control). Participants in study arms 1 and 2 were introduced, step by step, to their intervention, while those in study arm 3 were informed that they would be granted access to the intervention after they completed their follow-up assessment 3 months later.

Table 1. Inclusion and exclusion criteria with rationales.

Criterion	Rationale
Inclusion	
Informed consent via a web form	Ensure knowledge of procedures and declaration of consent
At least 18 years old	Ensure minimum age of participation
Cannabis use at least once weekly over the last 30 days	Include participants with less than daily cannabis use and increase validity
Internet access at least once weekly and a valid email address	Ensure access to the intervention
Good command of the German language	Ensure that participants understand the information provided
Exclusion	
Participation in other psychosocial or pharmacological treatments for the reduction or cessation of cannabis use	Avoid confounding treatment effects
Current pharmacologically treated psychiatric disease or any history of psychosis, schizophrenia, bipolar type I disorder, or significant current suicidal or homicidal thoughts	Prevent individuals with such problems from entering the study

Study Interventions

CANreduce version 1.0 [40,41] had already been shown to be effective at reducing cannabis use by combining an automated self-help program (web-based psychoeducation modules with a consumption diary) with the opportunity for individual chat counseling, both of which were grounded in motivational interviewing, self-control practices, and classical cognitive behavioral therapy. However, there were difficulties with adherence, retention, and high dropout rates.

The current version (CANreduce 2.0) was designed to overcome these difficulties, with the implementation of additional adherence-focused guidance—feedback on demand and enhanced adherence monitoring mainly through motivational automated emails with weekly reminders, encouragement, and suggestions for further self-help module interventions, and the constant opportunity for participants to ask any questions they might have throughout their participation in the web-based program.

The 2 active study arm groups received the same level of enhanced support, but only study arm 1 received specific enhancements with a social presence, based on Mohr's supportive accountability model [37]. This involved more intimate, personally addressed emails, texts, and videos from a constantly visualized eCoach, with the intention of creating a more personalized atmosphere, greater alliance between the user and eCoach, and a more considerate and caring participant–counselor connection, and with no explicit need for a certified therapist as constant backup.

Modules, Dashboard, and Consumption Diary

The CANreduce 2.0 self-help intervention consists of 8 modules that encompass motivational interviewing techniques, traditional cognitive behavioral therapy, self-control practices, and social problem-solving [32]. Modules 1 and 2 are an introduction to the program and its application, helping program users to work through and identify individual triggers and triggering situations, so they can avoid unintentional cannabis use (motivational interviewing techniques [42], cognitive behavioral therapy approach to relapse prevention [43]). In Modules 3-5, skills and techniques are taught to enhance social relationships, restore sleeping patterns, deal with ruminations (behavioral activation approach [44]), and overcome situations considered risky for relapse, such as feelings of discouragement (cognitive behavioral therapy approach for relapse prevention [43] and to handle cravings [45]). By establishing rules and rituals and applying mindful positive thinking, self-talk, envisioning consequences, and distraction techniques, the program's aim is for users to persevere in their attempts to reduce their cannabis consumption. Modules 6-8 teach problem solving (social problem-solving approach [46]) and rejection skills to help users resist cannabis use that exceed their individual cannabis consumption plan (based on cognitive behavioral therapy [45]); to meet challenges in daily life, such as manifestations of depression; and, upon program completion, to provide them with the opportunity for a personalized review (motivational interviewing techniques [43]).

The advanced version of CANreduce (2.0) is particularly tailored to patients with a common mental disorder such as attention deficit/hyperactivity disorder, to enhance program adherence and achieve better outcomes via improved module content that focuses on dis–related problems. These include associated common mental disorder symptoms such as depressed mood, and problem-solving skills (Module 6), as well as excessive ruminations and poor sleeping habits (Module 3). In addition, to help participants with attention deficit/hyperactivity disorder who are known to be easily distracted, coping strategies (such as focusing on cravings and letting them come and go) are evaluated (Module 4).

The concept of adherence-focused guidance was implemented to help participants who screened positive for attention deficit/hyperactivity disorder focus on the tasks in the modules. As an element of adherence monitoring, automated emails for motivation and emails for information on feedback on demand were sent to the participants.

The web-based program has a dashboard as its starting point that includes an overview of all 8 modules, providing useful information on a chronological timeline with follow-ups and data acquisition. Program users are advised to fill in their consumption diary at least once weekly, wherein participants in both active study arms were asked to define their individual standardized joint (on the basis of 36 photographs with specific cannabis doses) and their cannabis reduction goals at baseline. Participants could choose from 6 different fictional companions to accompany them through the program by communicating written thoughts and questions for further encouragement and reflection.

Ethics and Data Protection

The study was conducted in accordance with the Declaration of Helsinki, the European Directive on medical devices 93/42/EEC, and the ISO Norm 14155 and ISO 14971 of Swiss Law and Swiss Regulatory Authority requirements [32]. The study was approved by the ethics committee of the Canton of Zurich (2016-00264) and is registered (ISRCTN11086185).

Outcome Measures and Instruments

Sociodemographic data that were obtained included sex, age, country of origin, and highest level of education. Baseline characteristics pertaining to substance use included the number of cannabis joints consumed over the preceding 7 days, years of cannabis use, age at which regular cannabis use started, years of cannabis use, and any other substances consumed over the 30 days immediately preceding the study.

The study's primary outcome was change in the number of days in which cannabis was consumed over the preceding 30 days according to the timeline follow-back method [47], which was compared between baseline and the 3-month follow-up assessment. Secondary outcomes were changes in the Severity of Dependence Scale (SDS) score, which can range from 0 to 15, with a score >4 indicating cannabis dependence [48]; cannabis use disorder identification test score (CUDIT), which can range from 0 to 40, with a score ≥ 8 indicating hazardous cannabis use and a score ≥12 indicating possible cannabis use disorder [49]; symptoms (cut off score >14) reported using the ADHD Self-Report scale (ASRS) version 1.1, which can range from 0 to 24 [3]; depression cut off score >16, using Center for Epidemiologic Studies Depression (CES-D), which can range from 0 to 60 [50]; anxiety, using the Generalized Anxiety Disorder scale (GAD) cut off score >10, which can range from 0 to 21 [51]; and Posttraumatic Stress Disorder short screening scale score, which can range from 7 to 28, with a cutoff sum score ≥4 suggesting posttraumatic stress disorder. Additional outcomes of interest were participant retention throughout the course of the study (defined as each person's weekly diary completion rate), level of adherence to the intervention (defined as the number of modules each person completed), and any perceived adverse effects that participants attributed to the program. Study outcomes are described in greater detail elsewhere [31].

Statistical Analysis

Sociodemographic parameters and baseline clinical characteristics were compared between those screening positive and negative for attention deficit/hyperactivity disorder, using Pearson chi-square analysis for categorical variables, analysis

of variance for continuous variables, and the Kruskal-Wallis H test for ordinal variables.

Main outcomes of interest were compared between baseline and 3-month follow-up using paired *t* tests within each group—individuals with attention deficit/hyperactivity disorder (screened positive) and individuals without attention deficit/hyperactivity disorder (screened negative). An ASRS score >13 was considered attention deficit/hyperactivity disorder positive. Linear regression analyses were conducted with attention deficit/hyperactivity disorder screening group allocation as the predictor variable, controlled for the baseline value of the respective outcome variable to compare the individuals with attention deficit/hyperactivity disorder and individuals without attention deficit/hyperactivity disorder screen groups. To identify associations between attention deficit/hyperactivity disorder symptom severity and reductions in the primary and secondary outcomes after the intervention, Spearman rank correlation coefficients were calculated. The Fisher exact (2-tailed) test was used to compare the occurrence of adverse effects between individuals who screened positive and individuals who screened negative for attention deficit/hyperactivity disorder. Rates of retention, adherence, and perceived adverse effects were compared between individuals with attention deficit/hyperactivity disorder and individuals without attention deficit/hyperactivity disorder screen using 2-tailed paired *t* tests. We used intention-to-treat analysis. Missing values were imputed by means of chained equations with 20 sets of imputations [32]. The criterion for statistical significance was $P<.05$, and all inferential testing was 2-tailed.

Results

Baselines Characteristics of the Study Participants

Participants were predominantly male (263/367, 71.6%). The mean age was 27.9 years old (SD 7.5 years). No significant differences in age ($P=.15$) or the highest level of education ($P=.36$) between individuals with and without attention deficit/hyperactivity disorder were detected. The largest percentage of participants were from Switzerland (140/367, 38.1%), followed by Austria (134/367; 36.5%) and Germany (91/367, 24.7%).

All participants screened higher than the cutoff value (CUDIT ≥8) for cannabis use disorder. In the group with attention deficit/hyperactivity disorder, both the CUDIT and the SDS score were significantly higher ($P<.001$) than those in individuals screening negative. Scale scores for common mental disorder were also significantly higher (anxiety: $P<.001$; depression: $P<.001$; posttraumatic stress disorder: $P=.005$) in individuals with attention deficit/hyperactivity disorder than in those without.

The mean age of participants when they first started using cannabis was 20.0 years (SD 5.3 years), with a mean 7.9 years (SD 6.7 years) since the start of cannabis use. There were no statistically significant differences between groups for either starting age ($P=.45$) or duration of use ($P=.31$). Over the 7 days preceding the CANreduce 2.0 intervention, participants smoked a mean of 21.9 standard joints (SD 15.8 joints) per week. All participants reported cannabis consumption within the preceding 30 days, and 37% (134/367) reported risky alcohol use, which was defined as the consumption of 5 or more standard drinks on a single occasion (Table 2).

Table 2. Baseline characteristics of the study participants, by group and overall.

Characteristic	With attention deficit/hyperactivity disorder (n=94)	Without attention deficit/hyperactivity disorder (n=273)	All (n=367)	*F* test (*df1*,*df2*) or chi-square (*df*)[a]	*P* value
Sex, n (%)				0.58 (1)[a]	.45
Female	30 (31.9)	74 (27.1)	104 (28.3)		
Male	64 (68.0)	199 (72.8)	263 (71.6)		
Age, mean (SD)	26.9 (7.5)	28.2 (7.5)	27.9 (7.5)	2.10 (1,365)	.15
Highest education level, n (%)				13.16 (5)[a]	.36
Compulsory school	10 (10.6)	16 (5.8)	26 (7.0)		
Apprenticeship	17 (18.0)	46 (16.8)	63 (17.1)		
Middle school	35 (37.2)	79 (28.9)	114 (31.0)		
Higher professional education	11 (11.7)	44 (16.1)	55 (14.9)		
University	17 (18.0)	76 (27.8)	93 (25.3)		
Not stated	4 (4.2)	12 (4.3)	16 (4.3)		
Origin, n (%)				0.82 (2)[a]	.66
Switzerland	39 (41.4)	101 (36.9)	140 (38.1)		
Germany	24 (25.5)	67 (24.5)	91 (24.7)		
Austria	31 (32.9)	103 (37.7)	134 (36.5)		
Unknown	0 (0)	2 (0.7)	2 (0.5)		
Number of cannabis joints in preceding 7 days, mean (SD)	24.0 (14.4)	21.2 (16.2)	21.9 (15.8)	2.16 (1,365)	.14
Number of days cannabis (≥1 joint) was consumed in preceding 30 days, mean (SD)	26.6 (4.9	25.0 (6.4)	25.4 (6.1)	4.76 (1,365)	.03
Duration cannabis use (in years), mean (SD)	7.3 (6.1)	8.1 (6.9)	7.9 (6.7)	1.03(1,365)	.31
Age of onset of regular cannabis use (in years), mean (SD)	19.6 (5.5)	20.1 (5.2)	20.0 (5.3)	0.58 (1,365)	.45
Cannabis Use Disorder Identification Test score, mean (SD)	23.7 (4.6)	19.6 (5.6)	20.6 (5.7)	41.10 (1,365)	<.001
Adult ADHD Self-Report scale score, mean (SD)	15.8 (1.9)	9.0 (2.9)	10.7 (4.0)	438.60 (1,365)	<.001
Severity of Dependence Scale score, mean (SD)	9.1 (3.0)	7.1 (3.1)	7.6 (3.2)	31.01 (1,365)	<.001
Center for Epidemiological Studies Depression scale score, mean (SD)	27.4 (9.1)	19.8 (10.5)	21.7 (10.7)	39.40 (1,365)	<.001
Generalized Anxiety Disorder scale score, mean (SD)	10.7 (4.9)	6.6 (4.5)	7.6 (4.9)	55.90 (1,364)	<.001
Posttraumatic Stress Disorder Short Screening scale score, mean (SD)	14.9 (4.5)	12.2 (4.8)	12.9 (4.8)	8.18 (1,118)	.005
Years of substance use					
Cannabis, mean (SD)	7.3 (6.1)	8.1 (6.9)	7.9 (6.7)	1.03 (1,365)	.31
Alcohol, mean (SD)	4.5 (6.2)	5.3 (6.5)	5.1 (6.4)	1.08 (1,332)	.30
Alcohol, hazardous consumption[b], mean (SD)	1.4 (3.1)	1.8 (4.0)	1.7 (3.8)	0.66 (1,318)	.42
Cocaine, mean (SD)	1.1 (3.8)	0.2 (1.1)	0.4 (2.1)	11.15 (1,316)	<.001
Substance use in last 30 days, n (%)					
Cannabis	94 (100)	273 (100)	367 (100)	N/A[c]	N/A
Alcohol	74 (78.7)	213 (78.0)	287 (78.2)	0.26 (1)[a]	.61
Alcohol, hazardous consumption	35 (37.2)	99 (36.2)	134 (36.5)	0.38 (1)[a]	.54
Tranquilizer	8 (8.5)	16 (5.8)	24 (6.5)	0.73 (1)[a]	.39

Characteristic	With attention deficit/hyperactivity disorder (n=94)	Without attention deficit/hyperactivity disorder (n=273)	All (n=367)	*F* test (*df1*,*df2*) or chi-square (*df*)[a]	*P* value
Cocaine	18 (19.1)	32 (11.7)	50 (13.6)	2.73 (1)[a]	.10
Amphetamines	23 (24.4)	49 (17.9)	72 (19.6)	2.05 (1)[a]	.15
Hallucinogens	3 (3.1)	22 (8.0)	25 (6.8)	1.44 (1)[a]	.23
Heroin	0 (0)	0 (0)	0 (0)	N/A	N/A
Methadone	0 (0)	1 (0)	1 (0.2)	N/A	N/A
Other substances	5 (5.3)	8 (2.9)	13 (3.5)	0.79 (1)[a]	.37

[a]A chi-square test was used where indicated.

[b]Hazardous alcohol consumption is defined as 5 or more standard drinks (50 mL of spirits, 150-200 mL of wine, or 330-450 mL of beer) per day at least 3 days a week [52].

[c]N/A: not applicable.

Main Outcomes

Both individuals with attention deficit/hyperactivity disorder (mean difference 11.53, SD 21.87; *P*<.001) and individuals without attention deficit/hyperactivity disorder (mean difference 8.53, SD 9.4; *P*<.001) reported a significant reduction in days in the preceding month on which ≥1 joint was consumed. Likewise, a significant decrease was apparent in both the reported severity of dependence score (with: mean difference 3.57, SD 3.65, *P*<.001; without: mean difference 2.47, SD 3.39, *P*<.001) and cannabis substance use disorder score (with: mean difference 6.38, SD 5.96, *P*<.001; without: mean difference 5.33, SD 6.05, *P*<.001).

A similar pattern was evident for both anxiety (GAD score—with: mean difference 4.31, SD 4.71, *P*<.001; without: mean difference 1.84, SD 4.22, *P*<.001), and depression (CES-D score—with: mean difference 10.25, SD 10.54, *P*<.001; without: mean difference 4.39, SD 10.22, *P*<.001). However, the decrease in ASRS score was significant for the group with attention deficit/hyperactivity disorder (mean difference 4.65, SD 4.44; *P*<.001) but that for the group without attention deficit/hyperactivity disorder was not (mean difference 0.83, SD 4.10; *P*=.19).

There were no significant differences in change in use (*P*=.08), dependence (*P*=.95), use disorder (*P*=.85), attention deficit/hyperactivity disorder status (*P*=.84), depression (*P*=.84), and anxiety (*P*=.26) between baseline and final follow-up, dependent on ASRS score >14 (ie, having attention deficit/hyperactivity disorder) (Table 3 and Table 4).

Attention deficit/hyperactivity disorder symptom severity at baseline was not associated with reduced cannabis consumption (*P*=.19), severity of dependence (*P*=.14), or cannabis use disorder scores (*P*=.69); however, more severe attention deficit/hyperactivity disorder symptoms at baseline were associated with greater reductions in depression (*P*<.001) and anxiety (*P*<.001) after the intervention (Table 5).

Table 3. Comparison between outcomes at baseline and at the 3-month follow-up (intention-to-treat analysis)

Outcome	Without attention deficit/hyperactivity disorder (n=273)				With attention deficit/hyperactivity disorder (n=94)			
	Baseline, mean (SD)	Follow-up, mean (SD)	Effect size, Cohen *d*	95% CI	Baseline, mean (SD)	Follow-up, mean (SD)	Effect size, Cohen *d*	95% CI
Number of days cannabis (≥1 joint) was consumed in preceding 30 days	25.00 (6.40)	16.49 (10.04)	1.01	0.70 to 1.31	26.59 (4.93)	15.05 (9.76)	0.77	0.59 to 0.94
Severity of Dependence Scale	7.10 (3.06)	4.63 (3.02)	0.81	0.51 to 1.11	9.13 (3.02)	5.55 (2.86)	0.65	0.47 to 0.82
Cannabis Use Disorder Identification Test	19.59 (5.65)	14.26 (6.40)	0.88	0.58 to 1.18	23.72 (4.61)	17.34 (5.87)	0.50	0.33 to 0.67
Adult ADHD Self-Report scale	9.00 (2.95)	8.17 (4.01)	0.24	−0.05 to 0.52	15.80 (1.88)	11.15 (4.40)	0.58	0.41 to 0.75
Center for Epidemiological Studies Depression scale	19.77 (10.53)	15.38 (9.96)	0.59	0.30 to 0.88	27.41 (9.06)	17.16 (10.03)	0.75	0.57 to 0.92
Generalized Anxiety Disorder scale	6.57 (4.49)	4.71 (3.53)	0.46	0.17 to 0.75	10.68 (4.87)	6.37 (3.56)	0.76	0.59 to 0.93

Table 4. Regression analysis of changes in cannabis consumption, attention deficit/hyperactivity disorder, depression, and anxiety 3 months after initiating the intervention (intention-to-treat analysis, n=282)

Outcome variable	*B* (SE)	95% CI	*t* value	*P* value
Number of days cannabis (≥1 joint) was consumed in preceding 30 days	−3.72 (2.14)	−7.18 to 8.58	−1.73	.08
Severity of Dependence Scale	0.03 (0.68)	−1.31 to 1.39	0.05	.95
Cannabis Use Disorder Identification Test	0.26 (1.41)	−1.48 to 7.16	0.18	.85
Adult ADHD Self-Report scale	0.20 (1.03)	1.38 to 5.54	0.20	.84
Center for Epidemiological Studies Depression Scale	0.20 (1.03)	1.38 to 5.54	0.20	.84
Generalized Anxiety Disorder scale	0.89 (0.80)	1.24 to 3.40	1.11	.26

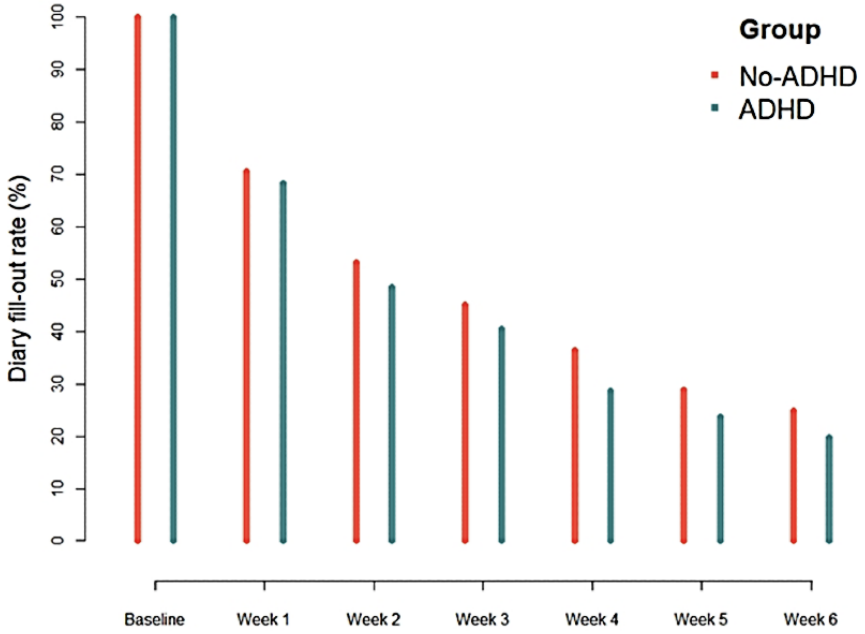
Table 5. Spearman rank correlation coefficients for baseline attention deficit/hyperactivity disorder symptom severity versus primary and secondary outcome change scores.

Outcome variable	Spearman ρ	95% CI	*t* value	*P* value
Number of days cannabis (≥1 joint) was consumed in preceding 30 days	0.14	−0.07 to 0.33	1.31	.19
Severity of Dependence Scale	0.11	−0.04 to 0.26	1.49	.14
Cannabis Use Disorder Identification Test	0.04	−0.16 to 0.24	0.39	.69
Center for Epidemiological Studies Depression Scale	0.33	0.19 to 0.46	4.33	<.001
Generalized Anxiety Disorder scale	0.28	−0.13 to 0.42	3.57	<.001

Retention

Figure 1 shows the rate at which participants made entries into the consumption diary from week 1 through week 6 after baseline, and at the final follow-up assessment for both study groups. A significant intergroup difference was apparent at the final 3-month follow-up assessment, with a lower percentage of individuals having filled out the diary in the individuals with attention deficit/hyperactivity disorder than the individuals without attention deficit/hyperactivity disorder screen group (χ_1^2=5.21, *P*=.02). Overall, there was no significant association between baseline attention deficit/hyperactivity disorder symptom severity and retention rate (ρ=−0.10, *P*=.13).

Figure 1. Retention throughout the study period of participants who screened positive for attention deficit/hyperactivity disorder (ADHD) (blue) versus those who screened negative (red).



Participant Adherence

There was no statistically significant difference ($t_{159.27}$=0.96, *P*=.34) in the number of modules completed among those screened positive for attention deficit/hyperactivity disorder (mean 2.10, SD 2.33) and those who screened negative (mean 2.36, SD 2.36). There also was no significant association between the magnitude of decrease in attention deficit/hyperactivity disorder symptoms and the number of completed modules (ρ=−0.09, *P*=.43).

Safety

Of 55 individuals who completed the questionnaire on adverse intervention effects, 44 (80%) answered that they had not experienced any negative effects during the study, while 7 people (12.7%) answered that an adverse effect had affected them somewhat negatively, 3 people (5.4%) answered that an adverse effect had affected them quite negatively, and 1 person (1.8%) an adverse effect had affected them to a great extent. There was no significant difference in the percentage of individuals screening positive with attention deficit/hyperactivity disorder and individuals screening negative with attention deficit/hyperactivity disorder who reported adverse effects (P=.33).

Discussion

Principal Results

In this study, we aimed to evaluate whether the CANreduce 2.0 program can reduce cannabis use in adults who screen either positive or negative for attention deficit/hyperactivity disorder, and whether individuals screening positive might benefit more from the program than those screening negative. Furthermore, we aimed to determine whether individuals with a positive attention deficit/hyperactivity disorder screen and more severe attention deficit/hyperactivity disorder symptoms might benefit most from the CANreduce 2.0 program. The study's main finding was that participation in the CANreduce 2.0 program reduced cannabis consumption from baseline to follow-up, both among individuals who screened positive and individuals who screened negative for attention deficit/hyperactivity disorder. Both SDS scores and CUDIT scores, indicating that the severity of dependence (both groups: P<.001) and frequency of cannabis use (both groups: P<.001), respectively, also were significantly lower after participation in the program; however, no significant differences (SDS: P=.14; CUDIT: P=.69) in the magnitude of reduction in these scores were apparent between participants who screened positive for attention deficit/hyperactivity disorder and those who screened negative. Similarly, psychological comorbidities—such as anxiety and depression—also improved after program participation, with significant changes observed in both the attention deficit/hyperactivity disorder positive and negative screening groups (P<.001). Participants with more severe attention deficit/hyperactivity disorder symptoms at baseline exhibited a greater reduction in depression and anxiety than those with milder symptoms. The rate of retention was significantly less (P=.02) in the individuals with attention deficit/hyperactivity disorder group at the end of the study period 3 months after starting the program, while attention deficit/hyperactivity disorder symptom severity was not significantly associated with the retention rate (P=.13). Those screening positive for attention deficit/hyperactivity disorder and those screening negative did not significantly differ in their adherence to the program (P=.43) or the number reporting adverse events (P=.33).

Although cannabis is a widely used psychoactive substance that may induce dependency and cause associated problems, many users do not seek help at outpatient addiction centers. Reasons for this include the relative inaccessibility of treatment centers, fear of stigmatization, and inadequate awareness that treatment is needed [40]. Web-based self-help programs have yielded beneficial results for alcohol and tobacco users [53-56], but information on the efficacy of such programs for cannabis users is limited. Users with comorbid psychiatric diseases, such as attention deficit/hyperactivity disorder, might particularly benefit from web-based self-help with adherence-focused guidance, as their attention deficit/hyperactivity disorder could innately aggravate their ability to adhere to a program lacking such support. However, no significant differences between the 2 screening groups in our study (those screening positive vs negative for attention deficit/hyperactivity disorder) were evident, in terms of reducing cannabis consumption (P=.08), the severity of dependence (P=.95), cannabis substance use disorder (P=.85), anxiety (P=.26), or symptoms of depression (P=.84), as all these outcomes were reduced similarly in the 2 groups. Hence, the program appears to provide similar benefits to cannabis users who screen positive and those who screen negative for attention deficit/hyperactivity disorder, which is contrary to our expectation of a particular benefit for users with attention deficit/hyperactivity disorder.

Participants who screened positive for attention deficit/hyperactivity disorder filled out the consumption diary to a lesser extent over the course of follow-up. It is, therefore, possible that participants screening positive for attention deficit/hyperactivity disorder used the intervention less than those without attention deficit/hyperactivity disorder over time and, as such, the program's effects might have been blunted in that group. At the same time, there were no differences in adherence between the 2 groups, which means that, over the course of the intervention, those screening positive and negative for attention deficit/hyperactivity disorder participated in the program to roughly the same degree. This is an encouraging outcome, as it potentially indicates that individuals with attention deficit/hyperactivity disorder can, indeed, participate in and adhere to a web-based remote intervention such as CANreduce 2.0. However, it was our assumption that those with more attention deficit/hyperactivity disorder symptoms would profit most from adherence-focused guidance. This assumption of ours was based on the beneficial effect of guidance in internet-based interventions [57] and the known deficits of individuals with attention deficit/hyperactivity disorder regarding impulse inhibition, working memory, organization, and planning skills [58].

Unfortunately, no conclusion can be drawn regarding possible withdrawal symptoms in our study; we did not assess withdrawal symptoms because we had previously found [50] that the vast majority of persons reduced their cannabis use slowly and gradually, and probably with no or only occasionally very mild withdrawal symptoms occurred.

In our study sample, we observed differences in the completion rates for individual modules in the CANreduce 2.0 program between participants who screened positive versus negative for attention deficit/hyperactivity disorder. Baseline differences between these groups in their CUDIT, ASRS, SDS, CES-D, and GAD-7 scores might explain these differences, given that individuals screening positive for attention deficit/hyperactivity

disorder entered the program with higher scores for all these measures.

Overall, the only significant difference between those with attention deficit/hyperactivity disorder and those without attention deficit/hyperactivity disorder was the reduction in mean attention deficit/hyperactivity disorder score between baseline and final follow-up ($P<.001$); however, individuals with attention deficit/hyperactivity disorder had higher initial attention deficit/hyperactivity disorder severity scores, which could explain this difference in the degree of improvement. That the severity of attention deficit/hyperactivity disorder-specific symptoms was reportedly reduced after the intervention is highly relevant, because it indicates that a cannabis-specific intervention can lead to reduced attention deficit/hyperactivity disorder severity. However, this might be a cannabis dose–dependent and not a simple direct effect, and verifying this would require controlled pharmacological research. That attention deficit/hyperactivity disorder severity at baseline was inversely correlated with the extent of change in comorbid anxiety and depression indicated that potential effects of the program might have been overlooked, as a result of the allocation of participants based on predefined threshold of symptoms.

The mean number of completed modules was very low for both, individuals who screened positive for attention deficit/hyperactivity disorder severity (mean 2.10, SD 2.33) and individuals who screened negative for attention deficit/hyperactivity disorder severity (mean 2.36, SD 2.36). This number is low because approximately one-third of participants only logged in briefly, worked through the first module, and never logged in again. These early dropouts are a common problem in web-based interventions [59], especially with interventions that set the hurdles for study participation low (exclusively web-based recruitment, broad inclusion criteria, and only a web-based baseline questionnaire).

On the other hand, a sizeable proportion of participants stayed in the program for the first 2 weeks and then worked through a large number of the modules. Nevertheless, efforts should continue to find ways to reduce the high dropout rates experienced with this web-based program. The next steps to improve the program could be, for example, identifying through qualitative interviews the background and evaluations of the program by participants screening positive versus those screening negative for attention deficit/hyperactivity disorder.

Limitations

This study has certain limitations that must be considered. First, there were baseline differences between the 2 study groups, including a higher number of cannabis use days in those screening positive for attention deficit/hyperactivity disorder. Second, the follow-up duration of 3 months might have been too short to establish any long-term effectiveness of the CANreduce 2.0 program; a longer period of follow-up could provide insights into how well program users retain whatever benefits they appear to achieve from the program. Third, the influence of personalized counseling on the observed outcome, being just one component of the program, cannot be ascertained. Fourth, formally diagnosing attention deficit/hyperactivity disorder requires at least one detailed face-to-face assessment; but given that this study targeted individuals who overuse cannabis in the general population who were not otherwise in treatment for their cannabis use, we had no choice but to rely on a web-based attention deficit/hyperactivity disorder screening instrument and no actual person-to-person contact. Fifth, potential, unrecognized confounders might have resulted from individuals not being randomly assigned to the 2 study groups of interest—those screening positive versus negative for attention deficit/hyperactivity disorder—and any one of these potential confounders could have biased results. Sixth, our analyses were limited by a sample size that was powered for the main study and not specifically for comparisons of those screening positive versus negative for attention deficit/hyperactivity disorder. Seventh, 2 items in the ASRS screening tool, which we used to measure attention deficit/hyperactivity disorder symptoms, might be prone to elevated scores in heavy cannabis users who do not suffer from attention deficit/hyperactivity disorder. False-positive results could, thereby, result from potential overlap of symptoms between attention deficit/hyperactivity disorder and cannabis use. Specifically, this refers to the items measuring appointment or obligation forgetfulness and task avoidance or delay. Furthermore, our decision to apply the ASRS [60] to measure attention deficit/hyperactivity disorder symptoms was based on our need to keep the assessment time brief. However, more extensive scales exist such as the KATE [61] or CAARS [62] would have most likely provided more reliable results.

Conclusions

The CANreduce 2.0 program appeared to benefit both individuals with and without attention deficit/hyperactivity disorder similarly. This web-based program offers a personalized counselling option designed to increase intervention adherence and may provide a viable option to reach cannabis users remotely, including those with attention deficit/hyperactivity disorder, and improve both psychiatric comorbidities and overall condition.

Acknowledgments

The study was funded by the Swiss National Science Foundation (grant 10001C_192301). We wish to thank Claudia Witt (Director of the Institute for Complementary and Integrative Medicine, University Hospital Zurich, Switzerland) for her assistance with the manuscript. This study would not have been possible without her support. Thanks also go to Kevin White, MD, PhD, for professional proofreading.

Conflicts of Interest

DE has been a consultant for Sanofi, Novartis, Minddistrict, Lantern, Schoen Kliniken, Ideamed. BARMER, Techniker Krankenkasse, and federal chambers for psychotherapy. DE is a stakeholder of HelloBetter.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 6422 KB - jmir_v24i4e30138_app1.pdf]

References

1. Polanczyk G, de Lima MS, Horta BL, Biederman J, Rohde LA. The worldwide prevalence of ADHD: a systematic review and metaregression analysis. Am J Psychiatry 2007 Jun;164(6):942-948. [doi: 10.1176/ajp.2007.164.6.942] [Medline: 17541055]
2. Polanczyk GV, Salum GA, Sugaya LS, Caye A, Rohde LA. Annual research review: a meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. J Child Psychol Psychiatry 2015 Mar 03;56(3):345-365. [doi: 10.1111/jcpp.12381] [Medline: 25649325]
3. Kessler RC, Adler L, Ames M, Demler O, Faraone S, Hiripi E, et al. The World Health Organization adult ADHD self-report scale (ASRS): a short screening scale for use in the general population. Psychol Med 2005 Feb 21;35(2):245-256. [doi: 10.1017/s0033291704002892] [Medline: 15841682]
4. Vogel M, Nordt C, Bitar R, Boesch L, Walter M, Seifritz E, et al. Cannabis use in Switzerland 2015-2045: a population survey based model. Int J Drug Policy 2019 Jul;69:55-59. [doi: 10.1016/j.drugpo.2019.03.008] [Medline: 31030906]
5. European drug report 2018: trends and developments. European Monitoring Centre for Drugs and Addiction. 2018. URL: https://www.emcdda.europa.eu/publications/edr/trends-developments/2018_en [accessed 2021-05-04]
6. 2015 national survey on drug use and health: detailed tables. Substance Abuse and Mental Health Services Administration. 2015. URL: https://www.samhsa.gov/data/sites/default/files/NSDUH-DetTabs-2015/NSDUH-DetTabs-2015/NSDUH-DetTabs-2015.pdf [accessed 2021-05-04]
7. Estévez N, Dey M, Eich-Höchli D, Foster S, Gmel G, Mohler-Kuo M. Adult attention-deficit/hyperactivity disorder and its association with substance use and substance use disorders in young men. Epidemiol Psychiatr Sci 2016 Jun;25(3):255-266 [FREE Full text] [doi: 10.1017/S2045796015000360] [Medline: 25989844]
8. Hurd YL, Michaelides M, Miller ML, Jutras-Aswad D. Trajectory of adolescent cannabis use on addiction vulnerability. Neuropharmacology 2014 Jan;76 Pt B:416-424 [FREE Full text] [doi: 10.1016/j.neuropharm.2013.07.028] [Medline: 23954491]
9. Deligianni, Studer, Daeppen, Gmel, Bertholet. Longitudinal associations between life satisfaction and cannabis use initiation, cessation, and disorder symptom severity in a cohort of young Swiss men. Int J Environ Res Public Health 2019 Apr 16;16(8):1372 [FREE Full text] [doi: 10.3390/ijerph16081372] [Medline: 30995797]
10. Rougemont-Bücking A, Jungaberle H, Scheidegger M, Merlo MC, Grazioli VS, Daeppen J, et al. Comparing mental health across Groups of users of psychedelics, MDMA, psychostimulants, and cannabis. J Psychoactive Drugs 2019;51(3):236-246. [doi: 10.1080/02791072.2019.1571258] [Medline: 30836844]
11. Henchoz Y, N'Goran AA, Baggio S, Deline S, Studer J, Gmel G. Associations of age at cannabis first use and later substance abuse with mental health and depression in young men. J Subst Use 2014 Oct 07;21(1):85-91. [doi: 10.3109/14659891.2014.966342]
12. Copeland J, Rooke S, Swift W. Changes in cannabis use among young people: impact on mental health. Curr Opin Psychiatry 2013 Jul;26(4):325-329. [doi: 10.1097/YCO.0b013e328361eae5] [Medline: 23689549]
13. Fergusson D, Boden J. Cannabis use and later life outcomes. Addiction 2008 Jun;103(6):969-976. [doi: 10.1111/j.1360-0443.2008.02221.x] [Medline: 18482420]
14. Weissenberger S, Ptacek R, Vnukova M, Raboch J, Klicperova-Baker M, Domkarova L, et al. ADHD and lifestyle habits in Czech adults, a national sample. Neuropsychiatr Dis Treat 2018;14:293-299 [FREE Full text] [doi: 10.2147/NDT.S148921] [Medline: 29391802]
15. Martínez-Luna N, Daigre C, Palma-Álvarez F, Perea-Ortueta M, Grau-López L, Roncero C, et al. Psychiatric comorbidity and addiction severity differences in patients with ADHD seeking treatment for cannabis or cocaine use disordersPsychiatric Comorbidity and Addiction Severity Differences in Patients With ADHD Seeking Treatment for Cannabis or Cocaine Use Disorders. J Atten Disord 2021 May 24;25(7):978-988. [doi: 10.1177/1087054719875787] [Medline: 31550967]
16. Romo L, Ladner J, Kotbagi G, Morvan Y, Saleh D, Tavolacci MP, et al. Attention-deficit hyperactivity disorder and addictions (substance and behavioral): prevalence and characteristics in a multicenter study in France. J Behav Addict 2018 Sep 01;7(3):743-751 [FREE Full text] [doi: 10.1556/2006.7.2018.58] [Medline: 30010409]
17. Maric M, Bexkens A, Bögels SM. Is clinical anxiety a risk or a protective factor for executive functioning in youth with ADHD? a meta-regression analysis. Clin Child Fam Psychol Rev 2018 Sep 26;21(3):340-353. [doi: 10.1007/s10567-018-0255-8] [Medline: 29484581]

18. Meinzer MC, Pettit JW, Viswesvaran C. The co-occurrence of attention-deficit/hyperactivity disorder and unipolar depression in children and adolescents: a meta-analytic review. Clin Psychol Rev 2014 Dec;34(8):595-607. [doi: 10.1016/j.cpr.2014.10.002] [Medline: 25455624]
19. Soler Artigas M, Sánchez-Mora C, Rovira P, Richarte V, Garcia-Martínez I, Pagerols M, ADHD Group of the Psychiatric Genomics Consortium, International Cannabis Consortium, et al. Attention-deficit/hyperactivity disorder and lifetime cannabis use: genetic overlap and causality. Mol Psychiatry 2020 Oct;25(10):2493-2503 [FREE Full text] [doi: 10.1038/s41380-018-0339-3] [Medline: 30610198]
20. Knouse LE, Teller J, Brooks MA. Meta-analysis of cognitive-behavioral treatments for adult ADHD. J Consult Clin Psychol 2017 Jul;85(7):737-750. [doi: 10.1037/ccp0000216] [Medline: 28504540]
21. Benyakorn S, Riley SJ, Calub CA, Schweitzer JB. Current state and model for development of technology-based care for attention deficit hyperactivity disorder. Telemed J E Health 2016 Sep;22(9):761-768 [FREE Full text] [doi: 10.1089/tmj.2015.0169] [Medline: 26985703]
22. Pettersson R, Söderström S, Edlund-Söderström K, Nilsson KW. Internet-based cognitive behavioral therapy for adults with ADHD in outpatient psychiatric care. J Atten Disord 2017 Apr 28;21(6):508-521. [doi: 10.1177/1087054714539998] [Medline: 24970720]
23. Patel R, Patel P, Shah K, Kaur M, Mansuri Z, Makani R. Is cannabis use associated with the worst inpatient outcomes in attention deficit hyperactivity disorder adolescents? Cureus 2018 Jan 07;10(1):e2033 [FREE Full text] [doi: 10.7759/cureus.2033] [Medline: 29535906]
24. van Emmerik-van Oortmerssen K, Vedel E, Kramer FJ, Blankers M, Dekker JJ, van den Brink W, et al. Integrated cognitive behavioral therapy for ADHD in adult substance use disorder patients: results of a randomized clinical trial. Drug Alcohol Depend 2019 Apr 01;197:28-36. [doi: 10.1016/j.drugalcdep.2018.12.023] [Medline: 30769263]
25. Taylor CB, Luce KH. Computer- and internet-based psychotherapy interventions. Curr Dir Psychol Sci 2016 Jun 24;12(1):18-22. [doi: 10.1111/1467-8721.01214]
26. Wallace A, Wade N, Hatcher K, Lisdahl K. Effects of cannabis use and subclinical ADHD symptomology on attention based tasks in adolescents and young adults. Arch Clin Neuropsychol 2019 Jul 26;34(5):700-705 [FREE Full text] [doi: 10.1093/arclin/acy080] [Medline: 30295694]
27. Brandt A, Rehm J, Lev-Ran S. Clinical correlates of cannabis use among individuals with attention deficit hyperactivity disorder. J Nerv Ment Dis 2018 Sep;206(9):726-732. [doi: 10.1097/NMD.0000000000000877] [Medline: 30124577]
28. Mitchell JT, Sweitzer MM, Tunno AM, Kollins SH, McClernon FJ. "I use weed for my ADHD": a qualitative analysis of online forum discussions on cannabis use and ADHD. PLoS One 2016 May 26;11(5):e0156614 [FREE Full text] [doi: 10.1371/journal.pone.0156614] [Medline: 27227537]
29. Boumparis N, Loheide-Niesmann L, Blankers M, Ebert DD, Korf D, Schaub MP, et al. Short- and long-term effects of digital prevention and treatment interventions for cannabis use reduction: a systematic review and meta-analysis. Drug Alcohol Depend 2019 Jul 01;200:82-94 [FREE Full text] [doi: 10.1016/j.drugalcdep.2019.03.016] [Medline: 31112834]
30. Olmos A, Tirado-Muñoz J, Farré M, Torrens M. The efficacy of computerized interventions to reduce cannabis use: A systematic review and meta-analysis. Addict Behav 2018 Apr;79:52-60. [doi: 10.1016/j.addbeh.2017.11.045] [Medline: 29248863]
31. Tait RJ, Spijkerman R, Riper H. Internet and computer based interventions for cannabis use: a meta-analysis. Drug Alcohol Depend 2013 Dec 01;133(2):295-304. [doi: 10.1016/j.drugalcdep.2013.05.012] [Medline: 23747236]
32. Amann M, Haug S, Wenger A, Baumgartner C, Ebert DD, Berger T, et al. The effects of social presence on adherence-focused guidance in problematic cannabis users: protocol for the CANreduce 2.0 randomized controlled trial. JMIR Res Protoc 2018 Jan 31;7(1):e30 [FREE Full text] [doi: 10.2196/resprot.9484] [Medline: 29386176]
33. Baumgartner C, Schaub MP, Wenger A, Malischnig D, Augsburger M, Walter M, et al. CANreduce 2.0 adherence-focused guidance for internet self-help among cannabis users: three-arm randomized controlled trial. J Med Internet Res 2021 Apr 30;23(4):e27463 [FREE Full text] [doi: 10.2196/27463] [Medline: 33929333]
34. Petker T, DeJesus J, Lee A, Gillard J, Owens MM, Balodis I, et al. Cannabis use, cognitive performance, and symptoms of attention deficit/hyperactivity disorder in community adults. Exp Clin Psychopharmacol 2020 Dec;28(6):638-648. [doi: 10.1037/pha0000354] [Medline: 32105137]
35. Kollins SH, Schoenfelder EN, English JS, Holdaway A, Van Voorhees E, O'Brien BR, et al. An exploratory study of the combined effects of orally administered methylphenidate and delta-9-tetrahydrocannabinol (THC) on cardiovascular function, subjective effects, and performance in healthy adults. J Subst Abuse Treat 2015 Jan;48(1):96-103 [FREE Full text] [doi: 10.1016/j.jsat.2014.07.014] [Medline: 25175495]
36. Volkow ND, Wang G, Telang F, Fowler JS, Alexoff D, Logan J, et al. Decreased dopamine brain reactivity in marijuana abusers is associated with negative emotionality and addiction severity. Proc Natl Acad Sci U S A 2014 Jul 29;111(30):E3149-E3156 [FREE Full text] [doi: 10.1073/pnas.1411228111] [Medline: 25024177]
37. 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]
38. CANreduce: Switzerland. URL: https://www.canreduce.ch [accessed 2021-05-04]

39. CANreduce: Austria. URL: https://www.canreduce.at [accessed 2021-05-04]
40. Schaub MP, Haug S, Wenger A, Berg O, Sullivan R, Beck T, et al. Can reduce--the effects of chat-counseling and web-based self-help, web-based self-help alone and a waiting list control program on cannabis use in problematic cannabis users: a randomized controlled trial. BMC Psychiatry 2013 Nov 14;13:305 [FREE Full text] [doi: 10.1186/1471-244X-13-305] [Medline: 24228630]
41. Schaub MP, Wenger A, Berg O, Beck T, Stark L, Buehler E, et al. A web-based self-help intervention with and without chat counseling to reduce cannabis use in problematic cannabis users: Three-arm randomized controlled trial. J Med Internet Res 2015 Oct 13;17(10):e232 [FREE Full text] [doi: 10.2196/jmir.4860] [Medline: 26462848]
42. Miller W, Rollnick S. Motivational Interviewing: Helping People Change. New York: Guilford Press; 2012.
43. Marlatt GA, Donovan DM. Relapse Prevention: Maintenance Strategies in the Treatment of Addictive Behaviors. 2nd ed. New York: Guilford Press; 2005:1222-1223.
44. Kanter JW, Puspitasari AJ, Santos MM, Nagy GA. Behavioural activation: history, evidence and promise. Br J Psychiatry 2012 May 01;200(5):361-363. [doi: 10.1192/bjp.bp.111.103390] [Medline: 22550329]
45. Steinberg K, Roffman R, Carroll K, McRee B, Babor T, Miller M. Brief Counseling for Marijuana Dependence: A Manual for Treating Adults. Rockville, Maryland: enter for Substance Abuse Treatment, Substance Abuse and Mental Health Services Administration, US Department of Health and Human Services; 2005.
46. Jaffee WB, D'Zurilla TJ. Personality, problem solving, and adolescent substance use. Behavior Therapy 2009;40(1):93-101. [doi: 10.1016/j.beth.2008.03.001] [Medline: 19187820]
47. Buuren SV, Groothuis-Oudshoorn K. MICE: multivariate imputation by chained equations in R. J Stat Soft 2011;45(3):1-67. [doi: 10.18637/jss.v045.i03]
48. Gossop M, Darke S, Griffiths P, Hando J, Powis B, Hall W, et al. The severity of dependence scale (SDS): psychometric properties of the SDS in English and Australian samples of heroin, cocaine and amphetamine users. Addiction 1995 May;90(5):607-614. [doi: 10.1046/j.1360-0443.1995.9056072.x] [Medline: 7795497]
49. Adamson SJ, Sellman JD. A prototype screening instrument for cannabis use disorder: the cannabis use disorders identification test (CUDIT) in an alcohol-dependent clinical sample. Drug Alcohol Rev 2003 Sep;22(3):309-315. [doi: 10.1080/0959523031000154454] [Medline: 15385225]
50. Radloff LS. The CES-D scale: A self-report depression scale for research in the general population. Appl Psychol Meas 1977;1(3):385-401. [doi: 10.1177/014662167700100306]
51. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
52. Hartz SM, Oehlert M, Horton AC, Grucza RA, Fisher SL, Culverhouse RC, et al. Daily drinking is associated with increased mortality. Alcohol Clin Exp Res 2018 Nov;42(11):2246-2255 [FREE Full text] [doi: 10.1111/acer.13886] [Medline: 30281161]
53. Civljak M, Sheikh A, Stead LF, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2010 Sep 08(11):1-58. [doi: 10.1002/14651858.CD007078.pub3] [Medline: 20824856]
54. Haug S, Meyer C, John U. Efficacy of an internet program for smoking cessation during and after inpatient rehabilitation treatment: a quasi-randomized controlled trial. Addict Behav 2011 Dec;36(12):1369-1372. [doi: 10.1016/j.addbeh.2011.08.004] [Medline: 21907496]
55. Riper H, Spek V, Boon B, Conijn B, Kramer J, Martin-Abello K, et al. Effectiveness of e-self-help interventions for curbing adult problem drinking: a meta-analysis. J Med Internet Res 2011 Jun 30;13(2):e42. [doi: 10.2196/jmir.1691] [Medline: 21719411]
56. Rooke S, Thorsteinsson E, Karpin A, Copeland J, Allsop D. Computer-delivered interventions for alcohol and tobacco use: a meta-analysis. Addiction 2010 Aug;105(8):1381-1390. [doi: 10.1111/j.1360-0443.2010.02975.x] [Medline: 20528806]
57. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on internet-based mental health interventions — a systematic review. Internet Interv 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
58. Willcutt EG, Doyle AE, Nigg JT, Faraone SV, Pennington BF. Validity of the executive function theory of attention-deficit/hyperactivity disorder: a meta-analytic review. Biol Psychiatry 2005 Jun 01;57(11):1336-1346. [doi: 10.1016/j.biopsych.2005.02.006] [Medline: 15950006]
59. 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]
60. Kessler RC, Adler L, Ames M, Demler O, Faraone S, Hiripi E, et al. The World Health Organization Adult ADHD Self-Report Scale (ASRS): a short screening scale for use in the general population. Psychol Med 2005 Feb;35(2):245-256. [doi: 10.1017/s0033291704002892] [Medline: 15841682]
61. Lauth GW, Minsel W. Kölner ADHS-Test für Erwachsene. URL: https://www.testzentrale.de/shop/koelner-adhs-test-fuer-erwachsene.html [accessed 2021-05-04]
62. Smyth AC, Meier ST. Evaluating the psychometric properties of the Conners adult ADHD rating scales. J Atten Disord 2019 Aug;23(10):1111-1118. [doi: 10.1177/1087054715624230] [Medline: 26838556]

Abbreviations

ASRS: Adult ADHD Self-Report Scale
CES-D: Center for Epidemiologic Studies Depression scale
CUDIT: Cannabis Use Disorder Identification Test
GAD: Generalized Anxiety Disorder scale
SDS: Severity of Dependence Scale

Edited by A Mavragani; submitted 03.05.21; peer-reviewed by F Moggi, M Sofis, N Wade, D Szinay; comments to author 13.06.21; revised version received 25.08.21; accepted 09.02.22; published 20.04.22.

Please cite as:
Ahlers J, Baumgartner C, Augsburger M, Wenger A, Malischnig D, Boumparis N, Berger T, Stark L, Ebert DD, Haug S, Schaub MP Cannabis Use in Adults Who Screen Positive for Attention Deficit/Hyperactivity Disorder: CANreduce 2.0 Randomized Controlled Trial Subgroup Analysis
J Med Internet Res 2022;24(4):e30138
URL: https://www.jmir.org/2022/4/e30138
doi:10.2196/30138
PMID:35442196

Original Paper

Impact of SMS Text Messaging Reminders on Helmet Use Among Motorcycle Drivers in Dar es Salaam, Tanzania: Randomized Controlled Trial

Benjamin Campbell[1], MA; Jesse Heitner[2], MPP, SD; Peter Amos Mwelelo[3], JD; Alexis Fogel[4], MD, MPH; Vaidehi Mujumdar[5,6], MD; Lisa V Adams[7], MD; Respicious Boniface[8], MD; Yanfang Su[9], MA, SD

[1]School of Education and Social Work, University of Sydney, Sydney, Australia

[2]International Clinical Research Center, Department of Global Health, University of Washington, Seattle, WA, United States

[3]Petmos Attorneys, Dar es Salaam, United Republic of Tanzania

[4]Benioff Children's Hospital Oakland, University of California San Francisco, San Francisco, CA, United States

[5]Department of Obstetrics and Gynecology, Ichan School of Medicine, Mount Sinai, New York City, NY, United States

[6]Department of Obstetrics and Gynecology, New York City Health and Hospitals Queens, New York City, NY, United States

[7]Section of Infectious Disease and International Health, Department of Medicine, Geisel School of Medicine at Dartmouth, Hanover, NH, United States

[8]Anaesthesiology Department, School of Medicine, Muhimbili University of Health and Allied Sciences, Dar es Salaam, United Republic of Tanzania

[9]Department of Global Health, University of Washington, Seattle, WA, United States

Corresponding Author:
Jesse Heitner, MPP, SD
International Clinical Research Center
Department of Global Health
University of Washington
908 Jefferson St
Seattle, WA, 98104
United States
Phone: 1 510 220 4266
Email: jah495@mail.harvard.edu

Abstract

Background: Road traffic injury is a pressing public health issue in Tanzania. Increasing helmet use among motorcycle drivers can help reduce the burden due to road traffic injuries in the country. Helmet adherence can be supported through mobile health interventions.

Objective: The aim of this study is to evaluate the comparative impact of two different types of SMS text messaging reminders on motorcycle helmet use.

Methods: Participants were 391 commercial motorcycle taxi drivers in Dar es Salaam, Tanzania. Participants were randomized into three groups, each receiving a different set of messages: (1) social norming messages aimed at emphasizing society's positive stance on helmet wearing, (2) fear appeal messages that emphasized the dangers of riding without a helmet, and (3) control group messages, which included basic road safety messages unrelated to helmet use. Every participant received the control messages. Adherence to helmet use was evaluated by self-report through surveys conducted at baseline, 3 weeks, and 6 weeks.

Results: At 6 weeks, the odds of self-reporting consistent helmet use were estimated to be 1.58 times higher in the social norming group than in the control group (P=.04), though this difference was not significant after accounting for multiple testing. There was little difference between fear appeal and control group recipients (odds ratio 1.03, P=.47). Subgroup analysis suggests that both fear appeal and social norming message types might have been associated with increased helmet use among participants who did not consistently wear helmets at baseline (odds ratio 1.66 and odds ratio 1.84, respectively), but this was not significant (P=.11 and P=.07, respectively). Among those who were consistent wearers at baseline, the social norming messages performed better than the fear appeal messages, and this difference reached traditional significance (P=.03), but was not significant after accounting for multiple testing.

Conclusions: The use of SMS text messaging reminders may improve helmet use among motorcycle drivers when framed as social norming messages. Given that nearly half of the drivers in our sample did not consistently wear their helmets on every trip, strategies to increase consistent usage could greatly benefit public safety.

Trial Registration: ClinicalTrials.gov NCT02120742; https://clinicaltrials.gov/ct2/show/NCT02120742

(J Med Internet Res 2022;24(4):e27387) doi:10.2196/27387

KEYWORDS

road traffic injury; behavior change; SMS reminders; mobile health; vehicle safety; mHealth; SMS; traffic injuries; transportation; public transportation; safety; automotive; automotive safety

Introduction

Road traffic injury is a pressing public health issue in Tanzania. According to the Global Burden of Disease 2017 Study, road traffic injury is the fourth leading cause of disability-adjusted life years for men aged 15-49 years in Tanzania [1]. Men are particularly at high risk of road traffic injury because nearly all drivers of motorcycle taxis ("bodabodas" in Kiswahili, or "bodas" for short), a major form of public transportation in the country, are men.

Studies have shown that helmet use can significantly reduce disability and death resulting from road traffic injuries [2]. Research conducted in Tanzania has shown that a lack of helmet wearing increases the probability of fatality in a motorcycle accident [3]. Efforts have been made by the Tanzanian government to develop tighter helmet use laws [4]. However, adherence to helmet use has remained dangerously low throughout the country [5,6]. This is partly because enforcement of laws is so limited [4].

One promising intervention to promote helmet wearing is the use of persuasive SMS text messaging reminders delivered to boda drivers. There is substantial evidence that mobile health interventions using SMS text messaging can lead to behavior change. For instance, in the largest study of its kind, texts reminding participants not to smoke significantly increased the likelihood that someone in a smoking cessation program would stop smoking [7]. Other studies have shown that text reminders can dramatically improve adherence to medication regimens [8,9]. Because of the high prevalence of cell phone and SMS text messaging use in Tanzania, especially among young people, the context is appropriate for such an intervention [10]. We implemented an innovative program that delivers SMS text messages to boda drivers over a 6-week period, reminding and persuading them to wear their helmets. To date, no program like this has been rigorously evaluated via randomized trial. The literature suggests that it takes approximately 21-42 days to form a new habit, so a 6-week study period was determined to be a sufficient time period to measure changes in helmet use [11,12]. The rate of consistent helmet wearing at the study's 6-week endpoint is the primary outcome of interest.

A key goal of this study is to measure which type of message leads to the greatest increase in helmet use. Messaging based on the social norming model, which promotes behavior change by informing the target population of how most people behave [13], could potentially be a more effective way of communicating road safety messages, particularly for men, compared to appealing to a fear of negative health outcomes, which is a longstanding health communication method. Findings from two recent studies support this idea by showing that road safety advertisements with threats of social consequences, such as the threat of losing one's driving license, were more effective at changing young males' driving behaviors than were advertisements depicting harsh physical consequences [14,15]. The fear appeal method, while historically prominent in the field of public health, has more recently been shown to be ineffective in leading to behavior change, especially among young men [16]. For example, a study by Woolley et al [16] demonstrated that men often dissociate their own speeding behaviors from a social problem and therefore perceive related fear appeals as being directed more toward others than themselves. This is consistent with a broader trend in the psychology of aging literature wherein younger adults are more motivated by potential rewards than loss aversion, a balance that reverses in older adulthood [17]. In this study, we aimed to test social norming and fear appeal messages against a control and against each other to see which, if either, has a greater impact on helmet use. Our trial implementation was successful and our results tentatively favor one type of message over the other.

Methods

Study Design

We conducted a randomized controlled trial to evaluate the impact of an SMS text messaging program on helmet use among boda drivers in Dar es Salaam, Tanzania. Participants were recruited in a convenience sample from the population of boda drivers in 3 districts in Dar es Salaam. Boda drivers were approached at boda stands with 3 or more boda drivers waiting for clients. The inclusion criteria required participants to be ≥18 years old, own a mobile phone with SMS text messaging capabilities, demonstrate the ability to retrieve SMS text messages, and have access to a helmet. Preintervention power calculations indicated that 385 participants would be needed to detect a 20 percentage point increase in consistent helmet use over an anticipated baseline of 32.4% being consistent users. In total, 391 participants were recruited. There were no incentives to join the study.

All participants were informed that they would receive 3 SMS text messages per week. A prestudy questionnaire indicated that the best time to deliver SMS text messages was between 6 AM and 7 AM, during off-peak hours. Participants were randomized into one of three study arms, with each group receiving a

different type of message: (1) social norming (eg, "Most of your peers properly wear their helmet every day – do you?"); (2) fear appeal (eg, "If you do not wear your helmet while driving, you will increase your chances of injury"); and (3) control, which included basic road safety messages (eg, "This is a short reminder to not speed while driving your boda"). Groups 1 and 2 received the control messages in addition to their group-specific messages. The information in the group-specific messages was designed to be both motivational and accurate, and was based on a literature review of motorcycle helmet use and road safety in Tanzania and the surrounding region. Participants received the intervention between May and June 2014. Texts were delivered in the local language, Kiswahili, using a mass-messaging platform called MightyText. Texts were sent Monday, Wednesday, and Friday mornings. For the complete message list and the literature source [2,18-23] for each message, see Multimedia Appendix 1.

Randomization proceeded in a 4-step process that created matched triplets of drivers and randomly assigned one member from each triplet to each study arm. First, a logistic regression of baseline consistent helmet use on demographic and driving-habit covariates was used to create a propensity score for predicting baseline helmet use. Second, participants were stratified into two groups: those who at baseline reported they had consistently worn their helmet on all trips in the past two weeks, and those who reported inconsistent use. Third, within the two strata, triplets were made, beginning by grouping the 3 individuals with the lowest use, then the next 3 lowest, and so on, in the same "propensity triplet." Only one individual with the highest score remained unmatched into a triplet. Finally, for each triplet, an integer from 1 to 6 was randomly drawn with replacement. Each integer represented one of the 6 permutations by which 3 (ordered) individuals may be assigned (one each) to 3 different treatments: ABC, ACB, BAC, BCA, CAB, and CBA. The individuals in the triplet were thereby simultaneously assigned to an arm of the study, with one member in each treatment arm. The last individual with the highest score was similarly assigned, and treated as the lowest score in their own triplet.

Matching in this fashion had two aims. First, it created equally sized treatment arms, which maximized statistical power across the planned group comparisons. Second, it was intended to balance the drivers' unobservable propensity to wear helmets across treatment arms by ensuring that baseline helmet use and predicted helmet use were balanced across treatment arms. Stratification assured that equal numbers of consistent wearers and inconsistent wearers were in each study arm, and matching into triplets based on close propensity scores prior to random assignment assured that estimated propensity to consistently wear helmets was also balanced across treatment arms. Matching on a propensity score constructed from observable covariates has been shown to be sufficient to remove bias from all covariates used to construct the propensity score [24]. Though this technique was originally conceived to improve causal inference in observational studies, matching on relevant covariates before treatment assignment in randomized experiments is now a common practice that can increase efficiency of estimation and the power of hypothesis tests [25]. Moreover, inadvertently matching on irrelevant covariates prior to a random assignment does not harm statistical efficiency or power [25].

Participant adherence to helmet use was captured through self-report surveys at baseline, at the 3-week midpoint of the experiment, and at 6 weeks (conclusion of the experiment).

Ethics Approval

This study was approved by the Institutional Review Board (Committee for the Protection of Human Subjects) at Dartmouth College, United States (00024570), and the Ethics Review Committee at Muhimbili University of Health and Allied Sciences, Tanzania.

Study Population Baseline Characteristics

The baseline characteristics of the study population are shown in Table 1. The mean age of participants was 28 years, all participants were men, and a majority had at most an elementary level education. At baseline, approximately 53% (207/391) of participants claimed that they wore their helmet on every trip, which was more than the 32% anticipated from previous literature review, on which our power calculations were based. There were no statistically significant differences across treatments for any observed variable. Self-reporting of consistent helmet wearing was perfectly balanced across treatment arms by the stratified design of the randomization method.

Table 1. Balance check for all observed baseline variables.

Baseline variable	Social norming	Fear appeal	Control	Test statistic[a]	*P* value
District (n=391), n (%)				$X^2(4)=1.71$	.79
District 1	51 (39.2)	48 (36.6)	42 (32.3)		
District 2	55 (42.3)	61 (46.6)	62 (47.7)		
District 3	24 (18.5)	22 (16.8)	26 (20)		
Age, years (n=384), mean (SD)	28.6 (6.4)	27.6 (6.7)	27.9 (5.9)	$F=0.86$	.43
Education (n=375), n (%)				$X^2(2)=0.57$	.75
Elementary or none	86 (67.7)	85 (69.1)	90 (72)		
Junior high or above	41 (32.3)	38 (30.9)	35 (28)		
Currently married (n=381), n (%)	83 (65.4)	85 (67.5)	84 (65.6)	$X^2(2)=0.15$	.93
Has children (n=378), n (%)	79 (62.7)	82 (66.1)	81 (63.3)	$X^2(2)=0.37$	.83
Cell phone self-owned (n=380), n (%)	128 (100)	126 (100)	124 (98.4)	$X^2(2)=4.05$	.13
Household size (n=386), mean (SD)	5.19 (2.4)	5.28 (3.2)	4.93 (2.2)	$F=0.59$	.56
Primary driving setting (n=380), n (%)				$X^2(4)=2.81$	.59
Urban/downtown	25 (19.7)	29 (22.3)	30 (24.4)		
Suburban/residential	24 (18.9)	32 (24.6)	23 (18.7)		
Both equally	78 (61.4)	69 (53.1)	70 (56.9)		
Night driving frequency (n=389), n (%)				$X^2(6)=3.00$	.81
Never	34 (26.4)	42 (32.3)	36 (27.7)		
Sometimes	41 (31.8)	43 (33.1)	45 (34.6)		
Usually	32 (24.8)	26 (20)	33 (25.4)		
Always	22 (17.1)	19 (14.6)	16 (12.3)		
Wears helmet consistently (n=391), n (%)	69 (53.1)	69 (52.7)	69 (53.1)	$X^2(2)=0.01$	>.99
Speeding frequency (n=390), n (%)				$X^2(6)=3.55$	.74
Never	15 (11.5)	18 (13.9)	20 (15.4)		
Sometimes	41 (31.5)	42 (32.3)	36 (27.7)		
Usually	65 (50)	64 (49.2)	70 (58.9)		
Always	9 (6.9)	6 (4.6)	4 (3.1)		
Weekend driving (n=391), n (%)				$X^2(6)=8.71$	.19
Never	9 (6.9)	3 (2.3)	7 (5.4)		
Sometimes	21 (16.2)	36 (27.5)	33 (25.4)		
Usually	30 (23.1)	33 (25.2)	32 (24.6)		
Always	70 (53.9)	59 (45.0)	58 (44.6)		

[a]Chi-square tests were conducted for all variables except for the age and household size variables, for which an analysis of variance was conducted.

Statistical Methods

The primary outcome of the study was self-reported adherence to helmet use as measured by the question, “In the past week, how often did you wear your helmet: (1) Every trip; (2) Not every trip.” We compared the adherence rate between experimental groups and between each experimental group and the control. A secondary outcome was heterogeneity of treatment effect by baseline helmet use habits.

A reliance on self-reports potentially introduces measurement error due to possible social desirability bias. Because helmet use is legally required, participants may have reported wearing them more frequently so as to be viewed positively. We aimed to overcome the social desirability bias by ensuring that the survey responses were anonymous. One indication that this strategy may have been successful is shown in survey respondents’ self-reported frequency of speeding; interestingly, 56% (218/390) of respondents were willing to admit to

exceeding speed limits “frequently” or “always.” Another 30% (129/390) reported speeding at least “sometimes.” Speeding would be expected to be subject to the same social desirability bias as helmet use, but many respondents were willing to self-report this behavior in the anonymous survey.

To investigate the effect of treatment arm assignment, several logistic regressions of consistent helmet use on treatment assignment were run. All statistical analyses were performed using R (version 3.0.2; R Foundation for Statistical Computing). All specifications were structured to estimate an intent-to-treat effect. For all group comparisons in all specifications, statistical significance of group difference was performed by permutation analysis as follows. First, the specification was run on all data using the true treatment assignment. The analysis was then rerun with each triplet of individuals (falsely) rerandomized with replacement to one of the 6 possible permutations of treatment assignments for that triplet. Performing this analysis with many permutations wherein analyzed treatment assignments have no relation to the intervention or associated outcomes recreates the distribution of the null hypothesis in which treatment and outcomes are unrelated. The analysis was run with 5000-10,000 permutations (depending on the specification), and significance was assessed by the percentage of runs in which the null distribution yielded results of larger magnitude than that of the true treatment assignment. Permutation tests have been shown to be valid for conducting any test of a null hypothesis of no treatment effect within an experimental sample, conditional on the single requirement that treatment has been randomly assigned [26].

Results

Our intervention was delivered over a 6-week period, with helmet use measurement at week 3 and week 6. The primary outcome of interest is the proportion of self-reported consistent helmet use at week 6. Unadjusted levels of reported helmet use for each group at both time points are shown in Table 2.

The final row represents the difference between a group’s week 6 difference from control and the group’s baseline difference from control.

Potential heterogeneity of effect by randomization strata was investigated by analyzing participants in two subgroups based on whether they were or were not consistent helmet wearers at baseline. Intervention effects had strong potential to be different in magnitude between these strata because the mechanism of effect was necessarily different between these groups. Among already consistent wearers, the only possible mechanism of effect is maintenance of adherence; conversely, for the inconsistent wearers, the only possible mechanism is promotion of adherence among the not yet adherent. Knowledge about heterogeneity or consistency of effect is important for future targeting of interventions. Unadjusted results are shown in Table 3.

The results in Table 2 show that the fear appeal and control groups had little change over the 6-week period. However, the group receiving social norming SMS text messages showed a final 11.1% lead over the control group in consistent helmet wearing despite their initial equal levels.

The results in Table 3 potentially indicate even more striking differences between treatment arms. Among drivers always wearing their helmets at baseline, the social norming arm had 9.6% more drivers stay consistent than the control arm, and the fear appeal group actually had 8.3% fewer drivers stay consistent, potentially denoting a detrimental effect of fear messages in this subgroup. Among drivers that began as inconsistent helmet wearers, 36% (18/50) of the control group became consistent helmet wearers, but the gains in the fear appeal and social norming arms were even larger (28/58, 48.3% and 29/57, 50.9%, respectively).

Table 2. Drivers reporting helmet use every trip (all time points).

Observations	Control	Fear appeal	Social norming
Baseline, n/N (%)	69/130 (53.1)	69/131 (52.7)	69/130 (53.1)
Difference from control, %	Reference	–0.4	0.0
Week 3, n/N (%)	62/113 (54.9)	62/117 (53)	70/122 (57.4)
Difference from control, %	Reference	–1.9	2.5
Week 6, n/N (%)	58/110 (52.7)	63/118 (53.4)	74/116 (63.8)
Difference from control	Reference	0.7	11.1
Week 6 difference in differences	Reference	1.1	11.1

Table 3. Drivers reporting helmet use for every trip at 6 weeks, by baseline answer.

Helmet use	Control	Fear appeal	Social norming
Subgroup: baseline "consistent wearers"			
Week 6, n/N (%)	40/60 (66.7)	35/60 (58.3)	45/59 (76.3)
Difference from control	Reference	–8.3	9.6
Subgroup: baseline "inconsistent wearers"			
Week 6, n/N (%)	18/50 (36)	28/58 (48.3)	29/57 (50.9)
Difference from control	Reference	12.3	14.9

Hypothesis testing of group-level differences was performed with logistic regression and *P* values were generated via nonparametric permutation testing to account for the correlations induced by the multistep randomization process. Regressions unadjusted for any covariates are shown in Table 4, which tests the odds ratios associated with the risk differences presented in Table 2 and Table 3. The first two columns display test results of whether each treatment arm statistically differed from the control arm. The final column of Table 4 shows tests of whether and how the effects of the two treatment arms statistically differ from each other. The first row of Table 4 presents these tests using all observations. The second and third rows present these same tests within the two subgroups of baseline "always wearers" and baseline "inconsistent wearers." Whether there was a heterogeneous effect of treatment assignment by this baseline subgrouping is displayed in the last row of Table 4, which tests for effect modification by taking the ratio of the odds ratios between the subgroups and testing whether this ratio is significant via permutation.

In comparing the two intervention arms to the control arm, 1-sided tests of significance were used, justified by the strong a priori expectation that the two message types would only encourage, not discourage, helmet wearing. However, because we had no such a priori expectation that one messaging intervention would work better than the other, a 2-sided test was used whenever comparing the social norming and fear appeal groups.

Using all observations in an unadjusted analysis, participants in the social norming arm had odds of consistently wearing their helmet that were 1.58 times the odds of the control group, which was the strongest measured association. Jointly testing all three possible group comparisons among all participants is this study's primary, trial-registered outcome, and it was preplanned to use a Holm-Bonferroni correction to account for this multiple testing. The 1-sided *P* value of .04 comparing the social norming arm to the control arm was not enough to satisfy the Holm-Bonferroni cutoff for simultaneously testing 3 null hypotheses, which requires that the most significant of 3 *P* values be less than or equal to .05/3 (.0167) to set a maximum Type I family-wise error rate of .05.

Within the subgroup of participants that started as consistent helmet wearers, neither intervention arm differed significantly from the control arm. The social norming group was measured to have 2.30 times the odds of the fear appeal group of consistently wearing their helmets (P=.03). However, this is nonsignificant under the Holm-Bonferroni correction for simultaneously testing 3 group differences in this subset. In the subgroup of participants that were not consistent users at baseline, both intervention arms outperformed the control group, but their gains, while perhaps clinically meaningful in size, were not statistically significant at a threshold of P<.05. Finally, the lowest section of Table 4 investigates whether the same message arms had different effects between the two subgroups (baseline "always wearers" and baseline "inconsistent wearers"). Although the measured effects had seemingly large differences across subgroups, these differences had *P* values well above .05.

After the unadjusted analysis, the same set of logistic regressions was performed including a set of demographic factors and baseline driving habits as controls. These controls were marital status, driving setting (primarily downtown or primarily suburban portions of the city), frequency of driving at night, and frequency of driving on the weekend. This list of controls was somewhat smaller than originally intended for several reasons. Originally, age and whether the driver had children were intended to be included in the controls, but strong multicollinearity between age, marital status, and having children precluded using all three simultaneously. Marital status was deemed to be the best summary indicator of the three as its effect was most consistent and interpretable across specifications. In addition, large amounts of missingness in self-reported income precluded its inclusion as a control variable. Table 5 reports the results of the adjusted logistic regressions.

The results in Table 5 follow those in Table 4 with relatively minor deviations. Given that the included variables were part of the original propensity score matching, it is unsurprising that their inclusion fails to alter the analysis in any meaningful way.

Table 4. Pairwise treatment group comparisons of odds of consistent helmet wearing (using coefficient results of unadjusted logistic regression).

Subgroup analysis	Fear appeal: control group comparison	Social norming: control group comparison	Social norming: fear appeal group comparison
All observations			
Odds ratio	1.03	1.58	1.54
P value[a]	.47[b]	.04[b]	.12[c]
Subgroup: baseline "always wearers"			
Odds ratio	0.70	1.61	2.30
P value[a]	.81[b]	.11[b]	.03[c]
Subgroup: baseline "inconsistent wearers"			
Odds ratio	1.66	1.84	1.11
P value[a]	.11[b]	.07[b]	.80[c]
Subgroup effect modification			
Ratio of odds ratios	0.42	0.87	2.07
P value[a]	.16[c]	.82[c]	.21[c]

[a]All *P* values determined by permutation analysis.

[b]One-sided test.

[c]Two-sided test.

Table 5. Pairwise treatment group comparisons of odds of consistent helmet wearing (using coefficient results of covariate-adjusted logistic regression).

Subgroup analysis	Fear appeal: control	Social norming: control	Social norming: fear appeal
All observations			
Odds ratio	1.01	1.57	1.55
P value[a]	.49[b]	.06[b]	.12[c]
Subgroup: baseline "always wearers"			
Odds ratio	0.62	1.58	2.54
P value[a]	.86[b]	.15[b]	.03[c]
Subgroup: baseline "inconsistent wearers"			
Odds ratio	1.84	1.90	1.03
P value[a]	.09[b]	.08[b]	.93[c]
Subgroup effect modification			
Ratio of odds ratios	0.34	0.83	2.46
P value[a]	.07	.76[c]	.15[c]

[a]All *P* values determined by permutation analysis.

[b]One-sided test.

[c]Two-sided test.

Discussion

Principal Findings

The results of our study show that social norming messages are potentially effective at increasing helmet use among motorcycle taxi "boda" drivers in Dar es Salaam, Tanzania. Over the 6-week period, the group receiving social norming SMS text messages showed an increase in helmet use from 53.1% to 63.8%, and that increase achieved traditional significance ($P<.05$) when compared to the control group ($P=.04$). However, accounting for multiple testing, we cannot reject the null of no association, as this P value is above the required $P\leq.0167$ to maintain a family-wise Type I error rate of at most .05 when making 3 group comparisons. In contrast, the fear appeal and control groups showed little change over the 6-week period.

Although the main finding shows that the group receiving social norming messages increased helmet adherence the most, though not to a statistically significant degree, the findings also suggest that responsiveness to messages may also have been determined

by participant baseline response. Specifically, for those who reported not wearing helmets all the time at baseline, both social norming and fear appeal messages were associated with higher adherence after the 6-week study period compared to the control group. Though shy of statistical significance due to the power limitation of restricting the sample, the associated odds ratios imply a near doubling of the odds of consistent usage, and the close similarity of the odds ratios between the two treatment arms suggests that initially inconsistent wearers are equally sensitive to both types of messages. However, among those who reported consistent helmet wearing at baseline, those recipients of social norming messages maintained high levels of adherence, while those receiving fear appeal messages actually decreased their level of consistent wearing compared to the control. Although neither treatment is associated with a statistical difference from the control in this subgroup, the combination of a positive association in the social norming arm and a deleterious association in the fear appeal arm results in a traditionally significant improvement of the social norming arm over the fear appeal arm (odds ratio 2.30, P=.03). However, this association does not meet the Holm-Bonferroni requirement of $P \leq .0167$.

These findings have important potential implications for policy makers as well as other stakeholders in road safety. Because social norming messaging overall showed a potentially greater association with consistent helmet use than fear appeal messaging, it could be strategic for regulators and nongovernmental organizations focusing on road traffic safety to use social norming messages for any mass message or media campaigns to promote road safety and behavior change among drivers. However, a larger and more highly powered study would be required to confirm this differential association. Finally, intervention designers should note that behavior change may take some time to set in among drivers; group-level differences were noticeable at 6 weeks, but not after 3 weeks.

Limitations

There are several limitations to this study. First, self-reports introduce the possibility of social desirability bias among the respondents thanks to the legal requirement that helmets be worn at all times. A second potential bias in this study is simply recall bias. Our main outcome question asks for an estimate of helmet use in the past week of boda driving. It is possible that drivers had difficulty remembering with accuracy the level of helmet wearing during that time. However, we believed that asking about behavior over the past week was a reasonable amount of time to ensure accuracy of estimates. Moreover, the recall burden is much lower when recalling consistency than when recalling the number of times something occurred or other numeric answers. Third, while the results can be useful in a Tanzanian urban context, they may not be applicable to other contexts. Fourth, the study was conducted in a convenience sample. The representativeness of the sample for Dar es Salaam boda drivers is left unknown. Fifth, our study measures effects of the intervention right after completion of the 6-week trial. How long measured effects persist into the future is unknown. Finally, our study is focused on helmet usage, while the ultimate goal of such an intervention is better health and safety for drivers on the road. This study was not structured or powered to detect differences in health outcomes by treatment arm, and further research would be necessary to determine if such a messaging intervention would improve health outcomes for drivers.

Conclusions

Though the evidence is not fully conclusive, this study suggests that SMS text messaging reminders can be an effective way to improve helmet use among motorcycle drivers. Specifically, social norming messages appear to be more effective than fear appeal messages when trying to increase helmet use among boda drivers. Furthermore, for drivers who already wear their helmet consistently, fear appeal messages may actually have a detrimental effect on helmet use. Future research should further investigate whether social norming messages are more effective than fear appeals when trying to change behavior.

Acknowledgments

The authors would like to acknowledge the leadership at road safety nongovernmental organization Amend, including Jeffrey Witte, Tom Bishop, and Kate Raum, who provided invaluable guidance and support in implementing the study; Neema Swai, for her assistance with survey translation; Deepani Jinadasa, for assistance on study design and logistics; and Simon Kalolo, for technical assistance.

We also would like to acknowledge the funders of this study, including Amend, Dartmouth College, University of Washington, and East West Alliance for Education and Health.

Authors' Contributions

BC acquired funding for the research; participated in the conceptualization, design, analysis, and interpretation of the study; led the implementation of the study, and participated in the writing of the manuscript. JH participated in the conceptualization and design of the study, led the data analysis, and led drafting of the manuscript. PA participated in study design and implementation and in reviewing the manuscript. AF participated in the conceptualization, design, analysis, and interpretation of the study and reviewed the manuscript. VM participated in the conceptualization, design, analysis, and interpretation of the study and conducted literature review. AS helped to conduct literature review and provided feedback on the manuscript. LVA and RB participated in the conceptualization, design, analysis, and interpretation of the study and reviewed and provided feedback on the manuscript.

YS acquired funding for the research; participated in the conceptualization, design, analysis, and interpretation of the study; and reviewed and provided feedback on the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full-text message bank.
[DOCX File , 26 KB - jmir_v24i4e27387_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1111 KB - jmir_v24i4e27387_app2.pdf]

References

1. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2016 (GBD 2016) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME); 2017. URL: http://ghdx.healthdata.org/gbd-results-tool [accessed 2018-11-25]
2. Liu B, Ivers R, Norton R, Blows S, Lo S. Helmets for preventing injury in motorcycle riders. Cochrane Database Syst Rev 2004;4:CD004333. [doi: 10.1002/14651858.cd004333.pub2]
3. Salum JH, Kitali AE, Bwire H, Sando T, Alluri P. Severity of motorcycle crashes in Dar es Salaam, Tanzania. Traffic Inj Prev 2019 Mar 19;20(2):189-195. [doi: 10.1080/15389588.2018.1544706] [Medline: 30888877]
4. United Republic of Tanzania. National Road Safety Policy. Dar es Salaam: Ministry of Infrastructure Development; 2009 Sep. URL: http://www.mow.go.tz/uploads/publications/24026c0843c2198084a6fff70d176bc1.pdf [accessed 2016-07-23]
5. Chalya PL, Mabula JB, Dass RM, Mbelenge N, Ngayomela IH, Chandika AB, et al. Injury characteristics and outcome of road traffic crash victims at Bugando Medical Centre in Northwestern Tanzania. J Trauma Manag Outcomes 2012 Feb 09;6(1):1-8 [FREE Full text] [doi: 10.1186/1752-2897-6-1] [Medline: 22321248]
6. Mwakapasa EG. Attitude towards and practice of helmet use among commercial motorcyclists in Dar es Salaam region, Tanzania. Academia. 2011. URL: http://www.academia.edu/5249218/ATTITUDE_TOWARDS_AND_PRACTICE_OF_HELMET_USE_AMONG_COMMERCIAL_MOTORCYCLISTS_IN_DAR_ES_SALAAM_REGION_TANZANIA [accessed 2013-12-15]
7. Free C, Knight R, Robertson S, Whittaker R, Edwards P, Zhou W, et al. Smoking cessation support delivered via mobile phone text messaging (txt2stop): a single-blind, randomised trial. Lancet 2011 Jul 02;378(9785):49-55 [FREE Full text] [doi: 10.1016/S0140-6736(11)60701-0] [Medline: 21722952]
8. Park L, Howie-Esquivel J, Chung M, Dracup K. A text messaging intervention to promote medication adherence for patients with coronary heart disease: a randomized controlled trial. Patient Educ Couns 2014 Feb;94(2):261-268 [FREE Full text] [doi: 10.1016/j.pec.2013.10.027] [Medline: 24321403]
9. Vervloet M, van Dijk L, Santen-Reestman J, van Vlijmen B, van Wingerden P, Bouvy M, et al. SMS reminders improve adherence to oral medication in type 2 diabetes patients who are real time electronically monitored. Int J Med Inform 2012 Sep;81(9):594-604. [doi: 10.1016/j.ijmedinf.2012.05.005] [Medline: 22652012]
10. Lancaster H. Tanzania Telecoms Market Report - Telecoms, Mobile and Broadband - Statistics and Analyses. Budde Comm. 2015 Jul. URL: http://www.budde.com.au/Research/Tanzania-Telecoms-Mobile-and-Broadband-Statistics-and-Analyses.html?r=51 [accessed 2022-03-03]
11. Maltz M. Psycho-cybernetics. New York: Pocket Books; 1969.
12. Kaushal N, Rhodes RE. Exercise habit formation in new gym members: a longitudinal study. J Behav Med 2015 Aug;38(4):652-663. [doi: 10.1007/s10865-015-9640-7] [Medline: 25851609]
13. Cialdini R, Kallgren C, Reno R. A Focus Theory of Normative Conduct: A Theoretical Refinement Reevaluation of the Role of Norms in Human Behavior. Advances in Experimental Social Psychology 1991:201-234 [FREE Full text] [doi: 10.1016/s0065-2601(08)60330-5]
14. Harré N, Foster S, O'Neill M. Self-enhancement, crash-risk optimism and the impact of safety advertisements on young drivers. Br J Psychol 2005;96(2):230. [doi: 10.1348/000712605x36019]
15. Lennon R, Rentfro R, O'Leary B. Social marketing and distracted driving behaviors among young adults: The effectiveness of fear appeals. Acad Mark Stud J 2010;14(2):95.
16. Wundersitz L, Hutchinson T, Woolley J. Best practice in road safety mass media campaigns: A literature review. Adelaide: University of Adelaide; 2010.
17. Ebner NC, Freund AM, Baltes PB. Developmental changes in personal goal orientation from young to late adulthood: from striving for gains to maintenance and prevention of losses. Psychol Aging 2006 Dec;21(4):664-678. [doi: 10.1037/0882-7974.21.4.664] [Medline: 17201488]

18. Nyoni JE, Masaoe EN. Factors contributing to high frequency of vulnerable road user injury in Dar Es Salaam. 2011 Jul Presented at: 30th Annual Southern African Transport Conference 'Africa on the Move'; July 11-14, 2011; Pretoria, South Africa URL: http://repository.up.ac.za/handle/2263/17298
19. Mcharo B. Motorcycle crash: injuries pattern and associated factors among patients treated at Muhimbili Orthopaedic Institute (MOI) [thesis]. Muhimbili University of Health and Allied Sciences. Dar es Salaam; 2012. URL: https://docplayer.net/517648-Motorcycle-crash-injuries-pattern-and-associated-factors-among-patients-treated-at-muhimbili-orthopaedic-institute-moi.html [accessed 2014-04-17]
20. Chalya PL, Mabula JB, Ngayomela IH, Kanumba ES, Chandika AB, Giiti G, et al. Motorcycle injuries as an emerging public health problem in Mwanza City, north-western Tanzania. Tanzan J Health Res 2010 Oct;12(4):214-221. [Medline: 24409627]
21. Saidi H, Mutisto BK. Motorcycle injuries at a tertiary referral hospital in Kenya: injury patterns and outcome. Eur J Trauma Emerg Surg 2013 Oct 13;39(5):481-485. [doi: 10.1007/s00068-013-0280-8] [Medline: 26815444]
22. Galukande JJ. Boda-boda Injuries a Health Problem and a Burden of Disease in Uganda: a Tertiary Hospital Survey. East Cent Afr J Surg 2016 Jun 6:33-37.
23. Chalya PL, Dass RM, Mchembe MD, Mbelenge N, Ngayomela IH, Chandika AB, et al. Citywide trauma experience in Mwanza, Tanzania: a need for urgent intervention. J Trauma Manag Outcomes 2013 Nov 11;7:1-9 [FREE Full text] [doi: 10.1186/1752-2897-7-9] [Medline: 24499566]
24. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70(1):41-55. [doi: 10.1093/biomet/70.1.41]
25. Greevy R, Lu B, Silber JH, Rosenbaum P. Optimal multivariate matching before randomization. Biostatistics 2004 Apr 01;5(2):263-275. [doi: 10.1093/biostatistics/5.2.263] [Medline: 15054030]
26. Anderson MJ. Permutation tests for univariate or multivariate analysis of variance and regression. Can J Fish Aquat Sci 2001 Mar 01;58(3):626-639. [doi: 10.1139/f01-004]

Edited by R Kukafka; submitted 23.01.21; peer-reviewed by M Peden, G Porter; comments to author 07.04.21; revised version received 20.04.21; accepted 10.11.21; published 07.04.22.

Please cite as:
Campbell B, Heitner J, Amos Mwelelo P, Fogel A, Mujumdar V, Adams LV, Boniface R, Su Y
Impact of SMS Text Messaging Reminders on Helmet Use Among Motorcycle Drivers in Dar es Salaam, Tanzania: Randomized Controlled Trial
J Med Internet Res 2022;24(4):e27387
URL: https://www.jmir.org/2022/4/e27387
doi:10.2196/27387
PMID:35389364

Original Paper

Weight Loss Trajectories and Related Factors in a 16-Week Mobile Obesity Intervention Program: Retrospective Observational Study

Ho Heon Kim[1], MS; Youngin Kim[1,2], MD; Andreas Michaelides[2], PhD; Yu Rang Park[1], PhD

[1]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea

[2]Noom Inc, New York, NY, United States

Corresponding Author:
Yu Rang Park, PhD
Department of Biomedical Systems Informatics
Yonsei University College of Medicine
50-1 Yonsei-ro, Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 010 5240 3434
Fax: 82 02 2227 8308
Email: yurangpark@yuhs.ac

Abstract

Background: In obesity management, whether patients lose ≥5% of their initial weight is a critical factor in clinical outcomes. However, evaluations that take only this approach are unable to identify and distinguish between individuals whose weight changes vary and those who steadily lose weight. Evaluation of weight loss considering the volatility of weight changes through a mobile-based intervention for obesity can facilitate understanding of an individual's behavior and weight changes from a longitudinal perspective.

Objective: The aim of this study is to use a machine learning approach to examine weight loss trajectories and explore factors related to behavioral and app use characteristics that induce weight loss.

Methods: We used the lifelog data of 13,140 individuals enrolled in a 16-week obesity management program on the health care app Noom in the United States from August 8, 2013, to August 8, 2019. We performed k-means clustering with dynamic time warping to cluster the weight loss time series and inspected the quality of clusters with the total sum of distance within the clusters. To identify use factors determining clustering assignment, we longitudinally compared weekly use statistics with effect size on a weekly basis.

Results: The initial average BMI value for the participants was 33.6 (SD 5.9) kg/m^2, and it ultimately reached 31.6 (SD 5.7) kg/m^2. Using the weight log data, we identified five clusters: cluster 1 (sharp decrease) showed the highest proportion of participants who reduced their weight by >5% (7296/11,295, 64.59%), followed by cluster 2 (moderate decrease). In each comparison between clusters 1 and 3 (yo-yo) and clusters 2 and 3, although the effect size of the difference in average meal record adherence and average weight record adherence was not significant in the first week, it peaked within the initial 8 weeks (Cohen d>0.35) and decreased after that.

Conclusions: Using a machine learning approach and clustering shape-based time series similarities, we identified 5 weight loss trajectories in a mobile weight management app. Overall adherence and early adherence related to self-monitoring emerged as potential predictors of these trajectories.

(J Med Internet Res 2022;24(4):e29380) doi:10.2196/29380

KEYWORDS

clustering; mobile health; weight loss; weight management; behavior management; time series analysis; mHealth; obesity; outcomes; machine learning; mobile app; adherence; prediction; mobile phone

Introduction

Background

The worldwide prevalence of overweight or obesity has doubled since 1980 [1]. In the United States, the prevalence of obesity has increased dramatically among both adults and children [2]. Meanwhile, epidemiologic research has identified high BMI values as a risk factor for an expanding set of chronic diseases, including cardiovascular disease, diabetes mellitus, kidney disease, and several cancers [1].

There is a need for effective obesity interventions that can reach large population sectors at low cost. With 66% of the world's population owning tablets or smartphones [3], web-based interventions can facilitate the implementation of wide-reaching, self-directed approaches to tackle obesity. Self-directed interventions require minimal contact with professionals and empower participants to control and regulate their thoughts themselves [4]. Relevant guidelines and systematic reviews have widely recognized self-directed interventions as effective intervention techniques for obesity treatment [5]. Several studies have examined the effectiveness of self-directed interventions [5-10] and have suggested that such approaches could be used to deliver improved and personalized care while reducing health care costs [5].

Weight Loss and Its Predictors

Individuals who lose 5%-10% of their initial weight are generally considered as responding to their obesity treatment: these proportions are clinically associated with improvement in cardiovascular risk factors [11]. In several weight loss intervention trials, participants who lost 5%-10% of their initial weight by the end were considered responders [12-14]. However, this approach does not distinguish between individuals with widely varying weight changes and those who steadily lose weight throughout the intervention period. Furthermore, little is known about patterns of individual week-to-week weight changes and about how these changes may relate to weight loss achievement (or a lack thereof) [14]. Such evidence could shed light on the relationship between weight loss and its predictors from a longitudinal perspective and would be of potential use in tailoring weight management and behavior change interventions, especially for groups at risk of suboptimal outcomes [15].

Research has yet to outline weight loss trajectories in individuals who have participated in self-directed interventions using mobile phones. Therefore, this study aims to (1) identify weight loss trajectories and (2) explore factors related to behavioral and health care app use characteristics eliciting weight loss.

Methods

Study Design and Participants

In this retrospective study, we obtained deidentified user log data from Noom, Inc, which provides mobile app services. Noom is a mobile-based wellness app that focuses on behavior changes to enable weight loss. Through this app, users can receive personalized and one-on-one coaching from health experts and self-monitor their food intake and exercise. Coaching encourages users to record their weight on a weekly basis and meals on a daily basis. In addition, this app enables users to record their energy intake with higher reliability than professional software [16]. Users select the volume of their intake or units from which calorie intake can be measured. By spurring users to record their meals, exercise sessions, and weight, this app keeps them aware of their weight status and dietary patterns. This program was modeled after the Diabetes Prevention Program, sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases, where participants received 16 weeks' worth of content over the first 6 months. There is additional curriculum past the 16-week mark to support users for approximately 1 year.

The retrospective cohort for this study included 93,814 users in the United States who were enrolled in the coaching program provided by Noom from August 8, 2013, to August 8, 2019. To target users with the same intervention duration, we only included those users who participated in the program for 16 weeks. Among these, we excluded app users without weight records because we could not evaluate their outcomes. We also excluded users with a height <125 cm or >230 cm, which is a loosened criterion of US Census protocol and these cutoff points correspond to 0.01% or 99.99%, respectively, of height distribution of our cohort [17]. Next, we excluded users with age data not meeting the inclusion criteria, such as young users or those who did not enter their age when signing up. To focus on individuals who needed to manage their weight, we included adult users with a BMI higher than the criterion for overweight (≥25 kg/m^2), which could potentially pose health problems such as cardiovascular disease based on the recommendations of the Centers for Disease Control and Prevention [18]. In addition, because this app (Noom) can include users who are underweight and want to gain weight from dietary management, this allowed us to focus on individuals who would be classified as overweight. Furthermore, we selected users with weight records available during the last week of the program or between the end of the program and 1 week after the end of the program to identify the final outcome of the weight loss program. Finally, we excluded users with inconsistent weight records that showed a difference in BMI of ≥3.5 kg/m^2 between consecutive time points within 1 month or of ≥7.0 kg/m^2 between consecutive time points within 2 months [19].

Data Acquisition and Preprocessing

We obtained deidentified log data for the users' dietary log, steps, weight, texting records (server logs of messages sent to, and received from, the coach absent any content), and demographic characteristics from Noom. Initially, we reviewed data for 93,814 users from five tables—meal logs, text sent and received log, step logs, weight logs, and user profile—and we extracted and included for analysis data for users who fulfilled the study criteria.

Dietary logs comprised records in which users entered details about consumption time and names of dishes. Weight logs contained manually entered records of their weight. Step logs included passive data automatically collected from the users' mobile devices. Message logs comprised server records of

message transmission when user messages were sent to, or received from, coaches.

In detail, from an individual i at a T_i length of univariate time series l_i, we obtained an i time series with a heterogeneous length ($l_i \in$ [illegible]^{Ti}). For a time series with a length <30 observations, linear interpolation was adopted because previous research suggested that linear interpolation of weight can estimate missing data [20]. As a moving average can make a time series shorter, the time series needs to be elongated in advance. Thus, for the long length of time series, we sampled values with the same interval and obtained the same length for each time series by resampling. In doing so, we were able to alleviate noise in the time series stemming from within-observer variability; for example, when weight was not measured at the same time every day, we smoothed the data after applying a centered moving average with a window size of 15 observations and finally obtained a time series with a length of 16 observations [21].

Exponential moving average methods such as double exponential smoothing or Holter-Winter smoothing were not considered because these methods use the most recent past value, which is often indicative of the near future rather than the remote past. Furthermore, these methods are more appropriate for economic indicators that entail a trend or seasonality in a long-term perspective [22,23]. To reduce time series noise in the distance between 2 time series, rather than considering the long-term trend or slope over time, a simple moving average was adopted, especially considering that this weight loss program was conducted only for a short period of 16 weeks.

Furthermore, we performed mean-variance scaling to normalize the weight range for each user by adjusting the mean of the time series as 0 and the SD as 1. Subsequently, to identify weight loss trajectories, we performed k-means clustering with dynamic time warping (DTW), which is considered the distance between 2 time series. k-means clustering is one of the most commonly used algorithms for partitioning clusters in which each cluster has a centroid (prototype), which reflects the mean value of its objects (Multimedia Appendix 1) [24].

This clustering algorithm minimizes the total distance between all objects in a cluster and the centroid. In this algorithm, we used DTW as the distance between 2 objects to calculate the total sum of distances. DTW is a similarity measure between time series that considers the shape of 2 time series [25,26]. Using DTW, we could cluster time series with similar patterns of change, regardless of the time points of the weight records [24]. As k-means clustering requires researchers to choose the number of clusters (called k), the quality of clusters may vary by k [27]. To evaluate the quality of clusters, we used elbow methods by inspecting the total sum of distances in all clusters (inertia) through a comparison of the number of clusters from 2 to 10 (Multimedia Appendix 2) [28,29].

Outcome

Our primary objective is to identify weight loss trajectories over 16 weeks among mobile app users who are overweight. The secondary objective is to determine factors related to behavioral and app use characteristics eliciting weight loss.

We defined both initial and final weights. Initial weight does not refer to the weight the user entered when signing up to use the app. Rather, it represents the first weight entered in the period before coaching and 1 week after the commencement of coaching. This first weight logged is closer to the users' actual initial weight because they may not have measured their actual weight before entering it during registration. Before coaching, they may have tended to underestimate their weight. Final weight refers to the last weight measured at week 16 (the last program week) or last weight measured at week 17 only if the user did not enter their weight at week 16. We defined final weight in terms of the percentage of weight loss after 16 weeks (stable weight [<2% change], gain of >2%, loss of 2%-5%, loss of 5%-10%, loss of 10%-15%, and loss of >15% compared with initial weight) [30].

Statistical Analysis

To identify behavioral and app use characteristics in each cluster, we first applied the Kolmogorov-Smirnov test to identify the normality of distribution. According to these results, we performed the Kruskal-Wallis test for normally distributed variables and 1-way ANOVA for normally distributed variables. We used ANOVA for continuous variables to identify whether the means in each k group were equal (H_0:$\mu_1 = \mu_2 , \ldots , = \mu_K$) and the Kruskal-Wallis test to determine whether population medians were equal (null hypothesis) [31].

The chi-square test was used to identify whether the distributions of categorical variables in each group were equal. Among the clusters, we chose three (clusters 1, 2, and 3) that showed a converging pattern over 8 weeks and compared the use characteristics of these 3 clusters before and after 8 weeks. To identify differences in app use and behavioral characteristics among these clusters, we compared characteristics by week, from week 1 to week 16, using 2-sample, 2-tailed t tests and ANOVA.

For each comparison, effect sizes were calculated depending on the type of variable and number of groups compared to determine the possibility of type 1 statistical error. As the statistical significance was not adequate enough to compare groups of a large population, which will almost always demonstrate a significant difference because of statistical power, we additionally calculated effect sizes referring to the magnitude of group differences as means or proportions [32]. For continuous variables, Cohen d was calculated when 2 independent means were compared and eta squared (η^2) was calculated when >2 independent means were compared (ANOVA) [33,34]. For categorical variables, phi (Φ) was calculated for a 2×2 contingency table and Cramer V was used for >2 categories [35].

In this study, a 2-sided P value of <.05 and an effect size greater than small, depending on its type (η^2≈0.01: small, Cramer V≈0.01: small, and Cohen d≈–0.20 to +0.20: small), were considered significant. Statistical analyses were performed with Python (version 3.68; Python Software Foundation) and with the tslearn (version 0.4.1) library [36,37].

Ethics Approval

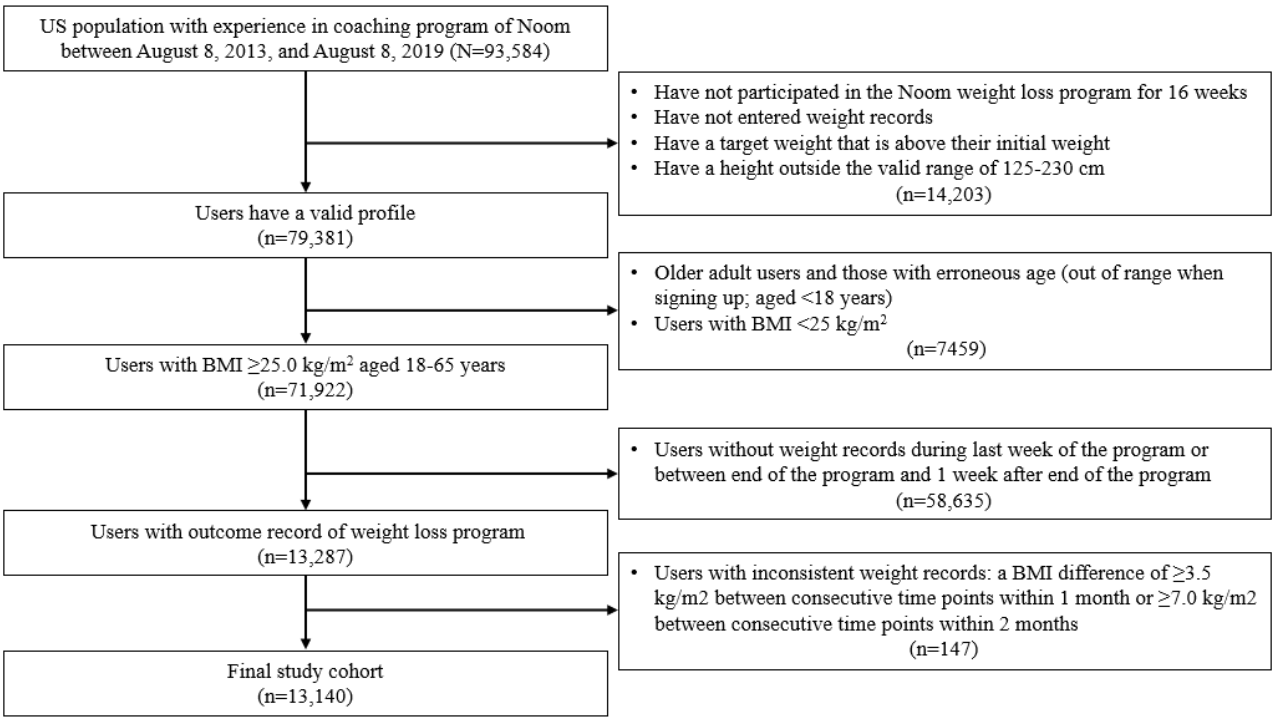
This study was approved by the institutional review board at Advarra (Columbia, Maryland, United States; CR00123125). The deidentified nature of the retrospective log data made obtaining informed consent unnecessary.

Results

Overview

Of the 93,584 unique users who have used Noom, 14,203 (15.18%) were excluded for the following reasons: did not participate in weight loss program, did not enter weight records, target weight more than initial weight, or height out of range. Of the remaining 79,381 users, 7459 (9.40%) were excluded because they were not overweight or they did not meet the age criteria, leaving 71,922 (90.60%) users. Of these 71,922 users, 58,635 (81.53%) were excluded for not completing the program, leaving 13,287 (18.47%) users. Finally, of these 13,287 users, 147 (1.11%) with inconsistent weight records were excluded, leaving 13,140 (98.89%) users who had exhibited adherence to the app (Figure 1).

Figure 1. Process of selection of eligible users.



Baseline Characteristics

At baseline, the proportion of female users (12,093/13,140, 92.03%; Table 1) was larger than that of male users (1047/13,140, 7.9%). The average age of the users was 43.9 (SD 10.9) years, and the average height was 166.4 (SD 7.4) cm. Overall, the mean initial BMI was 33.6 (SD 5.9) kg/m^2, which decreased to 31.6 (SD 5.7) kg/m^2 on average.

Table 1. Baseline characteristics (N=13,140).

Variables	Values
Sex, n (%)	
Female	12,093 (92.03)
Male	1047 (7.9)
Age (years), mean (SD)	43.9 (10.9)
Height (cm), mean (SD)	166.4 (7.4)
Initial weight (kg), mean (SD)	93.2 (18.1)
Initial BMI (kg/m^2), mean (SD)	33.6 (5.9)
Final weight (kg), mean (SD)	87.6 (17.2)
Final BMI (kg/m^2), mean (SD)	31.6 (5.7)
Weight loss (kg), mean (SD)	–5.7 (5.5)
BMI loss (kg/m^2), mean (SD)	–2.0 (1.9)

Weight Loss Trajectories

We explored the optimal number of clusters (k) by changing the number of clusters from 2 to 10. Using elbow methods with visualization, we determined the optimal number of clusters (k) to be 5 (Multimedia Appendix 2) for clustering users with adherence. Each user was assigned to a cluster: 85.96% (11,295/13,140) of the users were assigned to cluster 1, followed by 6.34% (833/13,140) assigned to cluster 2. Users in cluster 1 (sharp decrease) exhibited a decrease in their weight without plateaus. Users in cluster 2 (moderate decrease) initially showed a sharp reduction in weight, after which the slope of weight loss plateaued. Users in cluster 3 (yo-yo) exhibited a decrease in weight, but in the middle of the program, they exhibited a gain in weight. Although cluster 2 (moderate decrease) and cluster 3 (yo-yo) exhibited convergence in weight loss patterns for the initial 8 weeks, users in cluster 2 maintained their weight, whereas those in cluster 3 gained weight after 8 weeks. Users in cluster 4 (stable or increase) gained weight, and those in cluster 5 (other) did not show a convergence pattern because of the partitioned-clustering approach (Figure 2).

There were no significant differences in medians of the initial BMIs of the clusters (η^2<0.001; *P*=.43). However, there was a significant difference in weight reduction class (Cramer V=0.241; *P*<.001). In cluster 1 (sharp decrease), of the 11,295 users, 4541 (40.20%) exhibited weight loss of 5%-10%, which was the largest proportion among the clusters. In addition, this cluster comprised the largest proportion of users recording weight loss ranging from 10% to 15% (2206/11,295, 19.53%). In cluster 2 (moderate decrease), of the 833 users, 281 (33.7%) exhibited weight loss of 2%-5% (Table 2).

Furthermore, the 5 clusters showed different use characteristics in terms of frequency of meal and weight record adherence (η^2=0.056; *P*<.001, and η^2=0.024; *P*<.001, respectively). Cluster 1 users recorded their meals and weight most frequently (median 18.5, IQR 4.1 times per week and median 4.9, IQR 1.8 times per week, respectively), followed by cluster 2 users for meal record adherence (median 16.1, IQR 5.2 times per week). Cluster 2 users entered their weight a median of 4.4 (IQR 1.8) times per week.

The median number of sent and received messages among the clusters did not show a significant difference (η^2<0.001; *P*=.12, and η^2=0.001; *P*=.001, respectively). However, there was a significant difference in the median number of steps (η^2=0.002; *P*<.001).

Figure 2. Clustered weight loss trajectories using k-means with dynamic time warping. Each black line signifies an individual user's weight loss journey. The red line represents the weight loss trajectory of each cluster.

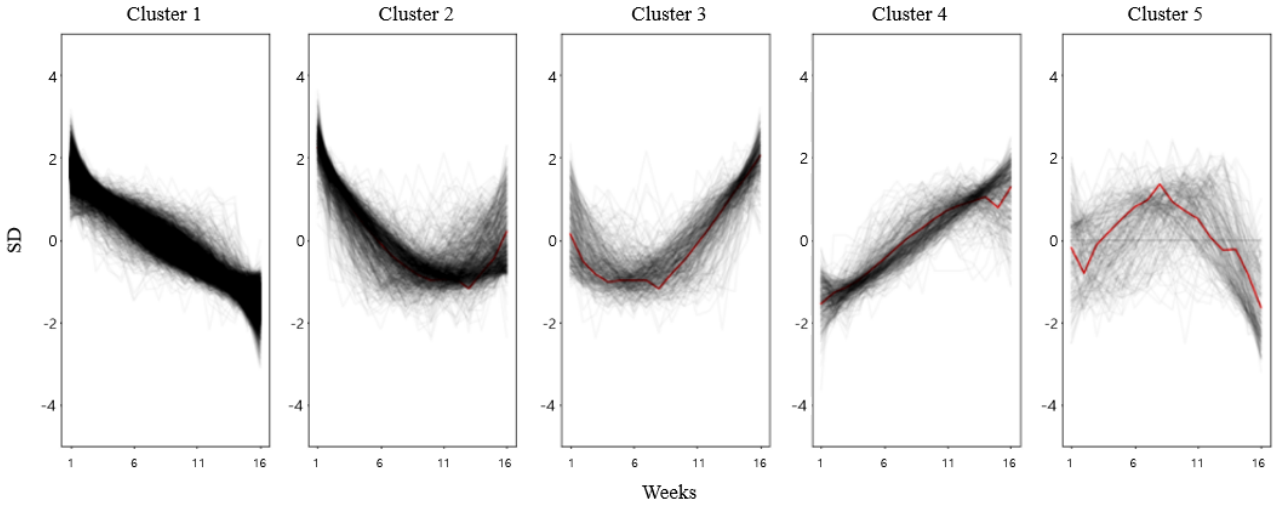


Table 2. Comparison of app use and behavioral characteristics among the 5 clusters (N=13,140).

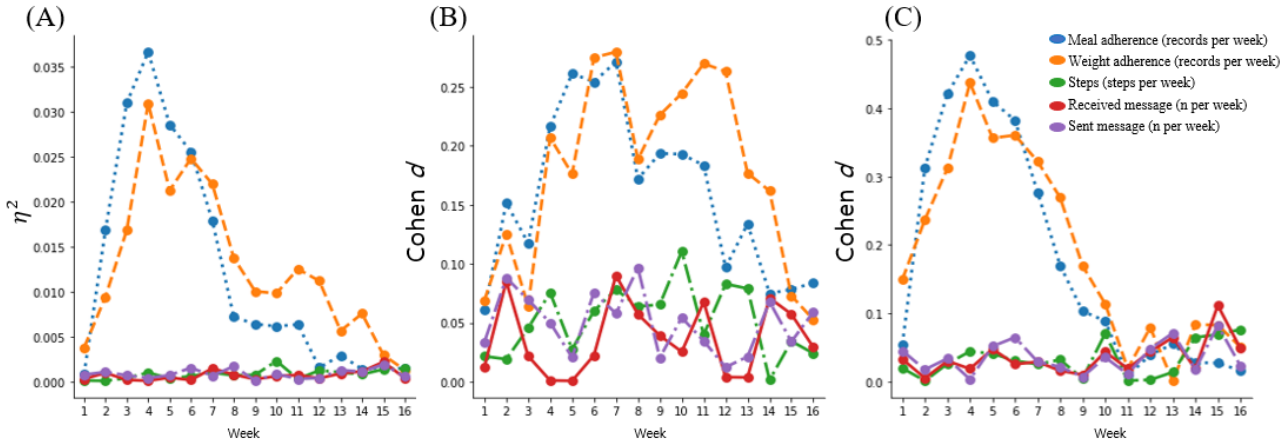
Variables	Clusters					*P* value	Effect size[a]
	Cluster 1 (sharp decrease), n=11,295	Cluster 2 (moderate decrease), n=833	Cluster 3 (yo-yo), n=384	Cluster 4 (stable or increase), n=401	Cluster 5 (other), n=227		
Initial BMI (kg/m^2), median (IQR)	32.5 (7.7)	32.6 (7.5)	32.6 (6.7)	33.5 (7.5)	32.0 (7.4)	.43	<0.001
Weight loss class, n (%)						<.001	0.241
Gained >2%	438 (3.87)	96 (11.52)	94 (24.48)	124 (30.92)	25 (11.01)		
Stable	1107 (9.80)	331 (39.74)	205 (53.39)	195 (48.63)	102 (44.93)		
Lost 2%-5%	2454 (21.73)	281 (33.73)	62 (16.15)	52 (12.97)	72 (31.72)		
Lost 10%-15%	2206 (19.53)	19 (2.28)	10 (2.60)	6 (1.50)	4 (1.76)		
Lost >15%	549 (4.86)	8 (0.96)	0 (0)	8 (2)	2 (0.88)		
Meal record adherence (records per week), median (IQR)	18.5 (4.1)	16.1 (5.2)	15.1 (6.2)	15.5 (6.1)	16.2 (6.7)	<.001	0.056
Weight record adherence (n per week), median (IQR)	4.9 (1.8)	4.4 (1.8)	4.0 (2.0)	4.162 (2.3)	3.6 (2.7)	<.001	0.024
Sent messages (n per week), median (IQR)	2.1 (1.6)	2.1 (1.6)	1.9 (1.4)	1.9 (1.7)	1.9 (1.8)	.12	<0.001
Received messages (n per week), median (IQR)	3.0 (1.7)	3.0 (1.6)	2.8 (1.5)	2.8 (1.925)	2.8 (1.954)	.001	0.001
Steps (per day), median (IQR)	5469.2 (4236.7)	5190.6 (3979.2)	5101.5 (3943.0)	5070.0 (3944.5)	4809.6 (3974.8)	<.001	0.002

[a]Effect size was calculated using eta squared (η^2) for continuous variables and Cramer V for categorical variables (η^2≈0.01: small, η^2≈0.09: moderate, and η^2≈0.25: large; Cramer V≈0.01: small, Cramer V≈0.30: moderate, and Cramer V≈0.50: large).

Weight Regain and Steady Loss

By the middle of the 16-week program, clusters 1, 2, and 3 showed similar weight loss slopes. However, after this point, the trajectory of cluster 3 showed a rebound. Therefore, we further compared the characteristics of these clusters. Longitudinal app use and behavioral characteristics of clusters 1, 2, and 3, including weight record, meal record, and steps, as well as messages sent and received, were plotted over 16 weeks (Figure 3). Overall, although no differences were observed in both meal and weight record adherence among the clusters in the first week, a significant difference peaked within the initial 8 weeks (Figures 3A, 3B, and 3C). The Cohen *d* values of clusters 1 and 2 were maintained between 0.17 and 0.28 even after 8 weeks, and there was a drastic decrease after 15 weeks; however, those of clusters 2 and 3 peaked at 4 weeks and decreased thereafter. Among these clusters, adherence peaked between 2 and 4 weeks and then showed divergence on average, with the highest adherence maintained in the clusters 1, 2, and 3, in that order (Multimedia Appendix 3). In contrast, there were no significant differences in average daily steps or messages sent and received for the entire 16-week period.

Figure 3. Comparison of longitudinal app use and behavioral characteristics for individual clusters. (A) Eta squared from analysis of variance test with clusters 1, 2, and 3. (B) Cohen *d* from *t* test between clusters 1 and 3. (C) Cohen *d* from *t* test between clusters 2 and 3 (η^2≈0.01: small, η^2≈0.09: moderate, and η^2≈0.25: large; Cohen *d*≈–0.20 to +0.20: small, Cohen *d*≈0.50: moderate, and Cohen *d*≈0.80: large).



Discussion

Principal Findings

In this retrospective study, we examined the weight loss trajectories of a large population of individuals who completed a mobile app intervention program. Using a machine learning approach and applying a clustering method and shape-based time series similarity, we found 5 primary clusters. By comparing the use characteristics of each cluster, we found that meal and weight record adherence affected weight trajectory more than the sent and received messages and daily steps.

To evaluate the validity of the clusters, we assessed both cross-sectional and longitudinal differences in use characteristics indicative of cluster attributes, as well as basic clustering validity indices such as inertia. In doing so, we noted that the magnitude of self-monitoring–related app use, such as adherence to recording one's meals and weight, depended more on clustering membership, especially between individuals with a steady weight loss and those who initially lost weight but later regained it. Longitudinally, app use between steadily losing or regaining weight differed greatly after initial use. During the first week, almost all users showed high adherence to the app; however, divergence in adherence began to appear from week 2. Cohen *d* values for meal and weight adherence between cluster 2 (moderate decrease) and cluster 3 (yo-yo) peaked at week 4 and gradually decreased until the end of the program. Although clusters 2 and 3 similarly showed high adherence to recording meal and weight until week 3, adherence decreased at the end of the program (Multimedia Appendix 3): adherence decreased more steeply in cluster 3 than in cluster 2. From these results, we discerned that cluster membership depended largely on how users maintained app and program adherence, reflecting a willingness to perform self-monitoring, from week 3 onward.

Unlike studies that used a variable-centered approach, which assumes that a set of *average* parameters can be estimated for all individuals drawn from a population, this study was based on a *person-centered approach*, categorizing individuals into common subpopulations and determining whether subgroups of similar participants exist [38]. Although our method did not assume any parameters, we found 5 subpopulations that showed differences in use patterns. Our method is consistent with previous research demonstrating that it is necessary to divide an entire population into subgroups and then perform more detailed analysis, rather than to analyze the entire population as a single group [14,39,40]. Considering the well-recognized heterogeneity of weight loss outcomes and variability of outcomes in the intervention period, these results lay the foundation for the design of an obesity management program using mobile phones [41].

Identification of Weight Loss Trajectories

Previously, to discover patterns of weight loss, research has applied statistical clustering methods, principal component analysis, and latent class analysis. In doing so, three primary weight loss patterns over time have been proposed (modest loss, moderate loss, and substantial loss) [14,42-44]. Most of these analyses were based on on-site interventions, although some studies focused on weight loss trajectories using samples from clinical trials. As there might be gaps in the efficacy of interventions between a clinical trial setting and a real-world setting [45], our results, which were obtained using a large data set, could be considered to broadly reflect a real-world context [46]. Thus, our study expounds on previous evidence of mobile-based interventions by considering the shape, and calculating the similarity, of time series with DTW and by comparing the behavioral characteristics of individual app users.

The methods used in this study can be adopted in research related to long-term weight loss or maintenance. It is known that it can be difficult to keep weight off because of various reasons, such as having an obesogenic environment or difficulty in managing physiological responses to weight loss [47]. In addition, hurdles such as the occurrence of a weight loss plateau appear during long-term maintenance. For long-term weight management, it may be possible to identify long-term weight loss trajectories using our method as a patient-centered approach. Although some studies have attempted to identify long-term weight loss trajectories in on-site management [48,49], no study has examined long-term weight trajectories using nonparametric analyses. Further research identifying weight loss plateaus or

rebounding trajectories in the long term could reveal factors contributing to, or mitigating, long-term weight loss.

Application of Weight Loss Trajectories

Obesity management using mobile health technology enables the design of just-in-time adaptive interventions for behavioral support that directly correspond to needs determined by using real-time data collected from user records [50]. Given the use of time series clustering at each week, our method was able to infer membership in a time series despite the different number of observations on each device. Using these data on membership at each time point, clinicians or counselors can estimate what a user's weight trajectory might look like. Accordingly, by providing feedback, they can support the user to self-monitor their weight timely.

Nevertheless, when the scale of 2 time series shows a large difference (eg, weight time series of users weighing 170 kg vs those weighing 70 kg), DTW, the distance between 2 time series, may be increased. To measure only the shape of a time series and not actual weight values, our method reflects factors that may be contributing to weight loss by rescaling. In this study, we comprehensively clustered weight loss patterns among participants who needed to lose weight (ie, those in the preobese or higher category) by adjusting the weight scale.

In addition, research has shown that weight loss trajectories can be related to clinical indicators associated with comorbidity, wherein weight loss patterns corresponded with improvements in blood pressure, triglyceride, and blood glucose levels [49]. Similarly, our approach provides added clinical meaning beyond mere weight loss. In addition, researchers have demonstrated that it is possible to predict the amount of weight a user will lose at the end of a program using interpretable artificial intelligence (AI), which can be used in the coaching process [51]. With explainable AI–used adherence as a feature, adherence becomes a contributing factor to weight loss. Furthermore, the membership of a cluster is interpreted as a contributing factor; for example, membership of cluster 3 (yo-yo) contributed to weight loss, but users in this cluster gained weight after 8 weeks. Interpreting the AI results, coaches can consult with users about their current app use, trajectory, and predicted weight loss, allowing them to intervene early in a program to improve otherwise suboptimal outcomes.

Use Factors Related With Weight Loss

In this study, although individuals were not classified by variables related to self-monitoring, the results revealed that two variables (meal and weight record adherence) showed an association with weight loss after clustering. Among behavioral strategies, self-monitoring is recommended for maintaining lost weight, even when using mobile apps [19,52,53]. It has been posited that self-monitoring serves as the initial step in a feedback loop that includes observation and recording, self-evaluation, and self-reinforcement, which can help individuals decide to adjust behaviors [54]. Individuals with a high meal adherence may review their dietary behavior more frequently and arrive at opportunities for self-evaluation and self-reinforcement.

Although we did not explicitly show that weight loss of 5% to 10% would be suitable as a cutoff for responders, our findings indicated that among individuals who used the app, those who adhered to the app more regularly lost more weight, which could be regarded as a weight loss response. Nevertheless, our findings also showed that the weight trajectory can rise again if compliance is not maintained, despite being high initially. Sufficient evidence suggests that early weight loss can be a predictive factor for long-term weight loss [55-58]. Similarly, our results showed that an early use pattern can also be a predictive factor for weight loss outcomes in a 16-week program. This study compared a group that lost and regained weight with a group that exhibited a moderate decrease in weight. The 2 groups showed a difference in app adherence after only 2 weeks, and this difference increased continuously for 8 weeks. In line with our results, a previous study also showed that early intervention affects short- and long-term weight loss in a weight loss program, although it did not examine web-based or mobile programs [59].

Limitations

Our study includes some inherent limitations because of its selection of participants and the retention rate, which could pose a selection bias. First, the participants in this study had purchased a subscription to a weight loss program. As our sampled participants may have had a higher intention to lose weight than individuals who use free web-based interventions, the former may have exhibited a relatively higher retention rate than the latter. The average 30-day retention rate for health and fitness apps in the United States is 3.4% and that for mental health apps is 3%-8% [60-62]. Considering these values, some selection bias may have occurred in this study.

Second, this work may have introduced latent sampling bias through the process of participant selection because of the criterion by which we included only users with weight records during the last week of the program or between the end of the program and 1 week after the end of the program. Although it was necessary to include this criterion to identify the final outcome of the weight loss program, it may have led to the inclusion of users who may have been more compliant and more likely to be successful with weight loss. Therefore, our findings need to be interpreted considering these limitations. In addition, our participants were selected from users who entered their records as a minimal requirement, which could affect the interpretation and generalization of our findings. For example, because some of the users who did not enter their records showed nonadherence, the factors related to use patterns may not predict their weight trajectories. Our results may be reliable for users who have at least minimal adherence. If the inclusion criteria were stricter, such as recording all 3 meals 1 day per week or recording 1 meal per day per week, approximately 3000 or 15,000, respectively, of the participants would have been additionally excluded. Therefore, we set inclusion criteria that required 1 recording of a meal per week as a condition for minimal self-monitoring.

In addition, our participants consisted of 92.03% (12,093/13,140) women and 7.9% (1047/13,140) men. Although it can be thought that there may be a bias in the selection process

of the participants, with a high proportion of women, statistics for participants in previous studies of mobile weight loss apps showed high proportions of women [63-65]. Although there are some magnitudes of difference in the proportion of women among the participants, our may be adopted in that the statistics of participants in previous studies were consistent with ours.

In addition, we conducted an experiment to identify whether a latent selection bias was present in the exclusion of participants with inconsistent weight in 2 consecutive records (Multimedia Appendix 4). Despite including these participants, the number of optimal clusters remained at 5, and each cluster showed the same 5 shapes presented in this paper. Although the membership of each cluster differed (clusters with participants with erroneous records comprised 11,184, 955, 363, 575, and 245 users, respectively, for clusters 1 to 5, whereas those excluding these participants comprised 11,295, 833, 384, 401, and 277 users, respectively), the comparative analysis showed consistent results, wherein (1) the mean initial BMI did not differ significantly; (2) after the program, cluster 1 users exhibited the highest weight loss; and (3) weight and meal record adherence was relatively higher in clusters 1 and 2 than in the other clusters.

In terms of data reliability, although we accounted for noise in the weight loss time series, it was difficult to determine whether each user had correctly entered meal skipping and weight (data reliability stems from user entries).

Finally, although sustainable weight loss in the long term is important in obesity management, we did not explore long-term outcomes because of missing values from the users' logs. Thus, a long-term cohort study is needed to obtain more reliable evidence on long-term outcomes. Despite these limitations, our work provides evidence indicating that weight loss trajectories depend on overall adherence and early adherence to self-monitoring in a limited observation period (16 weeks).

Conclusions

Using time series clustering, we identified 5 distinct profiles of weight change over a 16-week weight management intervention through a mobile app. We found that overall adherence and early adherence to self-monitoring could be predictive factors for greater weight loss success.

Acknowledgments

This study was supported by a grant from the National Research Foundation of Korea funded by the South Korean government (NRF-2019M3E5D4064682), the Foundational Technology Development Program (NRF2019M3E5D406468212) funded by the South Korean Ministry of Science and ICT, and the Bio-Industrial Technology Development Program (20014841) funded by the South Korean Ministry of Trade, Industry, and Energy.

Conflicts of Interest

AM is employed by Noom, Inc.

Multimedia Appendix 1
Pseudocode for weight loss trajectories: k-means clustering with dynamic time warping.
[DOCX File , 21 KB - jmir_v24i4e29380_app1.docx]

Multimedia Appendix 2
Sum of distance with reference to number of clusters (excluding users with inconsistent weight records).
[DOCX File , 89 KB - jmir_v24i4e29380_app2.docx]

Multimedia Appendix 3
Boxplot of the use factors by week.
[DOCX File , 91 KB - jmir_v24i4e29380_app3.docx]

Multimedia Appendix 4
Robustness of clustering with modified selection flow.
[DOCX File , 445 KB - jmir_v24i4e29380_app4.docx]

References

1. The GBD 2015 Obesity Collaborators. Health effects of overweight and obesity in 195 countries over 25 years. N Engl J Med 2017 Jul 06;377(1):13-27. [doi: 10.1056/nejmoa1614362]
2. Apovian CM. The obesity epidemic — understanding the disease and the treatment. N Engl J Med 2016 Jan 14;374(2):177-179. [doi: 10.1056/nejme1514957]
3. The device trends to know. GWI. URL: https://www.gwi.com/reports/device [accessed 2022-03-13]
4. Kessler D. The End of Overeating Taking Control of the Insatiable American Appetite. Pennsylvania, United States: Rodale Books; 2010.

5. Tang J, Abraham C, Greaves C, Yates T. Self-directed interventions to promote weight loss: a systematic review of reviews. J Med Internet Res 2014 Feb 19;16(2):e58 [FREE Full text] [doi: 10.2196/jmir.2857] [Medline: 24554464]
6. Nikolaou CK, Lean ME. Mobile applications for obesity and weight management: current market characteristics. Int J Obes (Lond) 2017 Jan;41(1):200-202. [doi: 10.1038/ijo.2016.186] [Medline: 27780974]
7. Haapala I, Barengo NC, Biggs S, Surakka L, Manninen P. Weight loss by mobile phone: a 1-year effectiveness study. Public Health Nutr 2009 Dec;12(12):2382-2391. [doi: 10.1017/S1368980009005230] [Medline: 19323865]
8. Liu F, Kong X, Cao J, Chen S, Li C, Huang J, et al. Mobile phone intervention and weight loss among overweight and obese adults: a meta-analysis of randomized controlled trials. Am J Epidemiol 2015 Mar 01;181(5):337-348 [FREE Full text] [doi: 10.1093/aje/kwu260] [Medline: 25673817]
9. Wickham C, Carbone ET. Who's calling for weight loss? A systematic review of mobile phone weight loss programs for adolescents. Nutr Rev 2015 Jun;73(6):386-398. [doi: 10.1093/nutrit/nuu018] [Medline: 26011913]
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. Douketis JD, Macie C, Thabane L, Williamson DF. Systematic review of long-term weight loss studies in obese adults: clinical significance and applicability to clinical practice. Int J Obes (Lond) 2005 Oct;29(10):1153-1167. [doi: 10.1038/sj.ijo.0802982] [Medline: 15997250]
12. Dhurandhar NV, Kyle T, Stevenin B, Tomaszewski K, ACTION Steering Group. Predictors of weight loss outcomes in obesity care: results of the national ACTION study. BMC Public Health 2019 Oct 30;19(1):1422 [FREE Full text] [doi: 10.1186/s12889-019-7669-1] [Medline: 31666040]
13. Michaelides A, Major J, Pienkosz E, Wood M, Kim Y, Toro-Ramos T. Usefulness of a novel mobile diabetes prevention program delivery platform with human coaching: 65-week observational follow-up. JMIR Mhealth Uhealth 2018 May 03;6(5):e93 [FREE Full text] [doi: 10.2196/mhealth.9161] [Medline: 29724709]
14. Yank V, Xiao L, Wilson SR, Stafford RS, Rosas LG, Ma J. Short-term weight loss patterns, baseline predictors, and longer-term follow-up within a randomized controlled trial. Obesity (Silver Spring) 2014 Jan 15;22(1):45-51 [FREE Full text] [doi: 10.1002/oby.20510] [Medline: 23740619]
15. de Vos BC, Runhaar J, Verkleij SP, van Middelkoop M, Bierma-Zeinstra SM. Latent class growth analysis successfully identified subgroups of participants during a weight loss intervention trial. J Clin Epidemiol 2014 Aug;67(8):947-951. [doi: 10.1016/j.jclinepi.2014.03.007] [Medline: 24774470]
16. Fallaize R, Zenun Franco R, Pasang J, Hwang F, Lovegrove JA. Popular nutrition-related mobile apps: an agreement assessment against a UK reference method. JMIR Mhealth Uhealth 2019 Feb 20;7(2):e9838 [FREE Full text] [doi: 10.2196/mhealth.9838] [Medline: 30785409]
17. National health and nutrition examination survey. Centers for Disease Control and Prevention. URL: https://wwwn.cdc.gov/Nchs/Nhanes/2017-2018/P_BMX.htm#BMIHT [accessed 2022-03-13]
18. Defining adult overweight and obesity. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/obesity/adult/defining.html [accessed 2022-03-13]
19. Chin SO, Keum C, Woo J, Park J, Choi HJ, Woo J, et al. Successful weight reduction and maintenance by using a smartphone application in those with overweight and obesity. Sci Rep 2016 Nov 07;6(1):34563 [FREE Full text] [doi: 10.1038/srep34563] [Medline: 27819345]
20. Turicchi J, O'Driscoll R, Finlayson G, Duarte C, Palmeira AL, Larsen SC, et al. Data imputation and body weight variability calculation using linear and nonlinear methods in data collected from digital smart scales: simulation and validation study. JMIR Mhealth Uhealth 2020 Sep 11;8(9):e17977 [FREE Full text] [doi: 10.2196/17977] [Medline: 32915155]
21. Ellis CA, Parbery SA. Is smarter better? A comparison of adaptive, and simple moving average trading strategies. Res Int Business Finance 2005 Sep;19(3):399-411. [doi: 10.1016/j.ribaf.2004.12.009]
22. Pozzi F, Di Matteo T, Aste T. Exponential smoothing weighted correlations. Eur Phys J B 2012 Jun 4;85(6). [doi: 10.1140/epjb/e2012-20697-x]
23. Time series based forecasting techniques using holt-winters. HCL. URL: https://www.hcltech.com/blogs/time-series-based-forecasting-techniques-using-holt-winters#:~:text=Holt%2DWinter%20is%20used%20for,gamma%20parameters%20in%20Holt's%20model [accessed 2022-03-13]
24. Aghabozorgi S, Seyed Shirkhorshidi A, Ying Wah T. Time-series clustering – A decade review. Inf Syst 2015 Oct;53:16-38. [doi: 10.1016/j.is.2015.04.007]
25. Sakoe H, Chiba S. Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans Acoust Speech Signal Process 1978 Feb;26(1):43-49. [doi: 10.1109/tassp.1978.1163055]
26. Ney H, Ortmanns S. Dynamic programming search for continuous speech recognition. IEEE Signal Process Mag 1999;16(5):64-83. [doi: 10.1109/79.790984]
27. Hamerly G, Elkan C. Learning the k in k-means. In: Proceedings of the 16th International Conference on Neural Information Processing Systems. 2003 Presented at: NIPS'03: Proceedings of the 16th International Conference on Neural Information Processing Systems; Dec 9 - 11, 2003; Whistler British Columbia Canada.

28. Kodinariya T, Makwana P. Review on determining number of cluster in K-means clustering. Int J Advance Res Comput Sci Manag Stud 2013;1(6):90-95 [FREE Full text]
29. Syakur MA, Khotimah BK, Rochman EM, Satoto BD. Integration k-means clustering method and elbow method for identification of the best customer profile cluster. In: Proceedings of the IOP Conference Series: Materials Science and Engineering and the 2nd International Conference on Vocational Education and Electrical Engineering (ICVEE). 2017 Presented at: IOP Conference Series: Materials Science and Engineering and the 2nd International Conference on Vocational Education and Electrical Engineering (ICVEE); Nov 9, 2017; Surabaya, Indonesia. [doi: 10.1088/1757-899x/336/1/012017]
30. Wing RR, Lang W, Wadden TA, Safford M, Knowler WC, Bertoni AG, Look AHEAD Research Group. Benefits of modest weight loss in improving cardiovascular risk factors in overweight and obese individuals with type 2 diabetes. Diabetes Care 2011 Jul;34(7):1481-1486 [FREE Full text] [doi: 10.2337/dc10-2415] [Medline: 21593294]
31. Kim H. Analysis of variance (ANOVA) comparing means of more than two groups. Restor Dent Endod 2014 Feb;39(1):74-77 [FREE Full text] [doi: 10.5395/rde.2014.39.1.74] [Medline: 24516834]
32. Sullivan GM, Feinn R. Using effect size-or why the P value is not enough. J Grad Med Educ 2012 Sep;4(3):279-282 [FREE Full text] [doi: 10.4300/JGME-D-12-00156.1] [Medline: 23997866]
33. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
34. Field A. Discovering Statistics Using SPSS. London: SAGE Publications; 2009.
35. Ellis P. The Essential Guide to Effect Sizes Statistical Power, Meta-Analysis, and the Interpretation of Research Results. Cambridge, UK: Cambridge University Press; 2010.
36. Osteen P, Bright C. Effect sizes and intervention research. In: Proceedings of the Society for Social Work and Research Conference. 2010 Presented at: Society for Social Work and Research Conference; Jan 14-17, 2010; San Francisco, CA.
37. Tavenard R, Faouzi J, Vandewiele G, Divo F, Androz G, Holtz C, et al. Tslearn, a machine learning toolkit for time series data. JMLR 2020;21(118):1-6 [FREE Full text]
38. Howard MC, Hoffman ME. Variable-centered, person-centered, and person-specific approaches. Org Res Methods 2017 Dec 10;21(4):846-876. [doi: 10.1177/1094428117744021]
39. Kumanyika S, Espeland M, Bahnson J, Bottom J, Charleston J, Folmar S, TONE Cooperative Research Group. Ethnic comparison of weight loss in the trial of nonpharmacologic interventions in the elderly. Obes Res 2002 Feb;10(2):96-106 [FREE Full text] [doi: 10.1038/oby.2002.16] [Medline: 11836455]
40. Teixeira PJ, Going SB, Sardinha LB, Lohman TG. A review of psychosocial pre-treatment predictors of weight control. Obes Rev 2005 Feb;6(1):43-65. [doi: 10.1111/j.1467-789X.2005.00166.x] [Medline: 15655038]
41. Marek RJ, Coulon SM, Brown JD, Lydecker JA, Marek S, Malcolm R, et al. Characteristics of weight loss trajectories in a comprehensive lifestyle intervention. Obesity (Silver Spring) 2017 Dec 31;25(12):2062-2067. [doi: 10.1002/oby.21942] [Medline: 29086487]
42. Neiberg RH, Wing RR, Bray GA, Reboussin DM, Rickman AD, Johnson KC, Look AHEAD Research Group. Patterns of weight change associated with long-term weight change and cardiovascular disease risk factors in the Look AHEAD Study. Obesity (Silver Spring) 2012 Oct;20(10):2048-2056 [FREE Full text] [doi: 10.1038/oby.2012.33] [Medline: 22327053]
43. Szabo-Reed AN, Lee J, Ptomey L, Willis E, Schubert M, Washburn R, et al. Longitudinal weight loss patterns and their behavioral and demographic associations. Ann Behav Med 2016 Feb 30;50(1):147-156 [FREE Full text] [doi: 10.1007/s12160-015-9740-1] [Medline: 26423446]
44. Morales KH, Kumanyika SK, Fassbender JE, Good J, Localio AR, Wadden TA. Patterns of weight change in black Americans: pooled analysis from three behavioral weight loss trials. Obesity (Silver Spring) 2014 Dec 24;22(12):2632-2640 [FREE Full text] [doi: 10.1002/oby.20904] [Medline: 25251464]
45. Tinetti ME. The gap between clinical trials and the real world: extrapolating treatment effects from younger to older adults. JAMA Intern Med 2014 Mar 01;174(3):397-398. [doi: 10.1001/jamainternmed.2013.13283] [Medline: 24424303]
46. Kim H, Lee S, Kim JH. Real-world evidence versus randomized controlled trial: clinical research based on electronic medical records. J Korean Med Sci 2018 Aug 20;33(34):e213 [FREE Full text] [doi: 10.3346/jkms.2018.33.e213] [Medline: 30127705]
47. 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]
48. Gray CM, Wyke S, Zhang R, Anderson AS, Barry S, Boyer N, et al. Long-term weight loss trajectories following participation in a randomised controlled trial of a weight management programme for men delivered through professional football clubs: a longitudinal cohort study and economic evaluation. Int J Behav Nutr Phys Act 2018 Jun 28;15(1):60 [FREE Full text] [doi: 10.1186/s12966-018-0683-3] [Medline: 29954449]
49. Kuk JL, Wharton S. Differences in weight change trajectory patterns in a publicly funded adult weight management centre. Obes Sci Pract 2016 Jun;2(2):215-223 [FREE Full text] [doi: 10.1002/osp4.35] [Medline: 29071099]
50. Hardeman W, Houghton J, Lane K, Jones A, Naughton F. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. Int J Behav Nutr Phys Act 2019 Apr 03;16(1):31 [FREE Full text] [doi: 10.1186/s12966-019-0792-7] [Medline: 30943983]

51. Kim HH, Kim Y, Park YR. Interpretable conditional recurrent neural network for weight change prediction: algorithm development and validation study. JMIR Mhealth Uhealth 2021 Mar 29;9(3):e22183 [FREE Full text] [doi: 10.2196/22183] [Medline: 33779574]
52. Jensen M, Ryan D, Apovian C, Ard J, Comuzzie A, Donato K, et al. 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American college of cardiology/American heart association task force on practice guidelines and the obesity society. J Am College Cardiol 2014 Jul;63(25):2985-3023. [doi: 10.1016/j.jacc.2013.11.004]
53. Jacobs S, Radnitz C, Hildebrandt T. Adherence as a predictor of weight loss in a commonly used smartphone application. Obes Res Clin Pract 2017 Mar;11(2):206-214. [doi: 10.1016/j.orcp.2016.05.001] [Medline: 27292942]
54. Laitner MH, Minski SA, Perri MG. The role of self-monitoring in the maintenance of weight loss success. Eat Behav 2016 Apr;21:193-197 [FREE Full text] [doi: 10.1016/j.eatbeh.2016.03.005] [Medline: 26974582]
55. Smith SR, O'Neil PM, Astrup A, Finer N, Sanchez-Kam M, Fraher K, et al. Early weight loss while on lorcaserin, diet and exercise as a predictor of week 52 weight-loss outcomes. Obesity (Silver Spring) 2014 Oct 18;22(10):2137-2146 [FREE Full text] [doi: 10.1002/oby.20841] [Medline: 25044799]
56. Fujioka K, Plodkowski R, O'Neil PM, Gilder K, Walsh B, Greenway FL. The relationship between early weight loss and weight loss at 1 year with naltrexone ER/bupropion ER combination therapy. Int J Obes (Lond) 2016 Sep 22;40(9):1369-1375. [doi: 10.1038/ijo.2016.67] [Medline: 27328752]
57. Fujioka K, O'Neil PM, Davies M, Greenway F, Lau DCW, Claudius B, et al. Early weight loss with liraglutide 3.0 mg predicts 1-year weight loss and is associated with improvements in clinical markers. Obesity (Silver Spring) 2016 Nov 02;24(11):2278-2288 [FREE Full text] [doi: 10.1002/oby.21629] [Medline: 27804269]
58. Tronieri J, Wadden T, Chao A, Pearl R, Alamuddin N, Berkowitz R. Early weight loss in behavioral treatment predicts later rate of weight loss and response to pharmacotherapy. Ann Behav Med 2019 Mar 01;53(3):290-295 [FREE Full text] [doi: 10.1093/abm/kay036] [Medline: 29800080]
59. Williamson DA, Anton SD, Han H, Champagne CM, Allen R, Leblanc E, et al. Early behavioral adherence predicts short and long-term weight loss in the POUNDS LOST study. J Behav Med 2010 Aug;33(4):305-314 [FREE Full text] [doi: 10.1007/s10865-010-9253-0] [Medline: 20195742]
60. App engagement: the matrix reloaded. Flurry. URL: https://www.flurry.com/post/113379517625/app-engagement-the-matrix-reloaded [accessed 2022-03-10]
61. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
62. Tison G, Hsu K, Hsieh J, Ballinger B, Pletcher M, Marcus G. Achieving high retention in mobile health research using design principles adopted from widely popular consumer mobile apps. Circulation 2017;136(suppl_1):A21029. [doi: 10.1161/circ.136.suppl_1.21029]
63. Kim Y, Oh B, Shin HY. Effect of mHealth with offline antiobesity treatment in a community-based weight management program: cross-sectional study. JMIR Mhealth Uhealth 2020;8(1):e13273 [FREE Full text] [doi: 10.2196/13273] [Medline: 31961335]
64. Krukowski RA, Harvey-Berino J, Ashikaga T, Thomas CS, Micco N. Internet-based weight control: the relationship between web features and weight loss. Telemed J E Health 2008;14(8):775-782 [FREE Full text] [doi: 10.1089/tmj.2007.0132] [Medline: 18954247]
65. Alexander E, Tseng E, Durkin N, Jerome GJ, Dalcin A, Appel LJ, et al. Factors associated with early dropout in an employer-based commercial weight-loss program. Obes Sci Pract 2018;4(6):545-553 [FREE Full text] [doi: 10.1002/osp4.304] [Medline: 30574348]

Abbreviations

AI: artificial intelligence
DTW: dynamic time warping

Edited by T Leung; submitted 05.04.21; peer-reviewed by A Leichtle, J Buekers; comments to author 27.05.21; revised version received 21.07.21; accepted 17.02.22; published 15.04.22.

Please cite as:
Kim HH, Kim Y, Michaelides A, Park YR
Weight Loss Trajectories and Related Factors in a 16-Week Mobile Obesity Intervention Program: Retrospective Observational Study
J Med Internet Res 2022;24(4):e29380
URL: https://www.jmir.org/2022/4/e29380
doi:10.2196/29380
PMID:35436211

Original Paper

Sleep Disturbance and Quality of Life in Rheumatoid Arthritis: Prospective mHealth Study

John McBeth[1,2], PhD; William G Dixon[1,2], PhD; Susan Mary Moore[1]; Bruce Hellman[3], MBA; Ben James[3], BA; Simon D Kyle[4], PhD; Mark Lunt[1], PhD; Lis Cordingley[5], PhD; Belay Birlie Yimer[1*], PhD; Katie L Druce[1*], PhD

[1]Centre for Epidemiology Versus Arthritis, University of Manchester, Manchester, United Kingdom

[2]National Institute for Health Research Manchester Musculoskeletal Biomedical Research Centre, Central Manchester University Hospitals National Health Service Foundation Trust, Manchester, United Kingdom

[3]uMotif, London, United Kingdom

[4]Sleep and Circadian Neuroscience Institute, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom

[5]Division of Musculoskeletal and Dermatological Sciences, Manchester University, Manchester, United Kingdom

[*]these authors contributed equally

Corresponding Author:
John McBeth, PhD
Centre for Epidemiology Versus Arthritis
University of Manchester
Stopford Building
Oxford Road
Manchester, M13 9PT
United Kingdom
Phone: 44 1612755788
Email: john.mcbeth@manchester.ac.uk

Abstract

Background: Sleep disturbances and poor health-related quality of life (HRQoL) are common in people with rheumatoid arthritis (RA). Sleep disturbances, such as less total sleep time, more waking periods after sleep onset, and higher levels of nonrestorative sleep, may be a driver of HRQoL. However, understanding whether these sleep disturbances reduce HRQoL has, to date, been challenging because of the need to collect complex time-varying data at high resolution. Such data collection is now made possible by the widespread availability and use of mobile health (mHealth) technologies.

Objective: This mHealth study aimed to test whether sleep disturbance (both absolute values and variability) causes poor HRQoL.

Methods: The quality of life, sleep, and RA study was a prospective mHealth study of adults with RA. Participants completed a baseline questionnaire, wore a triaxial accelerometer for 30 days to objectively assess sleep, and provided daily reports via a smartphone app that assessed sleep (Consensus Sleep Diary), pain, fatigue, mood, and other symptoms. Participants completed the World Health Organization Quality of Life-Brief (WHOQoL-BREF) questionnaire every 10 days. Multilevel modeling tested the relationship between sleep variables and the WHOQoL-BREF domains (physical, psychological, environmental, and social).

Results: Of the 268 recruited participants, 254 were included in the analysis. Across all WHOQoL-BREF domains, participants' scores were lower than the population average. Consensus Sleep Diary sleep parameters predicted the WHOQoL-BREF domain scores. For example, for each hour increase in the total time asleep physical domain scores increased by 1.11 points (β=1.11, 95% CI 0.07-2.15) and social domain scores increased by 1.65 points. These associations were not explained by sociodemographic and lifestyle factors, disease activity, medication use, anxiety levels, sleep quality, or clinical sleep disorders. However, these changes were attenuated and no longer significant when pain, fatigue, and mood were included in the model. Increased variability in total time asleep was associated with poorer physical and psychological domain scores, independent of all covariates. There was no association between actigraphy-measured sleep and WHOQoL-BREF.

Conclusions: Optimizing total sleep time, increasing sleep efficiency, decreasing sleep onset latency, and reducing variability in total sleep time could improve HRQoL in people with RA.

(J Med Internet Res 2022;24(4):e32825) doi:10.2196/32825

KEYWORDS

mobile health; sleep; rheumatoid arthritis; pain; fatigue; mood; sleep disturbance; HRQoL; quality of life; health-related quality of life; QoL; sleep efficiency; WHOQoL-BREF; mobile phone

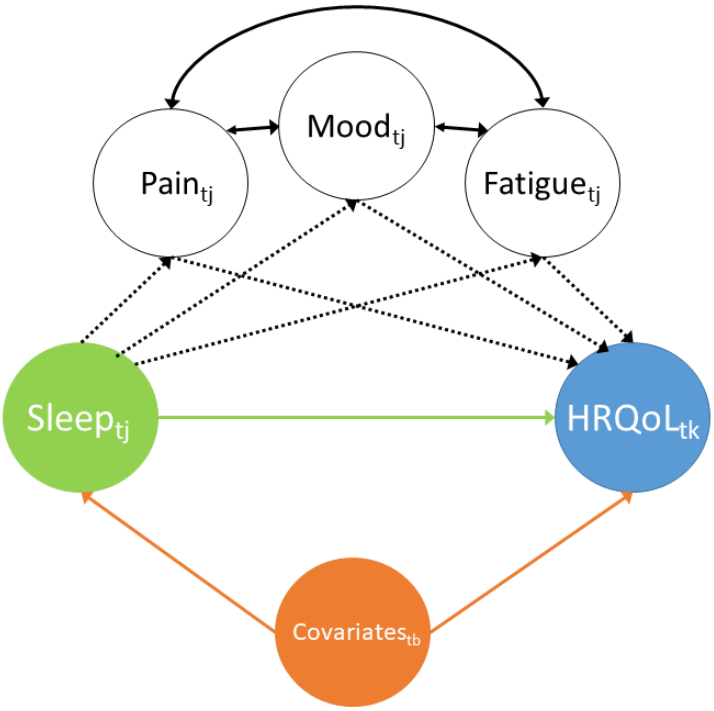
Introduction

Background

People living with rheumatoid arthritis (RA), a long-term progressive autoimmune disease, experience a significantly reduced health-related quality of life (HRQoL), which can be characterized as the impact a condition has on physical, emotional, and social well-being. People with RA have poorer HRQoL compared with patients with other rheumatic diseases [1] and the general population [2]. There are likely numerous causes for poor HRQoL. RA disease activity is a major contributor to lower HRQoL, although HRQoL remains significantly lower than that of the general population, even in those with well-controlled disease. Sleep disturbances are common in RA [3] and have been identified by patients as a possible driver of low HRQoL [4-6].

Studies of sleep in RA have reported less total sleep time, more waking periods after sleep onset, higher levels of nonrestorative sleep [5], and increased periods of mini arousal [4]. During a disease flare, people with RA experience more fragmented sleep, shorter total sleep time, and lower sleep efficiency [7-9]. However, few studies have determined the relationship between sleep variables and HRQoL and understanding whether these sleep disturbances reduce HRQoL remains a challenge. First, sleep is a multifaceted behavior comprising both objective and subjective components [10]. Thus, a comprehensive assessment of sleep health requires measurement of objective and self-reported sleep domains, including appraisals of sleep quality and quantitative estimates of sleep continuity and duration [11]. Despite this, it is only subjective sleep which has been commonly measured in epidemiological studies because, historically, it has been difficult to objectively measure sleep outside artificial laboratory settings. Many studies have also tended to be cross-sectional, despite the high degree of between-day variability, with individuals fluctuating between good and poor sleep states [12]. In addition, sleep disturbance increases the severity of common RA symptoms, including pain, mood, and fatigue, which are known to cause poor HRQoL [13]. Whether poor HRQoL in people with RA is a direct effect of sleep disturbance or an indirect consequence of changes in the severity of pain, mood, and fatigue (Figure 1) is not clear. Understanding these relationships would inform the development of interventions to improve HRQoL. Capturing these complex time-varying data with sufficiently high resolution to understand these relationships has been made possible by the widespread availability and use of mobile health (mHealth) technologies. mHealth technologies, including smartphone apps and wearables, allow frequent and repeated remote collection of patient-generated symptoms and other health data and objective assessments of sleep [14].

Figure 1. Directed acyclic graph for the relationship among sleep, health-related quality of life (HRQoL), pain, mood, and fatigue. The likelihood of reporting a particular level of health-related quality of life at days 10, 20, or 30 ($HRQoL_{tk}$) is directly predicted by sleep ($Sleep_{tj}$) in the previous 10 days (green arrow) as well as the effect of $Sleep_{tj}$ acting through pain ($Pain_{tj}$), mood ($Mood_{tj}$), and fatigue ($Fatigue_{tj}$) in the previous 10 days (black dashed lines). Pain, fatigue, and mood increase the likelihood of each other (black solid lines). The relationship may be confounded by covariates measured at baseline including age, sex, and disease severity ($Covariates_{tb}$; orange arrows).



Objectives

This prospective mHealth study tested the hypothesis that sleep disturbance (both absolute measures and variability) in people with RA would predict poor HRQoL. We then tested whether any observed relationship was explained by the effect of sleep on pain, mood, and fatigue severity.

Methods

Overview

The quality of life, sleep, and rheumatoid arthritis (QUASAR) study collected daily data from people living with RA for 30 days. The participants completed a baseline questionnaire and wore a triaxial accelerometer to assess sleep over the 30-day study period. On a patient co-designed smartphone app (Figure S1 in Multimedia Appendix 1) developed in collaboration with uMotif, participants completed a daily sleep diary; provided daily reports of the severity of their pain, fatigue, and mood; and completed a quality of life questionnaire every 10 days after baseline. QUASAR has been described in detail elsewhere [14], and the methods are summarized in the following sections.

Participants

Eligible participants included those aged ≥18 years, with RA (classified as self-reported clinical diagnosis of RA and currently using disease-modifying antirheumatic drugs [15]), who had access to an Android or iPhone operating system (Apple Inc) smartphone or tablet, and who were not employed in shift work.

Recruitment

Participants were recruited from May 1, 2017, to July 13, 2018, via an email sent to people registered on the electronic mailing list of the National Rheumatoid Arthritis Society, a UK-wide patient organization. The email contained an electronic study information pack, which included a participant information sheet, a copy of the study consent form, and a link to complete a web-based screening questionnaire. The screening questionnaire collected data on study eligibility criteria, contact information, and consent for further contact. Eligible participants were telephoned after at least 24 hours of questionnaire submission to discuss the project. Verbal consent was obtained, and a study pack (written consent form, baseline questionnaire, actigraph, and study instructions) was mailed to participants in time for the agreed study start date.

Data Collected

Baseline Questionnaire

Data were collected on sex (male or female), date of birth (day, month, and year), and date of RA diagnosis (month and year); BMI (self-reported weight in kilograms/height in meters^2) categorized as underweight (<18.5 kg/m^2), healthy (18.5-24.9 kg/m^2), overweight (24.9-30 kg/m^2), or obese (>30 kg/m^2); marital status (single, married or with partner, or separated); smoking (past, never, or current smoker); average weekly alcohol consumption (none: 0 units, moderate: 1-14 units, and high: ≥15 units); total number of medication types (range 0-4 from categories of painkillers, disease-modifying antirheumatic drugs, sleep medications, and others [free text]); and Index of Multiple Deprivation (English, 2015; Scottish, 2016; or Welsh, 2019, as appropriate) derived from the first part of participants' postcodes.

Baseline sleep quality was assessed using the Pittsburgh Sleep Quality Index (PSQI) [16] (score range 0-21, higher scores indicating worse sleep quality). Insomnia was assessed using the Sleep Condition Indicator (SCI) [17], and participants reported physician-diagnosed obstructive sleep apnea (OSA) and restless leg syndrome (RLS). Anxiety was measured using the anxiety subscale of the Hospital Anxiety and Depression Scale (score range: 0-21; categorized as *not a case*: 0-8, *borderline case*: 8-11, and *case*: 11-21) [18].

Disease severity was assessed using the Routine Assessment of Patient Index Data 3 (RAPID-3) [19]. The 15-item RAPID-3 measures three domains: physical function, pain, and global health in the past week. The first 10 items of the physical function domain were scored, transformed into a 0.3-10 scale and summed with the pain and global health domains to produce an overall score of 0-30. RAPID-3 scores are correlated with the disease activity score 28 and clinical disease activity index in clinical trials and clinical care [20].

Sleep Assessments

Consensus Sleep Diary

Each morning at 8 AM, participants were prompted via an alert in the study app to complete the 10-item Consensus Sleep Diary (CSD), which assesses the quantity and quality of sleep. CSD is widely considered the gold standard sleep diary [21]. The CSD variables were *time taken to fall asleep* (minutes), *total time asleep* (hours), and *sleep efficiency* (proportion of in-bed time spent sleeping). The CSD also assessed *sleep quality* (5-point Likert scale, ranging from 1 [*very poor*] to 5 [*very good*]) and *feeling refreshed* on awakening (5-point Likert scale, ranging from 1 [*not at all rested*] to 5 [*very-well rested*]).

Actigraphy

Participants were asked to wear the MotionWatch 8 actigraphy monitor (CamNtech), a Conformitè Europëenne–marked Class 1 medical device, on their nondominant wrist 24 hours a day for 30 days. MotionWatch 8 was configured to capture limb or bodily movements in 30-second epochs using a triaxial accelerometer. Actigraphy has been shown to provide reliable estimates of sleep compared with polysomnography [22]. CamNtech proprietary software was used to extract the sleep parameters of interest. Running the software requires the time participants get in to bed and out of bed in each 24-hour period to be recorded. In-bed and out-of-bed times were determined either via self-reported times in the CSD or by manual screening of actigraphy data (if CSD data were missing). Where manual screening took place, in-bed times were defined as the time of peak of activity count data immediately before continuous activity ceased for the day, and out-of-bed time was defined as the time of trough of activity count data immediately before continuous activity began for the subsequent day. To assess reliability, 20.1% (54/268) of the data streams were inspected by 2 raters. The actigraph sleep variables were *time taken to fall asleep* (minutes), *total time asleep* (hours), *sleep efficiency* (proportion of in-bed time spent sleeping), and *fragmentation*

index (the number of interruptions of sleep by physical movement with higher scores indicating more fragmented sleep).

Pain, Fatigue, and Mood

Participants were prompted once in the morning at 8 AM and once in the evening at 6 PM to complete the uMotif interface within the study app (Figure S1 in Multimedia Appendix 1) to report the presence and severity of their pain, mood, and fatigue on a 5-point ordinal scale. Pain and fatigue severity were scored from 1 (*none*) to 5 (*very severe*), and mood was scored from 1 (*depressed*) to 5 (*very happy*).

HRQoL Measurements

Participants completed the World Health Organization Quality of Life-Brief (WHOQoL-BREF) scale [23] using the study app at baseline and on days 10, 20, and 30. The recall period was 10 days to capture the changes since the previous assessment. Noncompleters received a reminder text to complete the assessment within 5 days of the original completion date. The WHOQoL-BREF captures an individual's HRQoL across 4 independent domains: physical (7 items), psychological (6 items), social relationships (3 items), and environmental (8 items). Domain items, individually scored from 1 to 5, were summed and transformed into a 0-100 score, with higher scores indicating better HRQoL [23]. For *healthy* people, median (SD) domain scores were physical 76.5 (16.2), psychological 67.8 (15.6), social 70.5 (20.7), and environmental 68.2 (13.8) [24]. To the best of our knowledge, minimal clinically important differences for the WHOQoL-BREF have not been established for RA. Others have reported minimal clinically important differences of approximately 10% in WHOQoL-BREF domain scores [25].

Ethical Approval

Approval was obtained in April 2017 from the National Research Ethics Service Committee North West—Liverpool Central Research Ethics Committee (reference: 17/NW/0217).

Statistical Methods

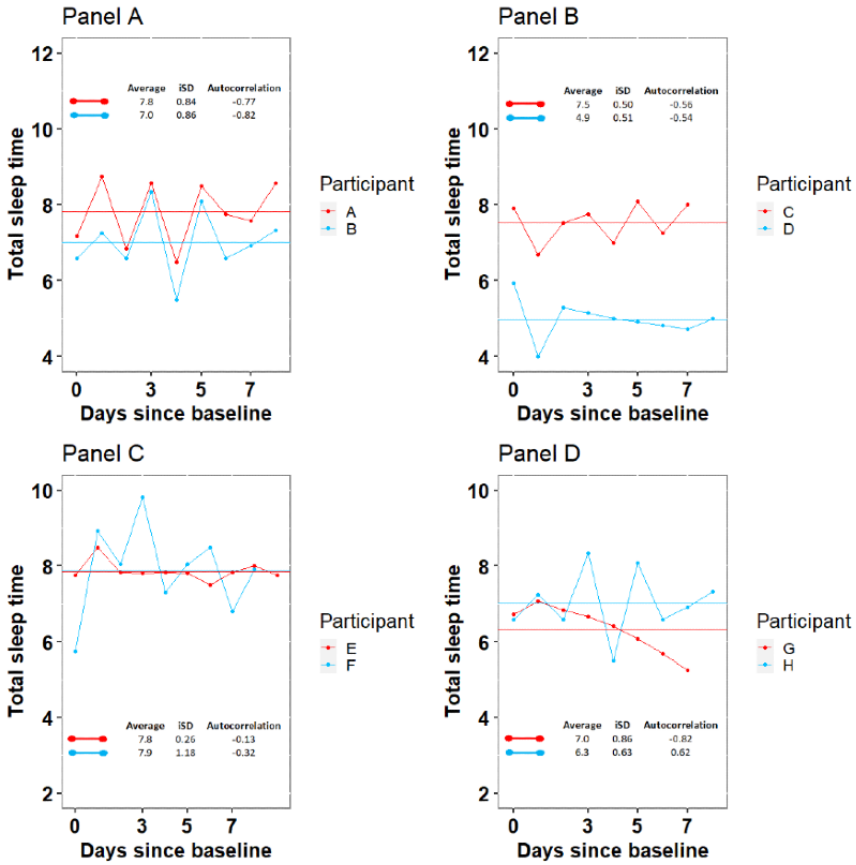
Participants were eligible for this analysis if they provided written consent for their data to be analyzed, completed the baseline questionnaire, returned the actigraph, provided symptom reports on at least 50% of eligible days, completed ≥50% of the CSDs, and completed the WHOQoL-BREF on at least two of the three possible follow-up time points. Descriptive statistics summarizing demographics and baseline measures are presented as frequencies and medians with IQRs.

Data Preparation

Within the study, data item collection frequencies differed: sleep and symptoms were measured daily, whereas HRQoL was measured at baseline and at approximately days 10, 20, and 30. To preserve temporal ordering, we examined the relationship between daily sleep and symptom data from baseline to day 10, from day 11 to 20, and from day 21 to 30, with HRQoL at days 10, 20, and 30.

For the sleep and symptom variables, we calculated the average score and 2 measures of score variability, intraindividual SD (iSD), and autocorrelation [26] (Figure 2). The *average* sleep and symptom scores were calculated as a simple arithmetic mean over each 10-day time window. *iSD* captures the amplitude of sleep or symptom score fluctuations, with higher values indicating higher amplitude and therefore increased between-day variability. *Autocorrelation* (temporal dependency) assesses the extent to which sleep or symptom scores can be predicted based on previous scores. Autocorrelation values ranged from −1 (indicating significant fluctuations around the mean value) to 1 (indicating stable scores at or above or below the mean value).

Figure 2. Examples of individual participants' actigraphy assessed daily total sleep time showing average, intraindividual SD (iSD), and autocorrelation scores over 10 days. In this figure, each panel plots the daily total sleep time for 2 selected participants over 10 days. The 10-day average sleep time is shown as a straight line. In all, 2 measures of variability of total sleep time across the 10-day period were calculated, the iSD and the autocorrelation. The individual panels show the following: (A) shows 2 participants with similar average iSD and autocorrelation scores, (B) shows 2 participants with different average but similar iSD and autocorrelation scores, and (C) shows 2 participants with similar average and autocorrelation scores. The higher iSD score of participant F reflects the higher amplitude of fluctuations in total sleep time when compared with the low amplitude of fluctuation in the total sleep time of participant E. (D) shows 2 participants with similar average scores. The autocorrelation score of participant H toward −1 reflects the fluctuation in total sleep time, whereas the autocorrelation score of participant G toward 1 reflects the day-to-day stability in total sleep time despite a decrease over the period of observation.



Data Analysis

The data for this study were organized at 2 levels. The first level was time (within-person), which was nested within the second level, individuals (between-person). Thus, we used a multilevel data analysis strategy in all analyses. First, univariable multilevel models were fitted to investigate the association between the average sleep scores and each of the 4 WHOQoL-BREF domains. To avoid multicollinearity, separate models were constructed for each sleep variable. To estimate the direct effect of average sleep scores on HRQoL (denoted by the green arrow in Figure 1), multivariable models were fitted to adjust for baseline factors (age, sex, Index of Multiple Deprivation, smoking status, alcohol consumption, marital status, number of medications, BMI, Hospital Anxiety and Depression Scale—anxiety subscale, RAPID-3, PSQI, SCI, OSA, and RLS). The models were then adjusted for sleep variability (iSD and autocorrelation) and consecutively for pain, mood, and fatigue. Finally, all variables were entered into the model. The results are presented as β-coefficients with 95% CIs. The variance in the outcome explained by fixed (marginal R^2) and combined fixed and random (conditional R^2) effects was used to assess model performance.

All analyses were performed using R (version 3.6.0; R Foundation for Statistical Computing).

Results

Study Cohort

A total of 9428 emails were sent to the nonmembers and registered members of the National Rheumatoid Arthritis Society. In total, 285 participants were recruited for the study (Figure S2 in Multimedia Appendix 1). Of 285 participants, 268 (94%) provided baseline data, consent, and returned the actigraph; and 254 (89.1% of recruited participants, 94.8% of eligible participants) were included in the analysis. The 254 participants provided 6731 person-days of CSD data (88.3% of the maximum possible; N=7620) and 7299 person-days of symptom reports (95.8% of the maximum possible; N=7620). The baseline characteristics of the study cohort are presented in Table 1. Sleep problems were common, with a median PSQI score of 11 (IQR 10-12); 32.3% (82/254) of the participants had probable insomnia (SCI score ≤16), 5.9% (15/254) had OSA, and 9.4% (24/254) had RLS.

Table 1. Cohort characteristics.

Characteristics	Participants with baseline data (N=268)	Participants in analysis (n=254)
Sex (female), n (%)	219 (81.7)	206 (81.1)
Age[a] (years), median (IQR)	57 (49-65)	57 (49-64)
Marital status, n (%)		
Single	21 (7.8)	19 (7.5)
Married or with partner	202 (75.4)	194 (76.4)
Separated, widowed, or divorced	44 (16.4)	39 (15.3)
Missing	1 (0.4)	2 (0.8)
Deprivation decile[a] (1=most deprived, 10=least deprived), median (IQR)	7 (4-8)	7 (4-8)
Disease duration[a] (years), median (IQR)	8.8 (3.6-13.9)	8.4 (3.34-13.8)
Baseline disease activity[a] (Routine Assessment of Patient Index Data 3), median (IQR)	14.3 (8.3-19.4)	14.2 (8.3-19.3)
Number of medications[a], median (IQR)	3 (2-4)	3 (2-4)
Possible insomnia (Sleep Condition Indicator; score ≤16), n (%)	83 (31)	82 (32.3)
Pittsburgh Sleep Quality Index[a], median (IQR)	11 (10-12)	11 (10-12)
Sleep apnea (yes), n (%)	16 (6)	15 (5.9)
Restless leg syndrome (yes), n (%)	25 (9.3)	24 (9.4)
Smoking, n (%)		
Current smoker	22 (8.2)	21 (8.3)
Ex-smoker	106 (39.6)	99 (39)
Never smoker	137 (51.1)	130 (51.2)
Missing	3 (1.1)	4 (1.6)
Alcohol[b], n (%)		
None	108 (40.3)	102 (40.2)
Moderate	138 (51.5)	130 (51.2)
Heavy	21 (7.8)	21 (8.3)
Missing	1 (0.4)	1 (0.4)
BMI[c] (kg/m^2), n (%)		
Underweight	5 (1.9)	5 (2)
Healthy	95 (35.5)	88 (34.6)
Overweight	79 (29.5)	73 (28.7)
Obese	81 (30.2)	79 (31.1)
Missing	8 (3)	9 (3.5)
Anxiety[d] (Hospital Anxiety and Depression Scale—anxiety subscale), n (%)		
Not case (0 to <8)	56 (21)	52 (20.5)
Borderline case (8 to <11)	146 (54.5)	138 (54.3)
Case (11 to 21)	65 (24.3)	62 (24.4)
Missing	1 (0.4)	2 (0.8)

[a]Missing values are not shown.

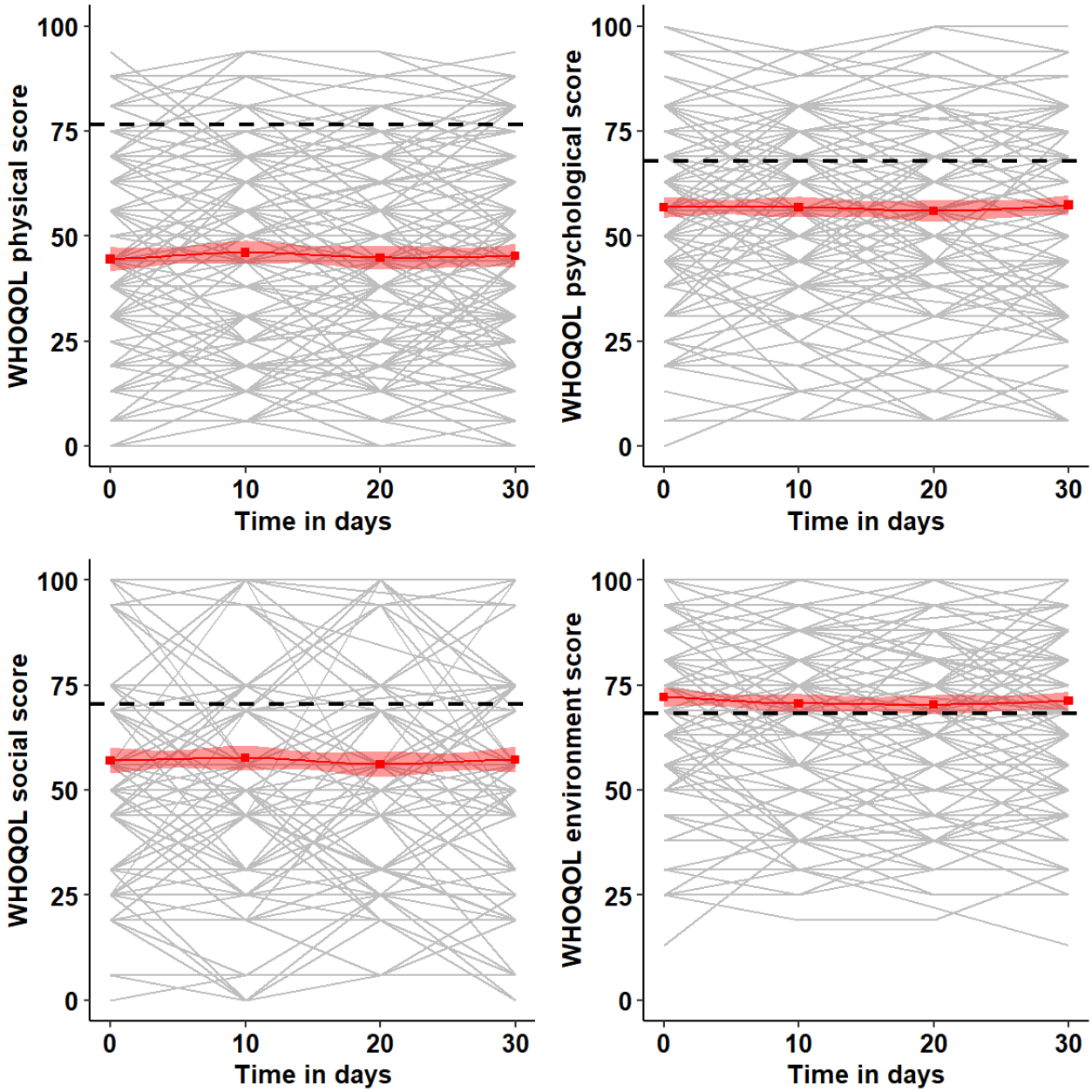
[b]None: 0 units; moderate: 1-15 units; and heavy: ≥16 units.

[c]Underweight: <18.5 kg/m^2; healthy: 18.5-24.9 kg/m^2; overweight: 24.9-30 kg/m^2; or obese: >30 kg/m^2.

[d]Not case: 0-8; borderline case: 8-11; and definite case: 11-21.

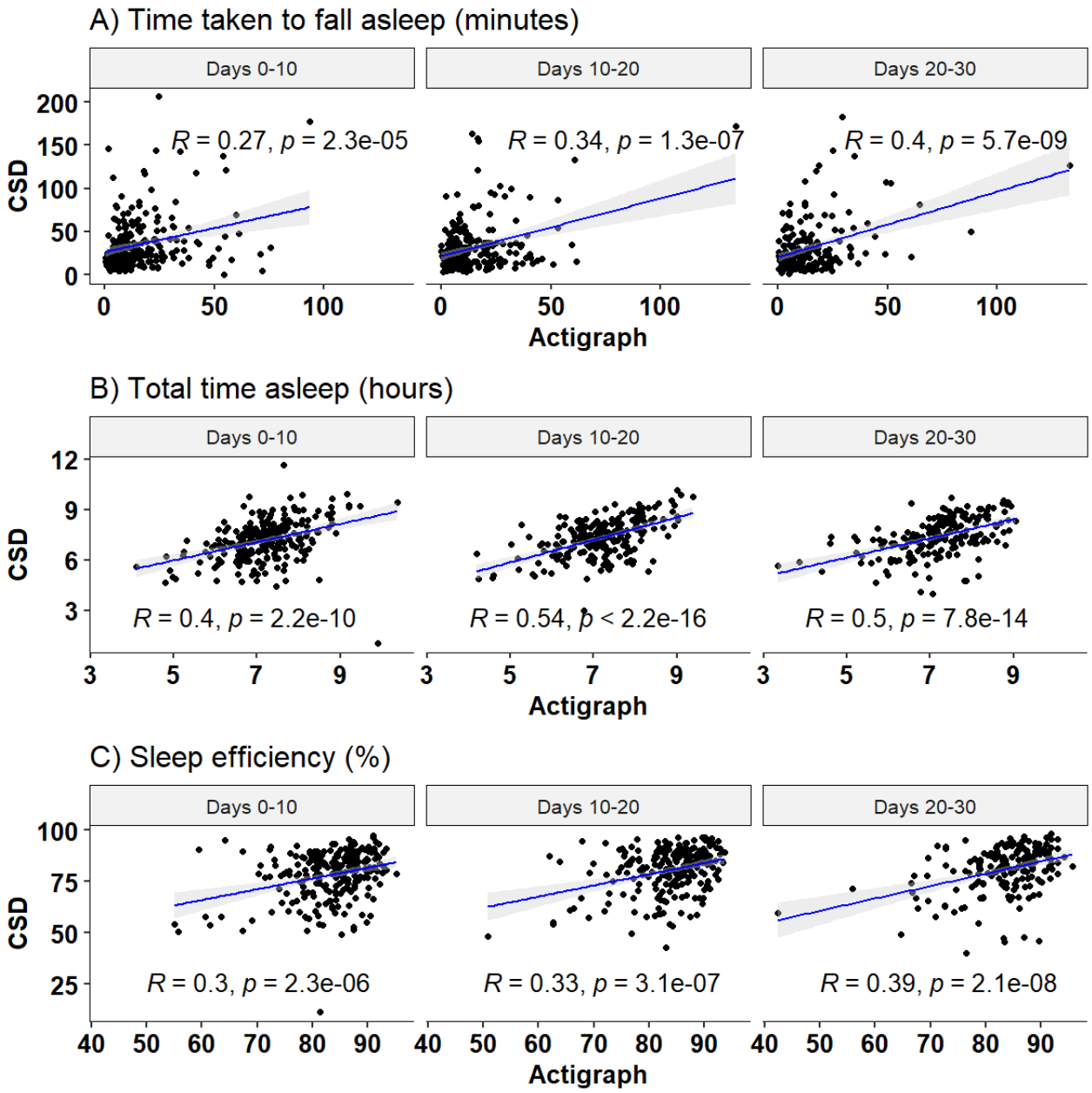
Individual participant's WHOQoL-BREF scores were plotted separately for each of the 4 domains (Figure 3). The cohort mean is shown in red, and the mean scores for healthy individuals are shown in black. At all time points, the mean scores across all participants were lower, that is, poorer, when compared with healthy individuals for the physical, psychological, and social HRQoL domains. The mean score for the environmental domain was similar for the RA cohort and general population. There was substantial variability in domain scores between individuals and change over time within individuals, with; for example, 20.1% (51/254) of the participants having a ≥10% decrease in the physical domain score between consecutive assessments.

Figure 3. Plot of individual participant World Health Organization Quality of Life-Brief (WHOQoL-BREF) domain scores across 30 days. The blue line is the cohort mean score, the black dashed line is the mean score for healthy individuals.



Participants' self-reported sleep patterns on the CSD did not correlate strongly with actigraph data. From baseline to day 10, the median (IQR) time taken to fall asleep was higher on the CSD (median 35.6, IQR 15.4-42.4 minutes) when compared with the actigraph (median 16.1, IQR 6.1-20.6 minutes; Table S1 in Multimedia Appendix 2), and the correlation between these 2 measures was weak (Pearson r=0.27; Figure 4). Similarly, low correlations were found for total time asleep (r=0.4) although the median total time asleep for both measures was 7.2 hours, and sleep efficiency (r=0.3) was lower on the CSD (median 78.0, IQR 71.9-86.9, cf. median 83.3, IQR 80.1-87.7). Participants reported poor sleep quality (median score 2.0, IQR 1.6-2.5) and did not feel refreshed on wakening (median score 1.6, IQR 1.1-2.1). The median actigraphy fragmentation index was 32.3 (IQR 22.0-38.6). These patterns were similar across days 11-20 and 21-30 (Figure 3; Table S1 in Multimedia Appendix 2).

Figure 4. Correlation between objective (actigraph) and subjective (Consensus Sleep Diary [CSD]) measured time taken to fall asleep (A), total time asleep (B), and sleep efficiency (C).



The Relationship Between Sleep and HRQoL and the Effect of Pain, Mood, and Fatigue

CSD Sleep Variables

Overview

The results of the multilevel models to examine the relationship between CSD sleep variables and the WHOQoL-BREF domains are shown in Table 2 and are summarized in the following sections.

Table 2. Consensus Sleep Diary and health-related quality of life.[a]

Sleep parameter and quality of life domain	Univariable β (95% CI)	Multivariable β (95% CI)				
		Baseline factors[b]	Baseline factors plus pain[c]	Baseline factors plus mood[c]	Baseline factors plus fatigue[c]	Baseline factors plus pain, mood, and fatigue[c]
Time taken to fall asleep (minutes)[d]						
Environmental	−0.29 (−0.63 to 0.04)	−0.082 (−0.42 to 0.23)	0.14 (−0.34 to 0.59)	0.13 (−0.34 to 0.58)	0.10 (−0.36 to 0.54)	0.43 (−0.10 to 0.95)
Physical	−0.47 (−0.86 to −0.08)[e]	−0.41 (−0.78 to −0.09)[e]	−0.47 (−0.97 to −0.01)[e]	−0.19 (−0.68 to 0.25)	−0.41 (−0.88 to 0.03)	0.02 (−0.49 to 0.51)
Psychological	−0.71 (−1.07 to −0.35)[e]	−0.55 (−0.91 to −0.22)[e]	−0.47 (−0.99 to 0.01)	−0.22 (−0.72 to 0.24)	−0.25 (−0.73 to 0.22)	0.13 (−0.40 to 0.66)
Social	−0.75 (−1.29 to −0.21)[e]	−0.53 (−1.06 to −0.02)[e]	−0.60 (−1.34 to 0.12)	−0.70 (−1.44 to 0.04)	−0.65 (−1.38 to 0.07)	−0.26 (−1.06 to 0.57)
Total time asleep (hours)[f]						
Environmental	0.49 (−0.41 to 1.40)	0.36 (−0.48 to 1.27)	−0.09 (−1.11 to 1.00)	0.33 (−0.67 to 1.39)	0.03 (−0.92 to 1.02)	0.23 (−0.96 to 1.46)
Physical	1.11 (0.07 to 2.15)[e]	0.93 (0.04 to 1.82)[e]	0.39 (−0.64 to 1.42)	0.47 (−0.52 to 1.47)	0.45 (−0.50 to 1.41)	0.14 (−0.94 to 1.22)
Psychological	0.45 (−0.52 to 1.43)	0.41 (−0.49 to 1.35)	−0.07 (−1.16 to 1.08)	0.28 (−0.74 to 1.36)	0.01 (−0.99 to 1.04)	0.30 (−0.84 to 1.52)
Social	1.65 (0.21 to 3.10)[e]	1.50 (0.18 to 2.93)[e]	0.012 (−1.55 to 1.68)	0.88 (−0.68 to 2.57)	1.001 (−0.49 to 2.57)	−0.03 (−1.77 to 1.84)
Sleep efficiency (%)[g]						
Environmental	1.10 (0.14 to 2.06)[e]	0.48 (−0.426 to 1.45)	−0.42 (−1.73 to 0.99)	−0.12 (−1.41 to 1.24)	−0.05 (−1.29 to 1.25)	−0.74 (−2.25 to 0.82)
Physical	2.92 (1.81 to 4.05)[e]	2.03 (1.10 to 3.02)[e]	1.22 (−0.08 to 2.60)	0.98 (−0.28 to 2.32)	1.42 (0.19 to 2.71)[e]	0.01 (−1.36 to 1.43)
Psychological	2.44 (1.41 to 3.49)[e]	1.80 (0.85 to 2.82)[e]	2.11 (0.71 to 3.61)[e]	1.57 (0.24 to 2.98)[e]	1.67 (0.38 to 3.04)[e]	1.28 (−0.20 to 2.84)
Social	2.97 (1.44 to 4.50)[e]	2.22 (0.78 to 3.76)[e]	1.53 (−0.50 to 3.72)	1.92 (−0.12 to 4.09)	2.02 (0.05 to 4.10)[e]	0.823 (−1.45 to 3.20)
Sleep quality (1-5)						
Environmental	1.65 (0.20 to 3.12)[e]	−0.05 (−1.46 to 1.47)	0.31 (−1.46 to 2.23)	0.03 (−1.82 to 2.01)	0.12 (−1.59 to 1.90)	0.10 (−2.21 to 2.54)
Physical	5.99 (4.34 to 7.69)[e]	4.58 (3.12 to 6.11)[e]	4.06 (2.29 to 5.91)[e]	3.97 (2.16 to 5.81)[e]	3.83 (2.11 to 5.61)[e]	3.13 (1.03 to 5.31)[e]
Psychological	2.47 (0.89 to 4.10)[e]	0.95 (−0.55 to 2.58)	1.06 (−0.84 to 3.17)	1.10 (−0.78 to 3.20)	1.12 (−0.68 to 3.09)	1.54 (−0.71 to 4.01)
Social	2.82 (0.44 to 5.24)[e]	0.04 (−2.25 to 2.56)	−0.25 (−2.97 to 2.69)	−0.22 (−3.15 to 3.00)	0.21 (−2.55 to 3.15)	−1.03 (−4.46 to 2.68)
Feeling refreshed (1-5)						
Environmental	2.12 (0.63 to 3.65)[e]	0.49 (−0.90 to 2.07)	0.22 (−1.48 to 2.15)	−0.05 (−1.80 to 1.88)	−0.38 (−2.09 to 1.44)	−0.85 (−3.07 to 1.56)

Sleep parameter and quality of life domain	Univariable β (95% CI)	Multivariable β (95% CI)				
		Baseline factors[b]	Baseline factors plus pain[c]	Baseline factors plus mood[c]	Baseline factors plus fatigue[c]	Baseline factors plus pain, mood, and fatigue[c]
Physical	6.28 (4.57 to 8.04)[e]	4.57 (3.14 to 6.10)[e]	3.99 (2.31 to 5.76)[e]	4.31 (2.59 to 6.06)[e]	3.44 (1.72 to 5.22)[e]	3.60 (1.58 to 5.65)[e]
Psychological	3.73 (2.07 to 5.44)[e]	2.43 (0.98 to 4.15)[e]	2.25 (0.47 to 4.35)[e]	2.13 (0.37 to 4.16)[e]	1.94 (0.18 to 3.93)[e]	1.87 (−0.32 to 4.34)
Social	2.97 (0.5 to 5.42)[e]	0.74 (−1.49 to 3.31)	−0.62 (−3.23 to 2.39)	−0.49 (−3.26 to 2.60)	−0.30 (−3.05 to 2.71)	−1.42 (−4.71 to 2.29)

[a]The relationship between average scores of Consensus Sleep Diary and health-related quality of life.

[b]Age, sex, Index of Multiple Deprivation, smoking status, alcohol consumption, marital status, number of medications, BMI (self-reported kg/m^2), Hospital Anxiety and Depression Scale—anxiety subscale, Routine Assessment of Patient Index Data 3, obstructive sleep apnea, and restless leg syndrome.

[c]Similar to footnote *b*, plus intraindividual SD and autocorrelation measures of sleep parameters, pain, mood, and fatigue.

[d]For each 10-minute increase in time taken to fall asleep.

[e]The results excluding zero.

[f]For each 1-hour asleep.

[g]For each 10% increase in sleep efficiency.

Time Taken to Fall Asleep

In the unadjusted models, an increase in the time taken to fall asleep was associated with lower, that is, poorer WHOQoL-BREF scores. For each 10-minute increase in the time taken to fall asleep, physical domain scores decreased by 0.47 points (β=−.47, 95% CI −0.86 to −0.08), psychological domain scores decreased by 0.71 points (95% CI −1.07 to −0.35), and social domain scores decreased by 0.75 points (95% CI −1.29 to −0.21). These associations persisted after adjusting for baseline factors (age, sex, deprivation, smoking status, alcohol consumption, marital status, number of medications used, BMI, and anxiety), baseline RAPID-3 scores, baseline PSQI and SCI scores, OSA, and RLS. When pain was included in the model, the relationship with physical domain scores was attenuated but persisted (β=−.05, 95% CI −0.10 to −0.001), whereas the relationship with psychological and social domain scores did not persist. When pain, mood, and fatigue were included in the final model, the time taken to fall asleep was not significantly associated with any of the WHOQoL-BREF domains (Table 2). The variability in time taken to fall asleep was not associated with the WHOQoL-BREF domains (Table S2 in Multimedia Appendix 2).

Total Time Asleep

An increase in the total time asleep was associated with higher, that is, better, physical (for each 1-hour increase: β=1.11, 95% CI 0.07-2.15) and social domain (β=1.65, 95% CI 0.21-3.10) scores. These associations were independent of baseline factors but were attenuated and no longer significant when pain, fatigue, and mood were included in the models. Variability in total time asleep was an important predictor of WHOQoL-BREF domains (Table S2 in Multimedia Appendix 2). An increase in the iSD score, indicating increased variability in the total sleep time, was associated with poorer physical (for each unit increase: β=−2.41, 95% CI −4.14 to −0.76) and psychological (β=−2.21, 95% CI −3.96 to −0.38) domain scores. These associations were not explained by the inclusion of baseline factors, such as pain, fatigue, and mood, into the models.

Sleep Efficiency

After adjusting for baseline factors, increased sleep efficiency was associated with better physical (for each 10% increase in sleep efficiency: β=2.03, 95% CI 1.10-3.02), psychological (β=1.80, 95% CI 0.85-2.82), and social domain scores (β=2.22, 95% CI 0.78-3.76). However, when pain, fatigue, and mood were included in the models, there were no significant associations between the sleep efficiency and HRQoL domains. The variability in sleep efficiency was not associated with the WHOQoL-BREF domains (Table S2 in Multimedia Appendix 2).

Sleep Quality

Sleep quality was associated with all 4 WHOQoL-BREF domains (Table 2). Although most associations were explained by baseline factors, the association between sleep quality and the physical domain scores persisted, with a unit increase in the sleep quality score being associated with a 4.58 (95% CI 3.12-6.11) increase in the physical domain score. This association was not explained by pain, mood, or fatigue. The variability in sleep quality was not associated with the WHOQoL-BREF domains (Table S2 in Multimedia Appendix 2).

Feeling Refreshed

A unit increase in the feeling refreshed score was associated with better physical (β=4.57, 95% CI 3.14-6.10) and psychological (β=2.43, 95% CI 0.98-4.15) HRQoL independently of baseline factors, including disease severity. The relationship with physical HRQoL persisted when pain, mood, and fatigue were included in the model. Variability in

feeling refreshed was not associated with WHOQoL-BREF domains (Table S2 in Multimedia Appendix 2).

Actigraphy Sleep Variables

Of the actigraph sleep parameters, only total sleep time was associated with physical domain scores and appeared in the opposite direction (β=−1.24, 95% CI −2.58 to −0.09; Table S3 in Multimedia Appendix 2). However, this association was attenuated and not significant when adjusted for pain, mood, and fatigue.

Model Performance

The marginal R^2 values of the final multivariable models ranged from 33% to 69% (Multimedia Appendix 2).

Discussion

Principal Findings

In this study, we report the findings of a prospective mHealth study that examined the role of sleep disturbance in people with RA and its impact on HRQoL. At all assessment points, the average HRQoL of the cohort was lower than the population average [24]. We observed consistent patterns of association between sleep and HRQoL. First, the CSD sleep variables predicted physical, psychological, and social HRQoL: increases in the time taken to fall asleep predicted poorer HRQoL, whereas increases in total time asleep, sleep efficiency, feeling refreshed, and sleep quality predicted better HRQoL. Increased variability in total sleep time was associated with poor physical and psychological quality of life. Second, most of these associations were independent of sociodemographic and lifestyle factors; disease activity; baseline medication use; levels of anxiety; sleep quality (PSQI score); and clinical sleep disorders, including insomnia, OSA, and RLS. Third, these data clearly show that the relationship between sleep variables and HRQoL (with the exception of feeling refreshed, sleep quality with physical HRQoL, and the relationships with variability in total sleep time) were mediated via changes in pain, mood, and fatigue. Finally, there was no consistent pattern of association between actigraphy-derived sleep variables and HRQoL.

Limitations

There are several limitations to consider when interpreting these results. First, our study design may have introduced a selection bias. For example, older age and higher disease severity have been shown to predict nonparticipation in digital health research [27]. Bias may have been introduced if nonparticipation was related to sleep disturbance and, independently of sleep, to HRQoL. However, our data indicate that this was unlikely to be the case because the rates of sleep disturbance were within the expected range. The median PSQI score was 11, which is comparable with that in other studies of people with RA [28]. The prevalence of poor sleep (PSQI score ≥5) was 90% (data not shown), which is in line with estimates from previous studies. Finally, one-third of our sample was classified as having probable insomnia, which is comparable with other estimates of patients with *arthritis* [29]. Our screening for OSA and RLS was limited, and their impact on RA-HRQoL was unclear in our study. Engagement in the study was high (241/254, 94.8%) among those participants who were recruited and successfully commenced data collection, and a few (14/268, 5.5%) were lost to follow-up after enrolling in the study. Therefore, it is unlikely that the loss to follow-up bias had a substantial influence on our results. Finally, we observed a poor correlation between the subjective (CSD) and objective (actigraphy) measures of sleep. The low correlation between subjective and objective measures is a common observation. They appear to measure different dimensions of sleep: the correlation between polysomnography measures of sleep and actigraphy is stronger than that between polysomnography and sleep diaries [30]. There may be different underlying biological (eg, inflammatory) and psychological mechanisms between subjective and objective measures, or self-reported sleep may reflect the reporting of a more chronic sleep problem, whereas actigraphy assesses acute sleep changes [30].

Implications of This Study

The data reported here support our hypothesis that sleep disturbances predict poor HRQoL in people with RA. Sleep efficiency was low compared with healthy people [31], and people with RA spent substantial periods in bed but not asleep. Optimizing total sleep time, increasing sleep efficiency, decreasing sleep onset latency, and reducing variability in total sleep time could improve HRQoL in people with RA. We also reported that pain, mood, and fatigue mediate these relationships. There is no high-quality evidence for the effectiveness of pharmacological [32] or nonpharmacological [33] sleep interventions in people with RA. The hybrid cognitive behavioral therapy (CBT) proposed by Tang et al [34] for people with chronic pain incorporates components of CBT for insomnia and CBT for pain and has been shown to improve sleep, pain interference, fatigue, and depression. Our data suggest that hybrid treatment models that simultaneously address sleep disturbances and associated symptoms, including pain, mood, and fatigue, may improve HRQoL in people with RA.

Conclusions

Sleep predicts poor HRQoL independent of disease severity. Sleep disturbance indirectly impacts poor HRQoL via its effects on pain, mood, and fatigue. These data should inform the development of complex interventions to improve sleep-related HRQoL in people with RA.

Acknowledgments

The authors would like to thank the National Rheumatoid Arthritis Society; National Health Service (NHS) Clinical Research Networks of Greater Manchester, West Midlands, North West Coast, Yorkshire, and Humber; East Cheshire NHS Trust; North Cumbria University Hospitals NHS Trust; Kent Community Health NHS Foundation Trust; Solent NHS Trust; and Gateshead Health NHS Foundation Trust for their support with participant recruitment. The quality of life, sleep, and rheumatoid arthritis

study was funded by Arthritis Research, United Kingdom (grant reference 21188). This work was also supported by infrastructure support from the Centre for Epidemiology versus Arthritis (grant reference 21755, formerly Arthritis Research UK Centre for Epidemiology [grant reference 20380]). The organization had no further input into how the study was executed or in the analysis or interpretation of data.

Conflicts of Interest

WGD has received consultancy fees from Google and Abbvie (unrelated to this work).

Multimedia Appendix 1
A figure of the main graphical interface used in the study and the flow chart of participants recruited and included in the study.
[DOCX File , 119 KB - jmir_v24i4e32825_app1.docx]

Multimedia Appendix 2
Tables summarizing the Consensus Sleep Diary and Actigraph derived sleep parameters, and the relationships between sleep parameters and health-related quality of life.
[DOCX File , 39 KB - jmir_v24i4e32825_app2.docx]

References

1. Dominick KL, Ahern FM, Gold CH, Heller DA. Health-related quality of life among older adults with arthritis. Health Qual Life Outcomes 2004 Jan 13;2:5 [FREE Full text] [doi: 10.1186/1477-7525-2-5] [Medline: 14720300]
2. Geryk LL, Carpenter DM, Blalock SJ, DeVellis RF, Jordan JM. The impact of co-morbidity on health-related quality of life in rheumatoid arthritis and osteoarthritis patients. Clin Exp Rheumatol 2015;33(3):366-374. [Medline: 25898121]
3. Zhang L, Shen B, Liu S. Rheumatoid arthritis is associated with negatively variable impacts on domains of sleep disturbances: evidence from a systematic review and meta-analysis. Psychol Health Med 2021 Mar;26(3):267-277. [doi: 10.1080/13548506.2020.1764597] [Medline: 32420757]
4. Drewes AM, Svendsen L, Taagholt SJ, Bjerregård K, Nielsen KD, Hansen B. Sleep in rheumatoid arthritis: a comparison with healthy subjects and studies of sleep/wake interactions. Br J Rheumatol 1998 Jan;37(1):71-81. [doi: 10.1093/rheumatology/37.1.71] [Medline: 9487254]
5. Roehrs T, Diederichs C, Gillis M, Burger AJ, Stout RA, Lumley MA, et al. Nocturnal sleep, daytime sleepiness and fatigue in fibromyalgia patients compared to rheumatoid arthritis patients and healthy controls: a preliminary study. Sleep Med 2013 Jan;14(1):109-115. [doi: 10.1016/j.sleep.2012.09.020] [Medline: 23149217]
6. Crosby LJ. Factors which contribute to fatigue associated with rheumatoid arthritis. J Adv Nurs 1991 Aug;16(8):974-981. [doi: 10.1111/j.1365-2648.1991.tb01803.x] [Medline: 1779087]
7. Drewes AM, Nielsen KD, Hansen B, Taagholt SJ, Bjerregård K, Svendsen L. A longitudinal study of clinical symptoms and sleep parameters in rheumatoid arthritis. Rheumatology (Oxford) 2000 Nov;39(11):1287-1289. [doi: 10.1093/rheumatology/39.11.1287] [Medline: 11085813]
8. Clarke LL, Wilson S, Kirwan JR. Using actigraphy to measure sleep patterns in rheumatoid arthritis: a pilot study in patients taking night-time prednisone. Musculoskeletal Care 2013 Sep;11(3):179-185 [FREE Full text] [doi: 10.1002/msc.1052] [Medline: 23695990]
9. Tang NK, Goodchild CE, Sanborn AN, Howard J, Salkovskis PM. Deciphering the temporal link between pain and sleep in a heterogeneous chronic pain patient sample: a multilevel daily process study. Sleep 2012 May 01;35(5):675-687 [FREE Full text] [doi: 10.5665/sleep.1830] [Medline: 22547894]
10. Tafti M. Genetic aspects of normal and disturbed sleep. Sleep Med 2009 Sep;10 Suppl 1:S17-S21. [doi: 10.1016/j.sleep.2009.07.002] [Medline: 19660984]
11. Buysse DJ. Sleep health: can we define it? Does it matter? Sleep 2014 Jan 01;37(1):9-17 [FREE Full text] [doi: 10.5665/sleep.3298] [Medline: 24470692]
12. Akerstedt T, Axelsson J, Lekander M, Orsini N, Kecklund G. The daily variation in sleepiness and its relation to the preceding sleep episode--a prospective study across 42 days of normal living. J Sleep Res 2013 Jun;22(3):258-265 [FREE Full text] [doi: 10.1111/jsr.12014] [Medline: 23205895]
13. Tang NK, McBeth J, Jordan KP, Blagojevic-Bucknall M, Croft P, Wilkie R. Impact of musculoskeletal pain on insomnia onset: a prospective cohort study. Rheumatology (Oxford) 2015 Feb;54(2):248-256 [FREE Full text] [doi: 10.1093/rheumatology/keu283] [Medline: 25125589]
14. Druce KL, Cordingley L, Short V, Moore S, Hellman B, James B, et al. Quality of life, sleep and rheumatoid arthritis (QUASAR): a protocol for a prospective UK mHealth study to investigate the relationship between sleep and quality of life in adults with rheumatoid arthritis. BMJ Open 2018 Jan 26;8(1):e018752 [FREE Full text] [doi: 10.1136/bmjopen-2017-018752] [Medline: 29374666]

15. Muller S, Hider SL, Raza K, Stack RJ, Hayward RA, Mallen CD. An algorithm to identify rheumatoid arthritis in primary care: a Clinical Practice Research Datalink study. BMJ Open 2015 Dec 23;5(12):e009309 [FREE Full text] [doi: 10.1136/bmjopen-2015-009309] [Medline: 26700281]
16. Mollayeva T, Thurairajah P, Burton K, Mollayeva S, Shapiro CM, Colantonio A. The Pittsburgh sleep quality index as a screening tool for sleep dysfunction in clinical and non-clinical samples: a systematic review and meta-analysis. Sleep Med Rev 2016 Feb;25:52-73. [doi: 10.1016/j.smrv.2015.01.009] [Medline: 26163057]
17. Espie CA, Kyle SD, Hames P, Gardani M, Fleming L, Cape J. The Sleep Condition Indicator: a clinical screening tool to evaluate insomnia disorder. BMJ Open 2014 Mar 18;4(3):e004183 [FREE Full text] [doi: 10.1136/bmjopen-2013-004183] [Medline: 24643168]
18. 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]
19. Anderson J, Caplan L, Yazdany J, Robbins ML, Neogi T, Michaud K, et al. Rheumatoid arthritis disease activity measures: American College of Rheumatology recommendations for use in clinical practice. Arthritis Care Res (Hoboken) 2012 May;64(5):640-647 [FREE Full text] [doi: 10.1002/acr.21649] [Medline: 22473918]
20. Pincus T, Yazici Y, Bergman MJ. RAPID3, an index to assess and monitor patients with rheumatoid arthritis, without formal joint counts: similar results to DAS28 and CDAI in clinical trials and clinical care. Rheum Dis Clin North Am 2009 Nov;35(4):773-viii. [doi: 10.1016/j.rdc.2009.10.008] [Medline: 19962621]
21. Carney CE, Buysse DJ, Ancoli-Israel S, Edinger JD, Krystal AD, Lichstein KL, et al. The consensus sleep diary: standardizing prospective sleep self-monitoring. Sleep 2012 Feb 01;35(2):287-302 [FREE Full text] [doi: 10.5665/sleep.1642] [Medline: 22294820]
22. Elbaz M, Yauy K, Metlaine A, Martoni M, Leger D. Validation of a new ActiGraph motion watch versus polysomnography on 70 healthy and suspected sleep-disordered subjects. J Sleep Res 2012;21(Suppl 1):218.
23. Programme on mental health: WHOQOL user manual. World Health Organization. 2012. URL: https://apps.who.int/iris/handle/10665/77932 [accessed 2019-01-01]
24. Skevington SM, McCrate FM. Expecting a good quality of life in health: assessing people with diverse diseases and conditions using the WHOQOL-BREF. Health Expect 2012 Mar;15(1):49-62 [FREE Full text] [doi: 10.1111/j.1369-7625.2010.00650.x] [Medline: 21281412]
25. de Mol M, Visser S, Aerts JG, Lodder P, de Vries J, den Oudsten BL. Satisfactory results of a psychometric analysis and calculation of minimal clinically important differences of the World Health Organization quality of life-BREF questionnaire in an observational cohort study with lung cancer and mesothelioma patients. BMC Cancer 2018 Nov 26;18(1):1173 [FREE Full text] [doi: 10.1186/s12885-018-4793-8] [Medline: 30477456]
26. Mun CJ, Suk HW, Davis MC, Karoly P, Finan P, Tennen H, et al. Investigating intraindividual pain variability: methods, applications, issues, and directions. Pain 2019 Nov;160(11):2415-2429. [doi: 10.1097/j.pain.0000000000001626] [Medline: 31145212]
27. Poli A, Kelfve S, Klompstra L, Strömberg A, Jaarsma T, Motel-Klingebiel A. Prediction of (Non)participation of older people in digital health research: exergame intervention study. J Med Internet Res 2020 Jun 05;22(6):e17884 [FREE Full text] [doi: 10.2196/17884] [Medline: 32501275]
28. Hughes M, Chalk A, Sharma P, Dahiya S, Galloway J. A cross-sectional study of sleep and depression in a rheumatoid arthritis population. Clin Rheumatol 2021 Apr;40(4):1299-1305 [FREE Full text] [doi: 10.1007/s10067-020-05414-8] [Medline: 32968925]
29. Budhiraja R, Roth T, Hudgel DW, Budhiraja P, Drake CL. Prevalence and polysomnographic correlates of insomnia comorbid with medical disorders. Sleep 2011 Jul 01;34(7):859-867 [FREE Full text] [doi: 10.5665/SLEEP.1114] [Medline: 21731135]
30. Jackowska M, Dockray S, Hendrickx H, Steptoe A. Psychosocial factors and sleep efficiency: discrepancies between subjective and objective evaluations of sleep. Psychosom Med 2011;73(9):810-816. [doi: 10.1097/PSY.0b013e3182359e77] [Medline: 22021463]
31. Tonetti L, Mingozzi R, Natale V. Comparison between paper and electronic sleep diary. Biol Rhythm Res 2016 Jun 10;47(5):743-753. [doi: 10.1080/09291016.2016.1191689]
32. Reynolds AC, Marshall NS, Hill CL, Adams RJ. Systematic review of the efficacy of commonly prescribed pharmacological treatments for primary treatment of sleep disturbance in patients with diagnosed autoimmune disease. Sleep Med Rev 2020 Feb;49:101232. [doi: 10.1016/j.smrv.2019.101232] [Medline: 31911367]
33. Santos EJ, Duarte C, Marques A, Cardoso D, Apóstolo J, da Silva JA, et al. Effectiveness of non-pharmacological and non-surgical interventions for rheumatoid arthritis: an umbrella review. JBI Database System Rev Implement Rep 2019 Jul;17(7):1494-1531. [doi: 10.11124/JBISRIR-D-18-00020] [Medline: 31169776]
34. Tang NK, Goodchild CE, Salkovskis PM. Hybrid cognitive-behaviour therapy for individuals with insomnia and chronic pain: a pilot randomised controlled trial. Behav Res Ther 2012 Dec;50(12):814-821. [doi: 10.1016/j.brat.2012.08.006] [Medline: 23123531]

Abbreviations

CBT: cognitive behavioral therapy
CSD: Consensus Sleep Diary
HRQoL: health-related quality of life
iSD: intraindividual SD
mHealth: mobile health
NHS: National Health Service
OSA: obstructive sleep apnea
PSQI: Pittsburgh Sleep Quality Index
QUASAR: quality of life, sleep, and rheumatoid arthritis
RA: rheumatoid arthritis
RAPID-3: Routine Assessment of Patient Index Data 3
RLS: restless leg syndrome
SCI: Sleep Condition Indicator
WHOQoL-BREF: World Health Organization Quality of Life-Brief

Edited by R Kukafka; submitted 11.08.21; peer-reviewed by A Tannoubi, D Gunasekeran, Z Liang; comments to author 08.11.21; revised version received 11.11.21; accepted 27.12.21; published 22.04.22.

Please cite as:
McBeth J, Dixon WG, Moore SM, Hellman B, James B, Kyle SD, Lunt M, Cordingley L, Yimer BB, Druce KL
Sleep Disturbance and Quality of Life in Rheumatoid Arthritis: Prospective mHealth Study
J Med Internet Res 2022;24(4):e32825
URL: https://www.jmir.org/2022/4/e32825
doi:10.2196/32825
PMID:35451978

Original Paper

Stigmatizing Attitudes Across Cybersuicides and Offline Suicides: Content Analysis of Sina Weibo

Ang Li[1,2], PhD; Dongdong Jiao[3], MSc; Tingshao Zhu[2,4], PhD

[1]Department of Psychology, Beijing Forestry University, Beijing, China

[2]Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China

[3]National Computer System Engineering Research Institute of China, Beijing, China

[4]Department of Psychology, University of Chinese Academy of Sciences, Beijing, China

Corresponding Author:
Tingshao Zhu, PhD
Key Laboratory of Behavioral Science
Institute of Psychology
Chinese Academy of Sciences
Hexie Building
No 16 Lincui Rd
Beijing, 100101
China
Phone: 86 150 1096 5509
Email: tszhu@psych.ac.cn

Abstract

Background: The new reality of cybersuicide raises challenges to ideologies about the traditional form of suicide that does not involve the internet (offline suicide), which may lead to changes in audience's attitudes. However, knowledge on whether stigmatizing attitudes differ between cybersuicides and offline suicides remains limited.

Objective: This study aims to consider livestreamed suicide as a typical representative of cybersuicide and use social media data (Sina Weibo) to investigate the differences in stigmatizing attitudes across cybersuicides and offline suicides in terms of attitude types and linguistic characteristics.

Methods: A total of 4393 cybersuicide-related and 2843 offline suicide-related Weibo posts were collected and analyzed. First, human coders were recruited and trained to perform a content analysis on the collected posts to determine whether each of them reflected stigma. Second, a text analysis tool was used to automatically extract a number of psycholinguistic features from each post. Subsequently, based on the selected features, a series of classification models were constructed for different purposes: differentiating the general stigma of cybersuicide from that of offline suicide and differentiating the negative stereotypes of cybersuicide from that of offline suicide.

Results: In terms of attitude types, cybersuicide was observed to carry more stigma than offline suicide (χ^2_1=179.8; *P*<.001). Between cybersuicides and offline suicides, there were significant differences in the proportion of posts associated with five different negative stereotypes, including *stupid and shallow* (χ^2_1=28.9; *P*<.001), *false representation* (χ^2_1=144.4; *P*<.001), *weak and pathetic* (χ^2_1=20.4; *P*<.001), *glorified and normalized* (χ^2_1=177.6; *P*<.001), and *immoral* (χ^2_1=11.8; *P*=.001). Similar results were also found for different genders and regions. In terms of linguistic characteristics, the *F*-measure values of the classification models ranged from 0.81 to 0.85.

Conclusions: The way people perceive cybersuicide differs from how they perceive offline suicide. The results of this study have implications for reducing the stigma against suicide.

(J Med Internet Res 2022;24(4):e36489) doi:10.2196/36489

KEYWORDS

stigma; cybersuicide; livestreamed suicide; linguistic analysis; social media

Introduction

Background

Suicide remains one of the leading causes of death worldwide according to the latest estimates released by the World Health Organization [1]. Mental illness stigma refers to the negative stereotyping of people with mental illnesses that may lead to discrimination [2]. Although there is often a stigma associated with all mental illnesses, suicide can be especially stigmatized, dismissed as "merely attention-seeking gesturers" [3]. Stigma against suicide can lead to a reduced likelihood of receiving treatment and an increased risk of death by suicide [4,5]. Therefore, lowering stigma can improve mental health outcomes. Stigma reduction efforts are more likely to be effective when targeting specific mental health problems [6-8]. Therefore, it is crucial to understand the negative stereotypes people often associate with suicide and then design antistigma campaigns accordingly.

The internet facilitates self-disclosure and social connection, giving rise to an emerging form of suicide (ie, cybersuicide). Unlike the traditional form of suicide that does not involve the internet (ie, offline suicide), cybersuicide covers a broad range of internet-mediated suicidal behaviors and phenomena, including livestreamed suicide [9,10]. The internet increases people's willingness to disclose more about themselves and offers highly interactive platforms for interpersonal communication (eg, Twitter and Sina Weibo). Therefore, on the internet, people with suicidal intentions are motivated to share and communicate their thoughts on suicide with others and are able to maintain active contact with their social network in the last moments of life, although members of social networks do not live close to them. In livestreamed suicide, for example, through various media (text messages, pictures, videos, and voice notes), the internet enables people with suicidal intentions to broadcast their suicides to their entire social network before death and allows real time interaction between people with suicidal intentions and their audience. This means that cybersuicide makes a very personal and private act highly public and greatly enhances the interaction between people with suicidal intentions and their audience, which raises challenges to long-held ideas and beliefs about offline suicide [11-13] and facilitates the creation of emerging cultures around suicide [9,10,14,15]. It is well known that social and cultural factors can influence attitudes toward mental health problems, including suicide [16-18]. Therefore, it is not surprising that cybersuicide, which has radically transformed the sociocultural context of suicide, may lead to changes in audience's attitudes, suggesting the need for further examination of the differences in stigmatizing attitudes across cybersuicides and offline suicides.

The livestreamed suicide is commonly considered as one of the most notable and representative types of cybersuicide, particularly in China [9,19]. From 2003 to 2016, at least 193 livestreamed suicide incidents occurred in China [20]. Therefore, considerable research efforts have been directed toward understanding the specific types of attitudes associated with livestreamed suicide. As social media allow users to freely disclose their feelings and thoughts and include vast quantities of publicly available data, many studies used human coders to analyze relevant social media data (eg, posts with relevant keywords and audience-generated messages related to relevant suicide incidents) and concluded that the public may react strongly against livestreamed suicide [21-24]. Besides, a study found a distinctive type of negative stereotype associated with livestreamed suicide (ie, false representation stigma, a misleading belief that people livestreaming their suicides do not really want to kill themselves) [25], which was not mentioned in previous studies regarding offline suicide. The results of these studies imply that the public may react differently to cybersuicides and offline suicides. However, to our knowledge, no research has been conducted to directly investigate the differences in stigmatizing attitudes toward cybersuicides and offline suicides. Furthermore, attitudes can manifest not only in specific attitude types (what it says exactly) but also in linguistic characteristics of language expressions (how it is said) [26-29]. Recent studies confirmed that health-related stigmatizing attitudes (eg, Alzheimer disease, depression, and livestreamed suicide) can be identified by analyzing linguistic characteristics of expressions in social media posts [30-32]. More importantly, the differences in stigmatizing attitudes across mental health problems can be reflected in different patterns of language use as well [33]. Therefore, apart from the differences in types of stigma, the differences in the linguistic characteristics of stigmatizing expressions also need to be investigated.

Objective

To address these concerns by analyzing social media data (Sina Weibo, a Chinese social media site that is similar to Twitter), this study attempts to directly and systematically investigate the differences in stigmatizing attitudes across cybersuicides and offline suicides in terms of attitude types and linguistic characteristics, respectively.

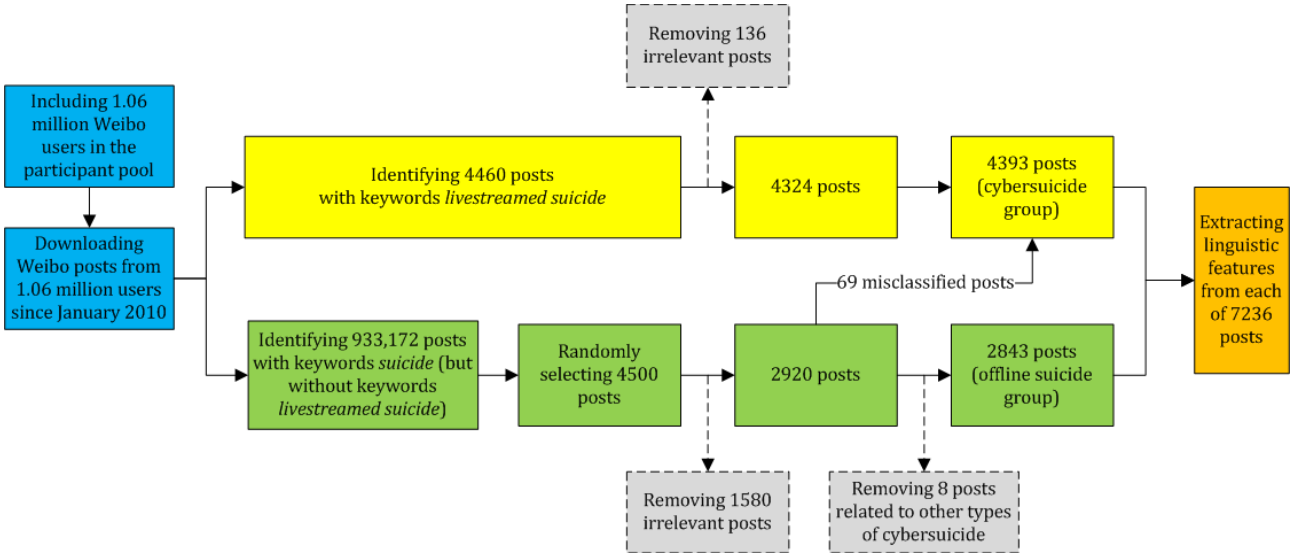
It is worth noting that cybersuicide is a new and developing form of suicide. The public may not be equally familiar with different types of cybersuicide. In China, because of the prevalence and media coverage of livestreamed suicide, compared with other types of cybersuicide, the public is expected to be more familiar with and more likely to discuss livestreamed suicide on social media. Therefore, to collect sufficient social media data for further analysis, this study aims to consider livestreamed suicide as a typical representative of cybersuicide and compare it with offline suicide.

Methods

Research Process

The research process included the following three steps: (1) data collection, (2) data preprocessing, and (3) data analysis. The data collection and preprocessing procedures are shown in Figure 1.

Figure 1. Procedures of data collection and preprocessing.



Data Collection

First, a participant pool of active Weibo users was created. According to a previous study, 1,953,485 active Weibo users were identified as potential participants [34]. The application programming interface (API) platform of Sina Weibo enables programmatic access to Weibo data that users choose to share with the public. As Sina Weibo places limits and quotas on API requests, among all potential participants, 1.06 million active users with available data were finally included in the participant pool.

Second, a database of Weibo posts was constructed. In May 2020, using API, a vast amount of publicly available Weibo posts from 1.06 million active users in the participant pool since their registration (the 2020 official numbers of monthly and daily active users: 511 million and 224 million, respectively [35]), were downloaded.

Third, several relevant Weibo posts were identified through database searches. To obtain the posts reflecting attitudes toward cybersuicides and offline suicides, two sets of keywords were used to search the database, including *livestreamed suicide* (直播自杀 and 自杀直播) and *suicide* (自杀). It is worth noting that although cybersuicide is a rapidly developing form of suicide, an overwhelming majority of suicide incidents still occur offline. For example, between 2003 and 2016, 193 incidents of livestreamed suicide occurred in China [20], whereas in 2019, the number of suicides in China had reached 116,324 [1]. Therefore, unless otherwise stated, the Chinese people commonly use the term *suicide* to refer to traditional offline suicide.

The process of database searches included the following three steps: (1) a total of 4460 posts with keywords *livestreamed suicide* were searched and obtained (cybersuicide group); (2) a total of 933,172 posts with keywords *suicide* (but without keywords *livestreamed suicide*) were searched and obtained (offline suicide group); (3) to balance the number of posts in each group, using simple random sampling, 4500 posts were randomly selected from 933,172 posts in the offline suicide group (cybersuicide: 4460 posts and offline suicide: 4500 posts).

Data Preprocessing

After data collection, preprocessing was performed on the raw data to prepare them for further analysis.

First, to exclude irrelevant posts and reclassify misclassified posts, manual scrutiny of the collected data was conducted.

In this study, irrelevant posts were considered as (1) posts that depicted suicides in fictional works (eg, movies, television programs, and novels), (2) posts that focused on suicides in nonhuman animals (eg, dogs), and (3) posts that used suicide-related keywords for nonsuicidal purposes (eg, making a bet). After the removal of irrelevant posts, 7244 posts (cybersuicide: 4460 – 136 = 4324 and offline suicide: 4500 − 1580 = 2920) remained.

Subsequently, 77 posts in the offline suicide group were reclassified as posts related to cybersuicide (livestreamed suicide: n=69, 90% posts; suicide *game*: n=4, 5% posts; prosuicide website and forum: n=2, 3% posts; and internet suicide pact: n=2, 3% posts). As this study primarily focused on livestreamed suicide rather than other types of cybersuicide, 8 posts related to the other 3 types of cybersuicide were excluded from further analysis.

Therefore, the final sample of this study included 7236 posts (cybersuicide: 4324 + 69 = 4393; offline suicide: 2920 – 69 – 8 = 2843). The demographic characteristics of the participants in the final sample are presented in Table 1.

Table 1. Demographics of participants.

Demographics	All Weibo posts (N=7236), n (%)	Cybersuicide (n=4393), n (%)	Offline suicide (n=2843), n (%)
Gender			
Male	4062 (56.14)	2473 (56.29)	1589 (55.89)
Female	3174 (43.86)	1920 (43.71)	1254 (44.11)
Regions			
North China	1312 (18.13)	812 (18.48)	500 (17.59)
Northeast China	299 (4.13)	191 (4.35)	108 (3.8)
East China	2277 (31.47)	1330 (30.28)	947 (33.31)
Central China	374 (5.17)	233 (5.3)	141 (4.96)
South China	1345 (18.59)	801 (18.23)	544 (19.13)
Southwest China	493 (6.81)	330 (7.51)	163 (5.73)
Northwest China	224 (3.1)	133 (3.03)	91 (3.2)
International and unspecified	912 (12.6)	563 (12.82)	349 (12.28)

Second, to extract psycholinguistic features from each of the 7236 posts automatically, the Simplified Chinese version of Linguistic Inquiry and Word Count software was used. The Simplified Chinese version of Linguistic Inquiry and Word Count is a reliable and valid text analysis tool for the automatic estimation of word frequency in multiple psychologically meaningful categories, including linguistic processes (eg, personal pronouns), psychological processes (eg, affective processes), personal concerns (eg, achievement), spoken categories (eg, assent), and punctuation categories (eg, periods) [36]. After feature extraction, the standardized values of psycholinguistic features were estimated for further analysis.

Data Analysis

Human Coding

To explore the differences in types of stigmatizing attitudes across cybersuicides and offline suicides, a content analysis was performed on all 7236 posts to determine whether each of them reflected stigma. The coding framework was developed based on expert consensus and available evidence. Specifically, in this study, a researcher (AL) reviewed relevant studies [25,37] and then performed an inductive content analysis on all 7236 posts to develop an initial coding framework. Subsequently, another two researchers (DJ and TZ) provided feedback to emend the initial framework. Using the amended framework (Table S1 in Multimedia Appendix 1), 2 independent human coders were recruited and trained to analyze all 7236 posts. The levels of agreement between the 2 coders were measured by Cohen *k* coefficient, and disagreements were resolved by the decisions of a researcher (AL). All individuals who participated in developing the coding framework and performing manual coding of posts had considerable experience in coding qualitative materials.

Construction of Classification Models

To explore the linguistic differences in stigmatizing expressions across cybersuicides and offline suicides, 2 groups of classification models were built using Waikato Environment for Knowledge Analysis (version 3.9.4; University of Waikato) software. Waikato Environment for Knowledge Analysis provides tools for developing machine learning techniques and applying them to practical data mining problems.

The first group of classification models was built to investigate whether linguistic differences existed in the expression of stigma in general (ie, cybersuicide-related or offline suicide-related stigma as a whole) between cybersuicides and offline suicides. The human coding results were considered as the ground truth for the validation of the classification models. In this study, an imbalanced data problem existed. For example, the number of stigmatizing posts belonging to the offline suicide class (minority class: 588 posts) was obviously lower than those belonging to the cybersuicide class (majority class: 1556 posts). Imbalanced data sets pose a challenge for machine learning modeling, as this problem may result in models with poor predictive performance, especially for the minority class. To handle this problem, using simple random sampling, a certain number of posts were randomly selected from the majority class to obtain a well-balanced data set. Subsequently, to improve classification accuracy, a subset of psycholinguistic features was selected for use in model construction. Specifically, a series of 2-tailed independent *t* tests were conducted to compare the values of all extracted psycholinguistic features between stigmatizing posts in the cybersuicide and offline suicide groups, and then effect sizes (Cohen *d* coefficient) were computed from the estimated *t* values. Features that were statistically significant at .05 and had a Cohen *d* >0.20 or <−0.20 were considered as key features. Finally, using four different machine learning algorithms (Naïve Bayes, support vector machine, multilayer perceptron neural network, and random forest [RF]), 4 classification models were constructed based on the selected key features. Using a 5-fold cross-validation technique, the classification performance of the established models was evaluated in terms of precision, recall, and *F*-measure.

It is worth noting that the good classification performance of models in the first group may be attributed to the existence of differences in the amount and distribution of negative stereotypes across cybersuicides and offline suicides rather than

the existence of linguistic differences in stigmatizing expressions across cybersuicides and offline suicides. To clarify this issue, the second group of classification models was built to investigate whether linguistic differences existed in the expression of certain negative stereotypes across cybersuicides and offline suicides. To obtain sufficient data for further analysis, in this study, posts reflecting two negative stereotypes (ie, *stupid and shallow* and *glorified and normalized*) were examined. Well-balanced data sets, key features, and classification models were obtained using the aforementioned methods.

Ethics Approval

The study protocol was reviewed and approved by the institutional review board of the Institute of Psychology, Chinese Academy of Sciences (protocol number: H15009). Informed consent was not obtained, as this study was based on publicly available data and involved no personally identifiable data collection or analysis.

Results

Human Coding

The Cohen *k* coefficients for *attitudes* and *negative stereotypes* reached 0.88 and 0.81, respectively, indicating almost perfect agreement [38]. The results of human coding are presented in Table 2.

Table 2. Results of human coding (N=7236).

Categories	Cybersuicide, n (%)	Offline suicide, n (%)
Attitudes	4393 (100)	2843 (100)
Stigmatizing	1556 (35.4)	588 (20.7)
Nonstigmatizing	2837 (64.6)	2255 (79.3)
Negative stereotypes	1556 (100)	588 (100)
Weak and pathetic	114 (7.3)	80 (13.6)
Self-centered	97 (6.2)	44 (7.5)
Stupid and shallow	528 (33.9)	129 (21.9)
False representation	387 (24.9)	13 (2.2)
Glorified and normalized	148 (9.5)	195 (33.2)
Immoral	111 (7.1)	69 (11.7)
Strange	59 (3.8)	14 (2.4)
Embarrassing	24 (1.5)	7 (1.2)
Vengeful	40 (2.6)	9 (1.5)
Mad	48 (3.1)	28 (4.8)

For stigma in general, posts on cybersuicide were more likely than posts on offline suicide to contain stigmatizing expressions (χ^2_1=179.8; *P*<.001). Similar results were found for different genders and regions, including men (χ^2_1=66.7; *P*<.001), women (χ^2_1=121.0; *P*<.001), North China (NC; χ^2_1=37.2; *P*<.001), East China (EC; χ^2_1=56.4; *P*<.001), Central China (CC; χ^2_1=10.4; *P*=.001), South China (SC; χ^2_1=37.6; *P*<.001), and Southwest China (SWC; χ^2_1=11.8; *P*=.001).

For negative stereotypes, posts on cybersuicide were often coded as *stupid and shallow* (528/1556, 33.93%) and *false representation* (387/1556, 24.87%), whereas posts on offline suicide were often coded as *glorified and normalized* (195/588, 33.2%) and *stupid and shallow* (129/588, 21.9%). Furthermore, posts on cybersuicide were more likely than posts on offline suicide to be coded as *stupid and shallow* (χ^2_1=28.9; *P*<.001) and *false representation* (χ^2_1=144.4; *P*<.001), whereas posts on offline suicide were more likely than posts on cybersuicide to be coded as *weak and pathetic* (χ^2_1=20.4; *P*<.001), *glorified and normalized* (χ^2_1=177.6; *P*<.001), and *immoral* (χ^2_1=11.8; *P*=.001). Similar results were found for different genders and regions. Specifically, significant differences in the proportions of posts coded as *weak and pathetic* were observed for posts by women (χ^2_1=21.0; *P*<.001) and from NC (χ^2_1=10.6; *P*=.001) and EC (χ^2_1=3.9; *P*=.048). Significant differences in the proportions of posts coded as *stupid and shallow* were observed for posts by men (χ^2_1=12.6; *P*<.001) and women (χ^2_1=16.0; *P*<.001) and from NC (χ^2_1=9.1; *P*=.003), EC (χ^2_1=7.9; *P*=.005), SC (χ^2_1=5.4; *P*=.02), and SWC (χ^2_1=8.0; *P*=.005). Significant differences in the proportions of posts coded as *false representation* were observed for posts by men (χ^2_1=86.4; *P*<.001) and women (χ^2_1=59.5; *P*<.001) and from NC (χ^2_1=25.5; *P*<.001), Northeast China (Fisher exact test: *P*=.004), EC (χ^2_1=48.2; *P*<.001), CC (Fisher exact test: *P*=.006), SC (χ^2_1=16.5; *P*<.001), SWC (χ^2_1=8.1; *P*=.004), and Northwest China (Fisher exact test: *P*=.003). Significant differences in the proportions of posts coded as *glorification and normalized* were observed for posts by men (χ^2_1=127.3; *P*<.001) and women

(χ^2_1=55.1; *P*<.001) and from NC (χ^2_1=37.5; *P*<.001), Northeast China (Fisher exact test: *P*<.001), EC (χ^2_1=60.7; *P*<.001), CC (Fisher exact test: *P*=.01), SC (χ^2_1=17.6; *P*<.001), and SWC (χ^2_1=21.5; *P*<.001). Significant differences in the proportions of posts coded as *immoral* were observed for posts by men (χ^2_1=7.8; *P*=.005) and women (χ^2_1=3.9; *P*=.048) and from EC (χ^2_1=4.8; *P*=.03) and Northwest China (Fisher exact test: *P*=.008). In addition, posts on offline suicide were more likely than posts on cybersuicide to be coded as *mad* for posts by women (χ^2_1=5.4; *P*=.02) and from CC (Fisher exact test: *P*=.04) and SC (Fisher exact test: *P*=.03).

Construction of Classification Models

Stigma in General

For exploring linguistic differences in the expression of stigma in general between cybersuicides and offline suicides, to achieve a balanced data set, 600 stigmatizing posts were randomly selected from posts in cybersuicide group (cybersuicide: 600 posts and offline suicide: 588 posts). A total of 6 key features were selected for use in the model construction (Table S2 in Multimedia Appendix 1). The RF model exhibited the best classification performance (precision=0.86, recall=0.85, and *F*-measure=0.85; Table 3).

Table 3. Performance of classification models.

Models	Stigma in general	Stupid and shallow	Glorified and normalized
Naïve Bayes			
Precision	0.73	0.72	0.72
Recall	0.73	0.72	0.72
F-measure	0.73	0.72	0.72
Support vector machine			
Precision	0.83	0.80	0.79
Recall	0.83	0.80	0.78
F-measure	0.83	0.79	0.78
Multilayer perceptron neural network			
Precision	0.85	0.84	0.78
Recall	0.85	0.83	0.77
F-measure	0.85	0.83	0.77
Random forest			
Precision	0.86	0.84	0.81
Recall	0.85	0.84	0.81
F-measure	0.85	0.84	0.81

Stupid and Shallow

For exploring linguistic differences in the expression of *stupid and shallow* between cybersuicides and offline suicides, to achieve a balanced data set, a total of 150 *stupid and shallow*–related posts were randomly selected from posts in cybersuicide group (cybersuicide: 150 posts and offline suicide: 129 posts). A total of 4 key features were selected for use in the model construction (Table S2 in Multimedia Appendix 1). The RF model exhibited the best classification performance (precision=0.84, recall=0.84, and *F*-measure=0.84; Table 3).

Glorified and Normalized

To explore the linguistic differences in the expression of *glorified and normalized* between cybersuicides and offline suicides (cybersuicide: 148 posts and offline suicide: 195 posts), 28 key features were selected for use in model construction (Table S2 in Multimedia Appendix 1). The RF model exhibited the best classification performance (precision=0.81, recall=0.81, and *F*-measure=0.81; Table 3).

Discussion

Principal Findings

To our knowledge, this study provides the first systematic analysis of the differences in stigmatizing attitudes toward cybersuicides and offline suicides. The results of this study have implications for reducing stigma against suicide.

First, it is necessary to confront and reduce stigma against cybersuicide. In this study, a large proportion of cybersuicide-related and offline suicide-related posts were coded as *stigmatizing* (1556/4393, 35.42% and 588/2843, 20.68%, respectively), although most posts about either type of suicide were not overtly stigmatizing. Cybersuicide allows people with suicidal intentions to interact with their audience in the last moments before death. If the audience responds appropriately to people with suicidal intentions, the likelihood of death would be reduced [39-41]. The prevalence of stigma may hamper audience from effectively responding to people with suicidal

intentions [25,42-44]. Therefore, cybersuicide deserves to focus on antistigma campaigns.

Second, the public reacts differently to cybersuicides and offline suicides. In terms of attitude types, in this study, cybersuicide was observed to carry more stigma than offline suicide (cybersuicide: 1556/4393, 35.42% and offline suicide: 588/2843, 20.68%; χ^2_1=179.8; *P*<.001), implying that the public may react more negatively to cybersuicide than to offline suicide. This is consistent with 2 previous studies that explored audience responses to cybersuicide and offline suicide incidents [24,45]. Furthermore, compared with offline suicide, cybersuicide was more likely to be considered as *stupid and shallow* (χ^2_1=28.9; *P*<.001) and *false representation* (χ^2_1=144.4; *P*<.001) and was less likely to be considered as *weak and pathetic* (χ^2_1=20.4; *P*<.001), *glorified and normalized* (χ^2_1=177.6; *P*<.001), and *immoral* (χ^2_1=11.8; *P*=.001). Similar results were also found for different genders and regions. This indicates that antistigma campaigns targeting offline suicide may not be effective in changing stigmatizing attitudes toward cybersuicide, suggesting the need for public awareness campaigns that specifically target cybersuicide. In addition, the most notable and obvious difference between the two types of suicide was found in *false representation* stigma (cybersuicide: 387/1556, 24.87% and offline suicide: 13/588, 2.2%). The misbelief that cybersuicide is not real may influence stigmatizing responses elicited mainly by the value judgment of suicide death itself but not by the perception of suicidal motives, causes, and methods. For example, the glorification and immorality of suicide represent two distinct values: that it is right or wrong to deliberately take one's own life. However, dismissing cybersuicide as a nonreal condition contradicts the assumption of such 2 stigmas that suicide is a real condition. This might be the reason why cybersuicide was less likely to be considered as *glorified and normalized* and *immoral* than offline suicide. It is also worth noting that cybersuicide was commonly considered as *stupid and shallow* (528/1556, 33.93%) and *false representation* (387/1556, 24.87%), whereas offline suicide was commonly considered as *glorified and normalized* (195/588, 33.2%) and *stupid and shallow* (129/588, 21.9%). According to previous studies, inappropriate audience responses are associated with greater false representation stigma [25], and both suicide ideation and suicide contagion are associated with greater glorification of suicide [46,47]. This indicates that the reduction of stigma against cybersuicide may contribute more in improving audience responses, whereas the reduction of stigma against offline suicide may contribute more in preventing suicide attempts.

Apart from the differences in attitude types, linguistic differences in the expression of stigma between cybersuicides and offline suicides also exist. Such differences existed not only at the level of stigma in general but also at the level of negative stereotypes. In this study, the *F*-measure values of the classification models ranged from 0.81 to 0.85. Compared with other similar studies [48,49], the classification models achieved satisfying accuracy. These results support the conclusion that the way people perceive cybersuicide is very different from the way people perceive offline suicide, implying that cybersuicide may have a different structure from offline suicide [10].

Third, the use of linguistic analysis methods can facilitate the identification of suicide-related stigma in mass media. Mass media is a major contributor to the dissemination of incorrect information, which may reinforce negative stereotypes surrounding mental illness [50-52]. It would be greatly helpful if mass media campaigns were developed to raise public awareness and challenge the stigma against mental illness. However, because of the sheer volume of information in mass media, it is difficult for human coders to identify and analyze stigmatizing information efficiently, suggesting the need for automatic detection of stigma. Linguistic analysis methods can be used to understand language use patterns of stigmatizing expressions and to construct computational models for the automatic detection of stigma [30,33,53]. For example, in this study, stigmatizing expressions of cybersuicide were associated with more frequent use of words related to leisure (eg, chat) and work (eg, dissemination), which may be attributed to the fact that cybersuicide makes suicidal acts highly public and contradicts the public perception of suicide. By contrast, stigmatizing expressions of offline suicide were associated with more frequent use of words related to achievement (eg, hero), which may be attributed to the higher prevalence of suicide glorification (cybersuicide: 148/1556, 9.51% and offline suicide: 195/588, 33.2%). The constructed machine learning models performed well in classifying stigma toward cybersuicides and offline suicides. This indicates that understanding the linguistic differences in stigmatizing expressions between cybersuicides and offline suicides would make automatic detection more precise. Automatic stigma detection would further improve antistigma efforts. The dissemination of internet messages can be helpful in promoting changes in health-related behaviors [54,55]. However, internet campaigns work best when targeted. With the help of automatic detection and machine learning algorithms, it should be easier to efficiently design and deliver target-specific messages to the population and would promote public mental health awareness.

Limitations

This study has some limitations. First, this study mainly focused on the stigma against livestreamed suicide. Therefore, it is unclear whether these findings are applicable to other types of cybersuicide. Second, social media users are not representative of all people in China. Therefore, the findings may not be applicable to nonusers. Third, the API of Sina Weibo only allowed us to download posts from a certain number of registered users. Therefore, these findings should be further confirmed on a larger scale and in more diverse populations in the future. Fourth, because of the lack of posts obtained from people with suicidal intentions, this study cannot analyze the stigma that people with suicidal intentions put on themselves (ie, self-stigma). Fifth, because of the lack of information about user types (eg, celebrities and general users), this study cannot investigate the differences in attitudes across different types of users and cannot compare attitude responses elicited by the deaths by suicide of different types of users.

Conclusions

This study used a nonintrusive method to directly and systematically examine the differences in stigmatizing attitudes toward cybersuicides and offline suicides. The results of this study support the conclusion that the way people perceive cybersuicide is very different from the way people perceive offline suicide and offer insight into the reduction of stigma against suicide.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary Tables S1 and S2.
[DOCX File , 22 KB - jmir_v24i4e36489_app1.docx]

References

1. World Health Organization. Suicide worldwide in 2019: global health estimates. Geneva, Switzerland: World Health Organization; Jun 16, 2021.
2. Corrigan PW, Penn DL. Lessons from social psychology on discrediting psychiatric stigma. Am Psychol 1999 Sep;54(9):765-776. [doi: 10.1037//0003-066x.54.9.765] [Medline: 10510666]
3. Sudak H, Maxim K, Carpenter M. Suicide and stigma: a review of the literature and personal reflections. Acad Psychiatry 2008;32(2):136-142. [doi: 10.1176/appi.ap.32.2.136] [Medline: 18349334]
4. Carpiniello B, Pinna F. The reciprocal relationship between suicidality and stigma. Front Psychiatry 2017;8:35 [FREE Full text] [doi: 10.3389/fpsyt.2017.00035] [Medline: 28337154]
5. Reynders A, Kerkhof AJ, Molenberghs G, Van Audenhove C. Attitudes and stigma in relation to help-seeking intentions for psychological problems in low and high suicide rate regions. Soc Psychiatry Psychiatr Epidemiol 2014 Feb;49(2):231-239. [doi: 10.1007/s00127-013-0745-4] [Medline: 23896893]
6. Carrara BS, Fernandes RH, Bobbili SJ, Ventura CA. Health care providers and people with mental illness: an integrative review on anti-stigma interventions. Int J Soc Psychiatry 2021 Nov;67(7):840-853. [doi: 10.1177/0020764020985891] [Medline: 33380251]
7. Siegel JT, Flores-Medel E, Martinez DA, Berger DE. Can mental health anti-stigma messages have untoward effects on some people with depression?: an exploratory study. J Health Commun 2019;24(11):821-828. [doi: 10.1080/10810730.2019.1672838] [Medline: 31621538]
8. Corrigan PW. Best practices: strategic stigma change (SSC): five principles for social marketing campaigns to reduce stigma. Psychiatr Serv 2011 Aug;62(8):824-826. [doi: 10.1176/ps.62.8.pss6208_0824] [Medline: 21807820]
9. Starcevic V, Aboujaoude E. Cyberchondria, cyberbullying, cybersuicide, cybersex: "new" psychopathologies for the 21st century? World Psychiatry 2015 Feb;14(1):97-100 [FREE Full text] [doi: 10.1002/wps.20195] [Medline: 25655165]
10. Fratini A, Hemer SR. Broadcasting your death through livestreaming: understanding cybersuicide through concepts of performance. Cult Med Psychiatry 2020 Dec;44(4):524-543. [doi: 10.1007/s11013-020-09671-9] [Medline: 32222946]
11. La Fontaine J. Explaining suicide: an afterword. Cult Med Psychiatry 2012 Jun;36(2):409-418. [doi: 10.1007/s11013-012-9256-0] [Medline: 22395949]
12. Salem T. Physician-assisted suicide. Promoting autonomy--or medicalizing suicide? Hastings Cent Rep 1999;29(3):30-36. [Medline: 10420304]
13. Jamison KR. Suicide and bipolar disorder. J Clin Psychiatry 2000;61 Suppl 9:47-51. [Medline: 10826661]
14. Seko Y. On Deathcasting: alone together on the edge of death. In: Nixon PG, Rawal R, Funk A, editors. Digital media across the life course. London, UK: Routledge; 2016:172-189.
15. Birbal R, Maharajh HD, Birbal R, Clapperton M, Jarvis J, Ragoonath A, et al. Cybersuicide and the adolescent population: challenges of the future? Int J Adolesc Med Health 2009;21(2):151-159. [Medline: 19702194]
16. Money TT, Batterham PJ. Sociocultural factors associated with attitudes toward suicide in Australia. Death Stud 2021;45(3):219-225. [doi: 10.1080/07481187.2019.1626943] [Medline: 31190630]
17. Luo A, He H, Mohamed S, Rosenheck R. Medical student attitudes towards people with mental illness in China: a qualitative study. Cult Med Psychiatry 2018 Sep;42(3):535-551. [doi: 10.1007/s11013-018-9568-9] [Medline: 29508204]
18. Chiles C, Stefanovics E, Rosenheck R. Attitudes of students at a US medical school toward mental illness and its causes. Acad Psychiatry 2017 Jun;41(3):320-325. [doi: 10.1007/s40596-016-0508-0] [Medline: 26951266]
19. Chen R, An J, Ou J. Suicidal behaviour among children and adolescents in China. Lancet Child Adolesc Health 2018 Aug;2(8):551-553. [doi: 10.1016/S2352-4642(18)30170-6] [Medline: 30119712]
20. Li A, Huang X, Zhu T. A systematic analysis of online broadcasts of suicidality in China. Asia Pac Psychiatry 2018 Sep;10(3):e12302. [doi: 10.1111/appy.12302] [Medline: 28976066]
21. Li A, Jiao D, Liu X, Sun J, Zhu T. A psycholinguistic analysis of responses to live-stream suicides on social media. Int J Environ Res Public Health 2019 Aug 09;16(16):2848 [FREE Full text] [doi: 10.3390/ijerph16162848] [Medline: 31404975]

22. O'Dea B, Achilles MR, Larsen ME, Batterham PJ, Calear AL, Christensen H. The rate of reply and nature of responses to suicide-related posts on Twitter. Internet Interv 2018 Sep;13:105-107 [FREE Full text] [doi: 10.1016/j.invent.2018.07.004] [Medline: 30206524]
23. Ma J, Zhang W, Harris K, Chen Q, Xu X. Dying online: live broadcasts of Chinese emerging adult suicides and crisis response behaviors. BMC Public Health 2016 Aug 11;16(1):774 [FREE Full text] [doi: 10.1186/s12889-016-3415-0] [Medline: 27515312]
24. Fu KW, Cheng Q, Wong PW, Yip PS. Responses to a self-presented suicide attempt in social media: a social network analysis. Crisis 2013 Jan 01;34(6):406-412 [FREE Full text] [doi: 10.1027/0227-5910/a000221] [Medline: 23871954]
25. Li A, Huang X, Hao B, O'Dea B, Christensen H, Zhu T. Attitudes towards suicide attempts broadcast on social media: an exploratory study of Chinese microblogs. PeerJ 2015;3:e1209 [FREE Full text] [doi: 10.7717/peerj.1209] [Medline: 26380801]
26. Su H, Hunston S. Language patterns and attitude revisited: adjective patterns, attitude and appraisal. Funct Lang 2019 Nov 25;26(3):343-371. [doi: 10.1075/fol.16030.su]
27. ŚLebioda A. Stigmatized by the language –linguistic labels of the disabled peoplein Poland and the United States. J Educ Cult Soc 2020 Jan 10;4(2):404-412. [doi: 10.15503/jecs20132.404.412]
28. Puhl R, Peterson JL, Luedicke J. Motivating or stigmatizing? Public perceptions of weight-related language used by health providers. Int J Obes (Lond) 2013 Apr;37(4):612-619. [doi: 10.1038/ijo.2012.110] [Medline: 22777543]
29. Pennebaker JW, Mehl MR, Niederhoffer KG. Psychological aspects of natural language. use: our words, our selves. Annu Rev Psychol 2003;54:547-577. [doi: 10.1146/annurev.psych.54.101601.145041] [Medline: 12185209]
30. Yu L, Jiang W, Ren Z, Xu S, Zhang L, Hu X. Detecting changes in attitudes toward depression on Chinese social media: a text analysis. J Affect Disord 2021 Feb 01;280(Pt A):354-363. [doi: 10.1016/j.jad.2020.11.040] [Medline: 33221722]
31. Li A, Huang X, Jiao D, O'Dea B, Zhu T, Christensen H. An analysis of stigma and suicide literacy in responses to suicides broadcast on social media. Asia Pac Psychiatry 2018 Mar;10(1):e12314. [doi: 10.1111/appy.12314] [Medline: 29383880]
32. Oscar N, Fox PA, Croucher R, Wernick R, Keune J, Hooker K. Machine learning, sentiment analysis, and tweets: an examination of Alzheimer's disease stigma on Twitter. J Gerontol B Psychol Sci Soc Sci 2017 Sep 01;72(5):742-751. [doi: 10.1093/geronb/gbx014] [Medline: 28329835]
33. Li A, Jiao D, Liu X, Zhu T. A comparison of the psycholinguistic styles of schizophrenia-related stigma and depression-related stigma on social media: content analysis. J Med Internet Res 2020 Apr 21;22(4):e16470 [FREE Full text] [doi: 10.2196/16470] [Medline: 32314969]
34. Li L, Li A, Hao B, Guan Z, Zhu T. Predicting active users' personality based on micro-blogging behaviors. PLoS One 2014;9(1):e84997 [FREE Full text] [doi: 10.1371/journal.pone.0084997] [Medline: 24465462]
35. Weibo reports third quarter 2020 unaudited financial results. Weibo Corporation. 2020. URL: http://ir.weibo.com/news-releases/news-release-details/weibo-reports-third-quarter-2020-unaudited-financial-results/ [accessed 2022-03-25]
36. Zhao N, Jiao D, Bai S, Zhu T. Evaluating the validity of simplified Chinese version of LIWC in detecting psychological expressions in short texts on social network services. PLoS One 2016;11(6):e0157947 [FREE Full text] [doi: 10.1371/journal.pone.0157947] [Medline: 27322382]
37. Batterham PJ, Calear AL, Christensen H. The Stigma of Suicide Scale. Psychometric properties and correlates of the stigma of suicide. Crisis 2013 Jan 01;34(1):13-21. [doi: 10.1027/0227-5910/a000156] [Medline: 22846447]
38. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
39. Armstrong G, Ironfield N, Kelly CM, Dart K, Arabena K, Bond K, et al. Re-development of mental health first aid guidelines for supporting Aboriginal and Torres Strait islanders who are experiencing suicidal thoughts and behaviour. BMC Psychiatry 2018 Jul 16;18(1):228 [FREE Full text] [doi: 10.1186/s12888-018-1809-5] [Medline: 30012118]
40. Ross AM, Kelly CM, Jorm AF. Re-development of mental health first aid guidelines for suicidal ideation and behaviour: a Delphi study. BMC Psychiatry 2014 Sep 13;14:241 [FREE Full text] [doi: 10.1186/s12888-014-0241-8] [Medline: 25213799]
41. Owens D, Horrocks J, House A. Fatal and non-fatal repetition of self-harm. Systematic review. Br J Psychiatry 2002 Sep;181:193-199. [doi: 10.1192/bjp.181.3.193] [Medline: 12204922]
42. Boukouvalas E, El-Den S, Murphy AL, Salvador-Carulla L, O'Reilly CL. Exploring health care professionals' knowledge of, attitudes towards, and confidence in caring for people at risk of suicide: a systematic review. Arch Suicide Res 2020;24(sup2):S1-31. [doi: 10.1080/13811118.2019.1586608] [Medline: 30856366]
43. Kresovich A. The influence of pop songs referencing anxiety, depression, and suicidal ideation on college students' mental health empathy, stigma, and behavioral intentions. Health Commun 2022 May;37(5):617-627. [doi: 10.1080/10410236.2020.1859724] [Medline: 33334191]
44. Osafo J, Akotia CS, Boakye KE, Dickson E. Between moral infraction and existential crisis: exploring physicians and nurses' attitudes to suicide and the suicidal patient in Ghana. Int J Nurs Stud 2018 Sep;85:118-125. [doi: 10.1016/j.ijnurstu.2018.05.017] [Medline: 29885548]
45. Larsen ME, Boonstra TW, Batterham PJ, ODea B, Paris C, Christensen H. We feel: mapping emotion on Twitter. IEEE J Biomed Health Inform 2015 Jul;19(4):1246-1252. [doi: 10.1109/JBHI.2015.2403839] [Medline: 25700477]

46. Batterham PJ, Han J, Calear AL, Anderson J, Christensen H. Suicide stigma and suicide literacy in a clinical sample. Suicide Life Threat Behav 2019 Aug;49(4):1136-1147. [doi: 10.1111/sltb.12496] [Medline: 30079479]
47. Beautrais A, Hendin H, Yip P, Takahashi Y, Chia BH, Schmidtke A. Improving portrayal of suicide in the media in Asia. In: Hendin H, Phillips M, Vijayakumar L, Pirkis J, Wang H, Yip P, editors. Suicide Prevention in Asia. Geneva, Switzerland: World Health Organization; Jan 2008:39-49.
48. Budenz A, Klassen A, Purtle J, Yom Tov E, Yudell M, Massey P. Mental illness and bipolar disorder on Twitter: implications for stigma and social support. J Ment Health 2020 Apr;29(2):191-199. [doi: 10.1080/09638237.2019.1677878] [Medline: 31694433]
49. Straton N, Jang H, Ng R, Vatrapu R, Mukkamala RR. Computational modeling of stigmatized behaviour in pro-vaccination and anti-vaccination discussions on social media. In: IEEE International Conference on Bioinformatics and Biomedicine. 2019 Feb 6 Presented at: BIBM '19; Nov 18-21, 2019; San Diego, CA, USA p. 2673-2681. [doi: 10.1109/bibm47256.2019.8983311]
50. Koike S, Yamaguchi S, Ojio Y, Ohta K, Ando S. Effect of name change of schizophrenia on mass media between 1985 and 2013 in Japan: a text data mining analysis. Schizophr Bull 2016 May;42(3):552-559 [FREE Full text] [doi: 10.1093/schbul/sbv159] [Medline: 26614786]
51. Morgan AJ, Jorm AF. Recall of news stories about mental illness by Australian youth: associations with help-seeking attitudes and stigma. Aust N Z J Psychiatry 2009 Sep;43(9):866-872. [doi: 10.1080/00048670903107567] [Medline: 19670060]
52. Dietrich S, Heider D, Matschinger H, Angermeyer MC. Influence of newspaper reporting on adolescents' attitudes toward people with mental illness. Soc Psychiatry Psychiatr Epidemiol 2006 Apr;41(4):318-322. [doi: 10.1007/s00127-005-0026-y] [Medline: 16424966]
53. Li A, Jiao D, Zhu T. Detecting depression stigma on social media: a linguistic analysis. J Affect Disord 2018 May;232:358-362. [doi: 10.1016/j.jad.2018.02.087] [Medline: 29510353]
54. Yom-Tov E, Brunstein-Klomek A, Mandel O, Hadas A, Fennig S. Inducing behavioral change in seekers of pro-anorexia content using internet advertisements: randomized controlled trial. JMIR Ment Health 2018 Feb 22;5(1):e6 [FREE Full text] [doi: 10.2196/mental.8212] [Medline: 29472176]
55. Cheng Q, Yom-Tov E. Do search engine helpline notices aid in preventing suicide? Analysis of archival data. J Med Internet Res 2019 Mar 26;21(3):e12235 [FREE Full text] [doi: 10.2196/12235] [Medline: 30912753]

Abbreviations

API: application programming interface
CC: Central China
EC: East China
NC: North China
RF: random forest
SC: South China
SWC: Southwest China

Edited by R Kukafka; submitted 16.01.22; peer-reviewed by L Zhang, A Budenz; comments to author 11.02.22; revised version received 19.02.22; accepted 13.03.22; published 08.04.22.

Please cite as:
Li A, Jiao D, Zhu T
Stigmatizing Attitudes Across Cybersuicides and Offline Suicides: Content Analysis of Sina Weibo
J Med Internet Res 2022;24(4):e36489
URL: https://www.jmir.org/2022/4/e36489
doi:10.2196/36489
PMID:35394437

Original Paper

Conditions for the Successful Integration of an eHealth Tool "StopBlues" Into Community-Based Interventions in France: Results From a Multiple Correspondence Analysis

Kathleen Turmaine[1], MSc, MPH; Agnès Dumas[1], PhD; Karine Chevreul[1,2,3], MD, PhD; PRINTEMPS Consortium[4]

[1]Université Paris Cité, ECEVE, UMR 1123, Inserm, Paris, France

[2]Assistance Publique-Hôpitaux de Paris, URC Eco Ile-de-France, Paris, France, Paris, France

[3]Assistance Publique-Hôpitaux de Paris, Hôpital Robert Debré, Unité d'épidémiologie clinique, Paris, France

[4]See Authors' Contributions

Corresponding Author:
Kathleen Turmaine, MSc, MPH
Université Paris Cité, ECEVE, UMR 1123, Inserm
10 avenue de Verdun
Paris, 75010
France
Phone: 33 1 57 27 86 89
Email: kathleen.turmaine@inserm.fr

Abstract

Background: For over a decade, digital health has held promise for enabling broader access to health information, education, and services for the general population at a lower cost. However, recent studies have shown mixed results leading to a certain disappointment regarding the benefits of eHealth technologies. In this context, community-based health promotion represents an interesting and efficient conceptual framework that could help increase the adoption of digital health solutions and facilitate their evaluation.

Objective: To understand how the local implementation of the promotion of an eHealth tool, StopBlues (SB), aimed at preventing psychological distress and suicide, varied according to local contexts and if the implementation was related to the use of the tool.

Methods: The study was nested within a cluster-randomized controlled trial that was conducted to evaluate the effectiveness of the promotion, with before and after observation (NCT03565562). Data from questionnaires, observations, and institutional sources were collected in 27 localities where SB was implemented. A multiple correspondence analysis was performed to assess the relations between context, type of implementation and promotion, and use of the tool.

Results: Three distinct promotion patterns emerged according to the profiles of the localities that were associated with specific SB utilization rates. From highest to lowest utilization rates, they are listed as follows: the privileged urban localities, investing in health that implemented a high-intensity and digital promotion, demonstrating a greater capacity to take ownership of the project; the urban, but less privileged localities that, in spite of having relatively little experience in health policy implementation, managed to implement a traditional and high-intensity promotion; and the rural localities, with little experience in addressing health issues, that implemented low-intensity promotion but could not overcome the challenges associated with their local context.

Conclusions: These findings indicate the substantial influence of local context on the reception of digital tools. The urban and socioeconomic status profiles of the localities, along with their investment and pre-existing experience in health, appear to be critical for shaping the promotion and implementation of eHealth tools in terms of intensity and use of digital communication. The more digital channels used, the higher the utilization rates, ultimately leading to the overall success of the intervention.

International Registered Report Identifier (IRRID): RR2-10.1186/s13063-020-04464-2

(J Med Internet Res 2022;24(4):e30218) doi:10.2196/30218

KEYWORDS

eHealth; internet-based intervention; community participation; health promotion; prevention; mental health

Introduction

Digital Health and Community-Based Health Interventions

For over a decade, digital health has held promise for enabling broader access to health information, education, and services for the general population, all at lower cost [1,2]. The unbridled development of the digital health market has led to more than 320,000 health apps currently available, including around 10,000 specifically aimed at promoting mental and behavioral health. In spite of this, adoption rates have been relatively poor, limiting visibility to their use and overall impact on health [3,4]. In fact, it is difficult to sort through all the existing eHealth tools to find evidence-based solutions for user needs. Indeed, an increasing number of studies on the subject have shown mixed results, and currently, there is a certain disillusionment regarding its benefits [5-8].

In recent years, agencies and governments, such as the World Health Organization (WHO) or the US Federal and Drug Administration, have elaborated guidelines and regulations in order to address the rather anarchic growth of eHealth technologies (eHT) [9-11]. However, it remains essential to conduct proper evaluations in order to assess overall eHT impact and benefits.

In this context, community-based health promotion (CBHP) represents an interesting and efficient conceptual framework, which could help increase the adoption of digital health solutions and promote greater opportunities for their evaluation [12,13]. While CBHP had been poorly theorized until very recently, it can be considered to have two defining characteristics. First, it is a community-based approach involving both professionals and local actors that emphasizes the holistic, preventative, and population levels rather than the pathogenic, curative, and individual levels. Importantly, it recognizes the social and organizational contexts in which people live, work, and interact [13,14]. Second, it relies "heavily on locally available channels of mass communication…with the potential to reach and change lifestyle behaviors of entire populations" [15]. Thus, through large-scale promotions, CBHP could allow eHT to reach more users, facilitating broader adoption and use within the population and thereby enhancing the level of evaluation results.

Since the successful development of the WHO Healthy Cities project [16,17], cities have been considered as one of the most suitable levels upon which CBHP can be established [18]. Therefore, environments in which CBHP is carried out should also be taken into full consideration. Further, it has been well-recognized that health status is influenced by social, economic, and environmental factors [19]. Thus, at the collective level, the availability of financial resources and health services, along with previous experience and expertise in implementing community-based interventions, can have a major impact on how CBHP is delivered [20-22]. Meanwhile, at the individual level, socioeconomic status (SES) and educational levels play a key role in how CBHP is accepted by the population. In digital health, particularly disparities in access and utilization patterns, notably across geographic and socioeconomic groups, have been highlighted [23]. Several studies have raised the risk of increasing health inequalities through the so-called "digital divide" [24-28].

The Case of StopBlues: a French Self-Help Tool to Prevent Mental Distress and Suicide

StopBlues (SB) is a first-of-its-kind website and mobile application in France that aims at preventing mental distress and suicide [29]. It was originally created in 2018 within a CBHP intervention, carried out in a sample of municipalities and associations of municipalities (referred to in this paper as "localities") where the tool was specifically promoted [29]. The idea was to anchor SB in local settings by offering the localities a ready-to-use eHealth tool, along with a promotional toolkit that they could customize and adapt to local needs. Similarly, they were invited to take part in the elaboration of the mental health resource locator included in the SB tool.

In each locality, an appointed delegate acted as the main point of contact with the research team and centralized the implementation of promotional actions locally. Two pre-experimentation meetings gathered the delegates to present the intervention and the SB tool and provide them with some suggestions regarding the implementation of promotion. Subsequently, the choice of promotional tools, the identification of public places and digital spaces for reaching out to the general population, as well as the launch date and the length of the promotion, were left to the discretion of the localities (Multimedia Appendix 2). This adaptable design added flexibility to the rigidity of standardized scientific experimentations [30,31]. Importantly, SB was included in a full evaluation program, in contrast to the limited research conducted via controlled trials or in real-world settings to evaluate the public health impact of eHT and mobile health applications, in particular [32,33]. The evaluation included a cluster-randomized, controlled trial (CRCT) [29] associated with qualitative research in order to better understand the implementation and identify the optimal conditions to ensure its utilization.

Objective

The objective of this paper was to analyze and understand how the implementation of the promotion varied according to the characteristics of different localities and how those differences influenced the utilization rate of SB. To this end, we assessed whether there were distinct profiles of localities based on their characteristics and promotion implementation patterns and whether those profiles were associated with a specific usage intensity of SB.

Methods

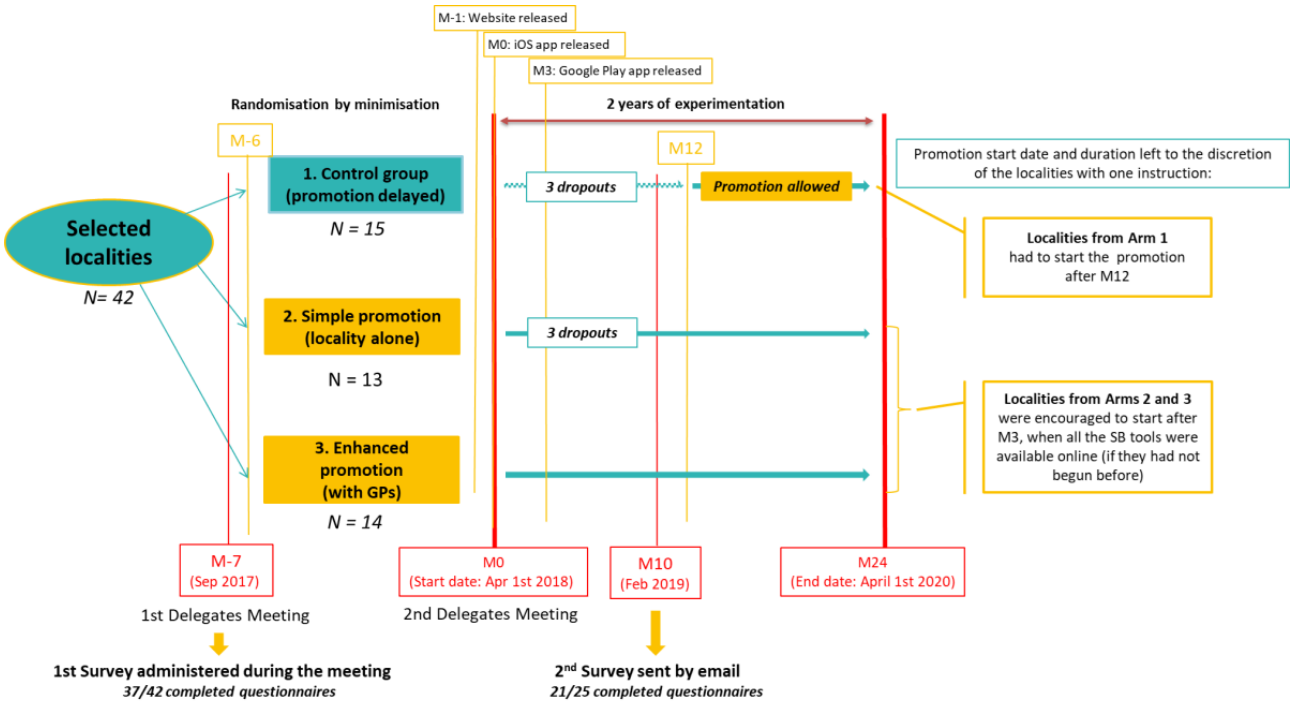
Setting

A three-arm, parallel-group CRCT was conducted to evaluate the effectiveness of the promotion with before and after observation [29]. 42 localities volunteered to be randomly allocated to one of the following three arms with a ratio of 1:1:1 (Figure 1; Multimedia Appendix 2):

1. Arm 1: “the control group” with no promotion (n=15). The localities included in this group started the promotion of SB with a one-year lag.
2. Arm 2: “the simple promotion group” with promotion by the locality only (n=13).
3. Arm 3: “the enhanced promotion group” with promotion by the locality and through general practitioner (GP) waiting rooms (n=14).

In total, 27 localities from Arms 2 and 3 that implemented the promotion of SB directly were included in the present analysis.

Figure 1. The intervention timelines. GP: general practitioner; SB: StopBlues.



Multiple Correspondence Analysis

In order to identify the localities’ profiles and find a potential association with the utilization rates of SB, a multiple correspondence analysis (MCA) was run using contextual and implementation characteristics of the 27 localities included in the CRCT. An MCA is a descriptive and exploratory statistical technique that is often used to help with the organization and classification of large amounts of data [34-36]. Notably, it allows researchers to deal with the complexity of qualitative data without impoverishing the richness of the reality [37-39].

In the current study, the MCA technique was also selected for the purpose of detecting all the possible patterns of relationship among the considered variables through a geometric approach. Indeed, each variable (unit of analysis) is located as a point in a low-dimensional Euclidian space [36,37,40].

Characteristics of the Localities Considered

All the characteristics included in the analysis were based on a review of the theoretical and empirical literature. Data collection included socioeconomic and demographic data from institutional sources, observations and discussions from the delegates' meetings, questionnaires, and web data extractions from SB analytic tools via Google Analytics (GA). The characteristics were divided into three sections:

1. The characteristics that were based on the pre-existing context.
2. The characteristics that derived from the program itself, its promotion, and implementation.
3. The utilization rate corresponds to the outcome variable of the promotion.

For the purpose of the MCA, these characteristics, detailed below, were transformed into a set of categorical variables. The characteristics from sections (1) and (2) were used as active variables that helped generate the MCA. The utilization rate was not used in the construction of the MCA but was added as a supplementary variable. The modalities of each variable are presented in Tables 1-3.

Table 1. Characteristics of the pre-existing context (context-based characteristics).

Variables and modalities (names in multiple correspondence analysis [MCA] if different)	Description	Frequency, n (%)
Level of urbanization and socioeconomic context of the area		
Rural area & low socioeconomic status (SES)	Rural localities with lower levels of SES	6 (27.3)
Urban area & low SES	Urban localities with lower levels of SES	8 (36.4)
Urban area & high SES	Urban localities with higher levels of SES	8 (36.4)
Local government interest and commitment to health matters		
Presence of a local health contract (LHC)		
No LHC	No LHC	7 (31.8)
LHC	LHC present	15 (68.2)
Presence and internal structure of a local mental health council (LMHC)		
Not having an LMHC (No LMHC)	No LMHC created at the time of the intervention	7 (31.8)
Unstructured LMHC	LMHC classified as unstructured	4 (18.2)
Structured LMHC	LMHC classified as structured	11 (50)
Experience in mental health project/policy (EXPInMH)		
No (No EXPInMH)	No experience	7 (31.8)
Yes (EXPInMH)	Experience in conducting projects in mental health	15 (68.2)

Table 2. Characteristics of the promotion and its implementation (promotion-related characteristics).

Variables and modalities (names in MCA if different)	Description	Frequency, n (%)
Delegates degree of understanding of and experience with mental health project as health professionals and coordinators of local mental health council (LMHC)		
Delegate was a health professional		
No (No health prof)	Delegate was not a health professional	10 (45.5)
Yes (health prof)	Delegate was a health professional	12 (54.5)
Delegate was the coordinator of the LMHC		
No (no coordinator LMHC)	Delegate was not the coordinator of an LMHC	11 (50)
Yes (coordinator LMHC)	Delegate was the coordinator of the LMHC	11 (50)
Provision of additional resources		
No (no added resources)	No additional resources were provided	11 (50)
Yes (added resources)	Promotion at the local level was provided with extra financial or human resources	11 (50)
Set-up of an ad-hoc working group		
No (no working group)	Delegate worked mainly alone	13 (59.1)
Yes (working group)	Ad-hoc group created locally to help with the promotion implementation	9 (40.9)
Involvement of general practitioners (GP) (promotion arm)		
Promotion arm including GP (Promo GP)	“Enhanced promotion group” with promotion by the locality and through GP waiting rooms	13 (59.1)
Promotion by the locality only (Promo locality)	“Simple promotion group” with promotion by the locality only	9 (40.9)
Intensity and type of promotion		
Small & traditional promotion	The number of promotional actions was below 5 and only traditional means were used	6 (27.3)
Large & traditional promotion	The number of promotional actions was above 5 and only traditional means were used	11 (50)
Large & digital promotion	The number of promotional actions was above 5 and digital means (websites, social media) were used	5 (22.7)

Table 3. The outcome variable of the promotion: Utilization rate of StopBlues.

Utilization rate of StopBlues (utilization rate) and its modalities	Description	Frequency, n (%)
Low	Rate < 25 per 100,000 residents or number of active users < 10 (for rate above 25 per 100,000 residents)	6 (27.3)
Medium	Rate ≥ 25 and < 50 per 100,000 residents or number of active users < 20 (for rate above 50 per 100,000 residents).	9 (40.9)
High	Rate ≥ 50 per 100,000 residents and number of active users ≥ 45.	7 (31.8)

Characteristics of the Pre-existing Context

The Level of Urbanization and Socioeconomic Context Of The Area

The geographic area and SES of the localities could have an impact on the utilization rate of SB. As a result, two characteristics were considered: the French Deprivation Index (FDep) and the urban unit, defined by the National Institute for Statistics and Economic Studies as a continuously built-up area with a minimum population of 2000 residents [41-45]. Localities were assigned a low or high SES, based on their FDep quintile: localities in quintiles 1 and 2 (least disadvantaged) were considered high SES, and those in quintiles 3 to 5 were considered low SES. Based on the combination of those two variables, the localities were grouped into three modalities: a rural area with low SES, an urban area with low SES, and an urban area with high SES (only one rural locality, in which the principal place of promotion was an urban unit, was characterized by high SES).

Considering the small number of localities, the number of categorical variables had to be limited and kept as low as possible. The goal was to ensure that the MCA did not lead to misinterpretation with the presence of rare variable modalities being disproportionately weighted in the model [37,46].

The below characteristics were collected through a questionnaire that was distributed to the delegates during the first pre-experimentation meeting and sent out by email to those who were absent (Multimedia Appendix 3).

Local Government Interest and Commitment to Health Matters

The Presence of a Local Health Contract

Local health contracts (LHCs) were created in 2009 as a roadmap with the objective of reducing health inequalities within a given territory by associating all the relevant local health actors around local governments and the regional health agency (RHA). The latter is the public administrative body of the French State responsible for implementing health policies at the regional level [47,48]. As a result, the presence of an LHC in a locality was considered a good indicator of local government interest and commitment to health matters.

The Presence and Internal Structure of a Local Mental Health Council

Local mental health councils (LMHCs), similarly to LHCs, bring together local health and social actors in order to tackle inequalities in the field of mental health by developing and implementing public policies [49]. One of the main goals of LMHCs is to improve mental health through community actions. Therefore, we hypothesized that their presence would have a positive impact on the implementation of the promotion [50].

However, there are no formal requirements on how these structures should run their activities. In fact, where present, there was great variability in the level of their respective internal structure. Thus, a distinction had to be made between "structured" LMHCs, ones that had a roadmap, translated into concrete actions and annual targets displayed in annual reports, and "unstructured" LMHCs, without specific actions and targets. Again, in order to keep the number of analyzed variables low, the presence and internal structure of LMHCs were merged into a single variable. Localities were categorized as not having an LMHC, having an unstructured LMHC, and having a structured LMHC.

The Experience in Mental Health Project/Policy

We also considered whether or not the locality had previous experience with conducting projects in mental health, particularly within the past six months [51]. Indeed, the existing literature indicates that the implementation of CBHPs requires a wide range of skills, notably in communication and management. These core competencies and capacities can be built and strengthened through multiple experiences [20].

The Delegates Degree of Understanding of and Experience With Mental Health Projects, as Health Professionals and Coordinators of LMHC

Two characteristics concerning the delegate profiles were also considered in the analysis: whether they were health professionals and, where possible, whether they were coordinators of the LMHC. The hypothesis behind this was that the delegates who received their initial education in health and those who were directly involved in the design and implementation of mental health programs would have more experience in conducting mental health projects. We also assumed that they would benefit from a larger network in the field [22,52].

Characteristics and Implementation of the Promotion

The next four characteristics were based on data collected through a web-based questionnaire that was sent to all the delegates (Figure 1; Multimedia Appendix 4).

The Provision of Additional Resources

We presumed that the availability of resources is an important factor that can influence the outcome of an intervention [21,22]. This characteristic therefore comprised all additional resources—human or financial—provided by local governments or RHAs for the implementation of the promotion.

The Set-Up of an Ad-Hoc Working Group

No compulsory guidelines for the implementation were given; however, interestingly, some localities put in place dedicated working groups. We hypothesized that the creation of a working group was indicative of long-term assimilation of best practices on collaborative processes, including the ability to co-construct projects with extended networks of partnerships [21,53]. We were interested to see if the presence of working groups could be associated with higher utilization rates.

The Involvement of GPs

Because previous studies have shown that the involvement of GPs could positively influence the outcome of an intervention with a focus on primary prevention [16,17,54], the participation of GPs in the promotion was introduced in the analysis. This was dependent on the promotion arm of the trial to which the localities were assigned.

The Intensity and Type of Promotion

We hypothesized that the effectiveness of a promotion would depend on the number of promotional actions developed: the more actions put in place, the more effective the promotion would be. However, effectiveness could also depend on the variety of these actions. As a result, we postulated that the use of digital channels (localities website, social media, etc) for promotion would have a positive impact on effectiveness because it targeted the users of eHealth tools directly.

The modalities were built according to the number and type of promotional actions put in place. Therefore, if there were fewer than five actions, the promotion was classified as "small," and if there were five or more actions, the promotion was classified as "large." Regarding the type of actions put in place, when conventional and paper-based promotion materials, such as

flyers, posters, and leaflets were used, the promotion was classified as "traditional." Meanwhile, when three and more digital means (websites, newsletters, social media, etc.) were used to promote SB, the promotion was classified as "digital."

Because no localities with fewer than five promotional actions (small) used three or more digital means, the localities were divided into three categories as follows: small and traditional, large and traditional, and large and digital.

The Outcome of the Promotion: Utilization Rate of Stopblues

We assessed the effectiveness of the promotion of SB through SB usage frequency using GA during the two-year experimentation period. Indeed, previous studies have shown that information from GA can be used to evaluate promotional campaigns, determine the geographic distributions of users, and analyze their use of online tools [55-57]. Data were extracted at month 24 directly from GA. This provided general data regarding the users: number (active and new) and location (country and city).

For the purpose of the analysis, the utilization rate variable was categorized as follows: low (< 25 per 100,000 residents or number of active users < 10), medium (between ≥ 25 and < 50 per 100,000 residents or number of active users < 20), and high (≥ 50 per 100,000 residents and number of active users ≥ 45).

Statistical Analyses

The MCA was performed using all the characteristics of the pre-existing context and those that derived from the program itself (the promotion and its implementation) as active variables. The latter contributed to the construction of a multiple-dimensional coordinate system. Only the first two dimensions were considered. All the variables and individuals were then displayed as plot points in the resulting two-dimensional coordinate system. The outcome variable, the utilization rate, was added to the MCA as a supplementary variable to distinguish whether groups of localities and their implementation characteristics were associated with different levels of utilization rate. Finally, the strength of the correlation between the variables was obtained by running a Pearson correlation test, using the x- and y-dimensional axis scores of the variable modalities produced by the MCA (the categorical variables were transformed into continuous ones). Only correlations above 0.50 ($|r|\geq 0.50$) (moderate positive or negative correlation) were considered meaningful for the purpose of the analysis [58], and the statistical significance value was set at $P<.05$. All statistical analyses were performed using packages: *FactoMineR* (version 2.3), *Factoshiny* (version 2.3), *Factoextra* (version 1.0.6), and *Ade4* (version 1.7-16) [59-62] from the software R version 3.6.3 (R Foundation).

Ethical Considerations

The study was granted ethical approval by the relevant ethics committees: the French National Institute for Health and Medical Research (INSERM; approval 15-240 on July 7, 2015), the French Advisory Committee for Data Processing in Health Research (approval 15-793 on September 30, 2015), and the French Data Protection Authority (decision DR-2016-421 on November 3, 2016) [29].

At the user level, the research, its purpose, and outcomes were described in an introductory section, and users had to acknowledge their participation in the intervention by signing a written informed consent. They were also informed that they could withdraw their consent at any time. At the locality level, all local governments signed a convention with the INSERM promoter.

Results

Multiple Correspondence Analysis: General Features

The analysis was performed on 22 localities out of the 27 included in the two arms of the CRCT promoting SB: 9/13 (69.2%) and 13/14 (92.9%) in the simple and the enhanced promotion groups, respectively. Three localities dropped out, and two failed to provide data regarding the characteristics necessary for the analysis. Among those localities that dropped out and did not complete the study, two abandoned the study before the beginning of the trial, citing political difficulties and excessive workload, while the third locality terminated its participation when the appointed delegate left her position and was not replaced.

The total inertia of the MCA model was equal to 1.3. The first two dimensions accounted for 43.1% of the cumulative projected inertia: 23.6% of projected inertia (inertia = 0.307/1.3) for the first dimension (dimension 1) and 19.5% (inertia = 0.254/1.3) for the second (dimension 2) and displayed distinct groups (Figure 2). These dimensions were therefore considered the most relevant for the analysis.

The contributions of the variables for the first two dimensions are presented in Table 4. The closer the value is to 1, the more the variable contributes to the definition of the dimension (1 being the maximum value). The values above the inertia are considered high and meaningful.

Figure 2. Biplot of the explored variable modalities and typology of the localities.

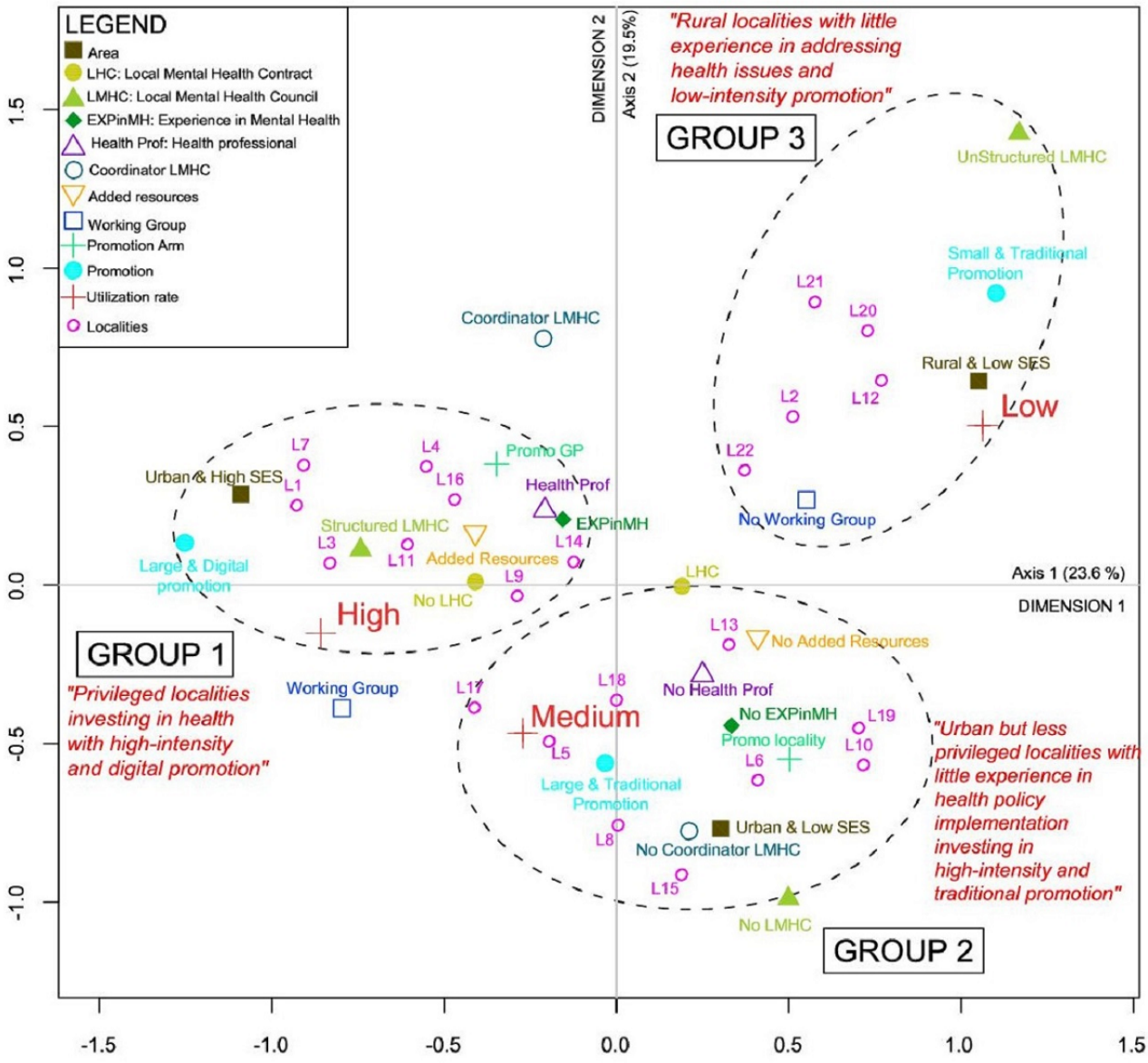


Table 4. Contribution of active variables.

	Dimension 1 (axis 1/x-axis)		Dimension 2 (axis 2/y-axis)		Mean
Name of the variables	Discrimination	Contribution (%)	Discrimination	Contribution (%)	
Level of urbanization and socioeconomic context of the area	*0.766*[a]	24.976	*0.357*[a]	14.061	0.562
Presence of a local health contract (LHC)	0.078	2.543	0.000	0.000	0.039
Presence and internal structure of a local mental health council (LMHC)	*0.603*[a]	19.661	*0.686*[a]	27.019	0.645
Experience in mental health project/policy	0.052	1.695	0.092	3.623	0.072
Delegate was a health professional	0.052	1.695	0.067	2.639	0.060
Delegate was the coordinator of the local mental health council	0.045	1.467	*0.604*[a]	23.789	0.325
Provision of additional resources	0.168	5.478	0.027	1.063	0.098
Set-up of an ad-hoc working group	*0.440*[a]	14.346	0.104	4.096	0.272
Involvement of GPs	0.175	5.706	0.210	8.271	0.193
Type and intensity of promotion	*0.688*[a]	22.432	*0.392*[a]	15.439	0.540
Total	3.067	100.000	2.539	100.000	2.803
% of projected inertia	23.580		19.530		21.550
Inertia	0.307		0.254		

[a]The italicized values are considered high and meaningful (above the inertia).

Three Different Profiles of Localities and Their Respective Promotion Implementation

The MCA identifies three distinct groups of localities characterized by different implementation profiles (Figure 2).

In the MCA plot, the statistical strength of the relationship between the variable modalities is determined by their spatial proximity and their location within the quadrants. The closer the variables are to each other, the stronger the relationship among the localities sharing these features. At the same time, the greater the distance of a modality from the intersection of the axes, the stronger its significance in the interpretation of results.

Four variables specifically helped shape the groups (Tables 4 and 5): the level of urbanization and socioeconomic context of the area in which the localities were situated; the presence and internal structure of LMHCs; the intensity and type of promotion implemented; and the set-up of an ad hoc working group.

Table 5. Correlations between meaningful variables in both dimensions.

Variables	Level of urbanization and socioeconomic context of the area	Presence and internal structure of a local mental health council	Delegate was the coordinator of the local mental health council	Set-up of an ad-hoc working group	Type and intensity of the promotion implemented in the localities	SB utilization rate
Level of urbanization and socioeconomic context of the area	1					
Presence and internal structure of a local mental health council	Dimension 1: r=0.626 P=.002; dimension 2: r=0.291 P=.02					
Delegate was the coordinator of the local mental health council	Dimension 1: r=0.289 P=.22; dimension 2: r=0.073 P=.14	Dimension 1: r=0.061 P=.82; dimension 2: r=0.711 P<.001				
Set-up of an ad-hoc working group	Dimension 1: r=0.448 P=.29; dimension 2: r=0.252 P=.26	Dimension 1: r=0.499 P=.02; dimension 2: – r=0.027 P=.76	Dimension 1: – r=0.092 P=.68; dimension 2: r=0.092 P=.68			
Type and intensity of the promotion implemented in the localities	Dimension 1: r=0.651 P=.001;dimension 2: r=0.216 P=0.33	Dimension 1: r=0.674 P<.001; dimension 2: r=0.659 P=.002	Dimension 1: – r=0.057 P=.83; dimension 2: r=0.265 P=.22	Dimension 1: r=0.712 P<.001; dimension 2: r=0.235 P=.60		
SB utilization rate	Dimension 1: r=0.790 P<.001; dimension 2: r=0.436 P=.04	Dimension 1: r=0.650 P<.001; dimension 2: r=0.321 P=.27	Dimension 1: r=0.000 P=1.0; dimension 2: r=0.000 P=1.0	Dimension 1: r=0.660 P<.001; dimension 2: r=0.534 P=.01	Dimension 1: r=0.651 P<.001; dimension 2: r=0.676 P=.003	1

Privileged Localities Investing in Health With High-Intensity and Digital Promotion

A first group (Group 1 in Figure 2) located in the upper left quadrant is comprised of the eight most urbanized and wealthiest localities (L1, L3, L4, L7, L9, L11, L14, and L16).

These localities were characterized by strong supportive environments with solid networks of local health service organizations, such as LMHCs, and significant experience in mental health project management and policy implementation. Moreover, the delegates were mostly health professionals and coordinated the LMHCs.

The promotion pattern of SB in those localities was characterized by the investment of additional resources—financial, material, or human—and a working group aimed at coordinating the local implementation. They also included digital strategies in the promotion. Besides, they tended to more frequently involve GP waiting rooms as promotion channels for SB.

Urban but Less Privileged Localities With Little Experience in Health Policy Implementation Investing in High-Intensity and Traditional Promotion

The second group (see Group 2 in Figure 2), located in the lower quadrants of the plot, includes ten localities (L5, L6, L8, L10, L13, L15, L16, L17, L18, and L19).

These were mostly situated in an urban area with low SES and had no experience in conducting projects in mental health, indicating that the environment was less supportive. This was illustrated notably by the absence of an LMHC. Additionally, in most cases, the appointed delegates from these localities were not health professionals. Yet, they still managed to put in place a large-scale traditional promotion, backed by the set-up of a working group.

Rural Localities With Little Experience in Addressing Health Issues and Low-Intensity Promotion

This third group (see Group 3 in Figure 2), located on the upper right quadrant, is comprised of five localities (L2, L12, L20, L21, and L22) mostly situated in rural areas with low SES levels.

Here, the environment was the least supportive, as illustrated by the fact that their LMHC was not structured with a clear roadmap. The localities also lacked experience in conducting mental health projects, and their appointed delegates were not health professionals.

In this group, the promotion pattern of SB was characterized by no investment of additional resources. They implemented a low-intensity promotion, defined by a few actions put in place, and did not create a working group for the purpose of the promotion.

Association Between Localities Groups and Utilization Rate of SB

The addition of the utilization rate to the MCA as a supplementary variable displayed a strong relationship between the profiles of the localities and the utilization of SB (Figure 2).

The privileged localities investing in health with high-intensity digital promotion were associated with the highest utilization rates. The urban, but less privileged localities, with little experience in health policy implementation and investing in high-intensity traditional promotion, were associated with medium utilization rates. Finally, rural localities with little experience in addressing health issues and low-intensity promotion were associated with the lowest utilization rates.

Besides, correlations between the meaningful variables, displayed in Table 5, were determined from the coordinate x- and y- dimensional axis scores. In dimension 1, the SB utilization rate was strongly correlated with the level of urbanization and socioeconomic context of the area, the LMHC (presence and internal structure), the set-up of an ad-hoc working group, and the type and intensity of the promotion implemented in the localities. In dimension 2, the SB utilization rate was also strongly correlated with the type and intensity of the promotion and, to a lesser extent, with the set-up of an ad-hoc working group.

Discussion

Principal Findings

Three distinct promotion patterns emerged, according to the profiles of localities, that were associated with the utilization rates of SB. These patterns were identified in the MCA, where, given the small number of localities included in the analysis, only correlations above 0.50 ($|r|\geq 0.50$) were considered significant. From highest to lowest utilization rates, the promotion patterns are listed as follows:

The urban "privileged localities investing in health" that implemented high-intensity promotion with a digital component demonstrated a greater capacity to take ownership of the project, understand its ins and outs, and anticipate potential challenges.

The "urban but less privileged localities" that, despite having relatively little experience in health policy implementation, managed to implement a high-intensity, traditional promotion.

The "rural localities with little experience in addressing health issues" that implemented low-intensity promotion could not overcome the challenges associated with their context-based characteristics (low SES, rural setting, and unstructured LMHCs, in particular).

Overall, contextual characteristics related to local government interest and commitment to health matters and the experience in mental health policy implementation were decisive to the success of the intervention and utilization of SB. Localities that were deeply involved in the structuring of health programs and with greater experience managed to implement promotional campaigns that resulted in higher utilization rates. The presence of dedicated local health service organizations to help with the implementation of public health interventions was crucial, although not sufficient to guarantee their success. Indeed, they also needed to be structured, with a roadmap that was translated into concrete actions and annual targets. As previously noted, this shows that the collaboration between local partners represents a fundamental parameter to tackle public health issues within a community [63]. Hence, dedicated structures, like LHCs and LMHCs, could not handle problems when pre-existing shortcomings, such as lack of resources or local partners, were not addressed in the first place [48].

Localities where working groups were implemented had higher SB utilization rates, indicating the importance of being able to not only mobilize existing partners but also organize the implementation processes of an intervention. Stronger networks and a history of partnerships between local stakeholders could more naturally lead to a greater willingness to work together and, thus, collaborate in working groups [21,53,63].

Finally, the involvement of GPs in the promotion was globally associated with higher SB utilization rates. More precisely, 75% of the localities characterized by high utilization rates were included in the enhanced promotion group that involved GPs. This is consistent with previous literature findings regarding the important role of primary care in the prevention of suicide [16,17,54].

Among the findings related to the program itself, the type of promotion was particularly noteworthy. The implementation of a digital promotion was associated with the highest SB utilization rates. Other researchers have already pointed out the utility of digital marketing and, more specifically, the use of social media to promote health [23,64,65]. In effect, it seems logical to use the internet to promote eHealth tools. This is notably true when considering eHT in mental health care. Indeed, because of the stigma associated with mental illness, people with poor mental health are more likely to use the internet to find information and possible solutions to their problems [66-68]. Likewise, social media have many advantages such as cost-effectiveness, 24/7 availability, and potential broad audience reach. They also allow for interactive engagement [69,70].

In addition, those who support the utilization of eHT in mental health have previously acknowledged the benefits of employing digital channels for user engagement [71,72]. However, notwithstanding the growing body of literature on social media, the evaluation of the effectiveness of their utilization in health is lacking [73,74]. In fact, to our knowledge, this is the first

study in a French setting that confirms the potential of social media for the promotion of eHT.

Another important factor to be considered is the population targeted by the promotion. Regarding the sociodemographic context of the implementation, we notably found that the more deprived and rural the localities were, the lower the utilization rates of SB. Previous studies have highlighted similar results, where promotion interventions were less effective among populations with lower SES [27,28,75]. Two reasons for this can be put forward: First, these populations generally lack access to health information and do not have the ability to interpret and process the information [27]. This explains why the segments of a population with the lowest incomes tend to be less likely to benefit from preventive care. Second, they are less likely to use eHealth tools, such as mobile health apps, either because of lack of adequate access or lack of skills and training [1,76,77].

Several studies have stressed that the effectiveness of health promotion interventions varied between rural and urban settings [78-80]. Indeed, rural communities face many specific challenges related to the difficulty of accessing a range of health care services and an increasing shortage of health professionals [71,78]. However, to have a deeper understanding of the differences between rural and urban areas in terms of SB utilization rates, three factors should be carefully examined. Firstly, rural localities tend to cover larger and more remote areas; therefore, the implementation of a homogenous and uniform promotion is more challenging [71,79,81]. Secondly, while digital health promotion could be a solution to engage and reach the rural populations [71], these areas are still characterized by poorer broadband access, which can limit the adoption of eHT [82,83]. Thirdly, rural populations tend to be older than urban ones [84,85]. Previous studies have indicated that older adults have more limited access to technology in general and, consequently, tend to possess poorer digital health literacy skills than younger age groups [86-88].

Thus, lower digital literacy levels are amplified through the joint effect of SES and age in rural localities. This supports the idea raised by Frohlich and Potvin that population-level approaches to eHT should be complemented with interventions focusing directly on vulnerable populations in order to alleviate the digital divide [89].

Limitations

Limitations in this study are primarily due to the sample size of localities, the data used, and the method of the MCAs.

The analysis was conducted on a sample size of only 22 localities. While no rule exists regarding the minimum number of observations needed to run an MCA, small sample sizes affect the reliability and interpretation of the results [38,46]. However, Di Franco [46] indicates that a sufficient sample consists of 20 observations per single active categorical variable, and CRCTs that involve less than 30 clusters [90,91] are also commonly used in the assessment of interventions in prevention.

Other limitations of this study are related to the data used to perform the MCA. The first is associated with how the delegates responded to the self-report questionnaires that were used to collect data on the type and intensity of promotion that they implemented in their locality. Possible classical information biases can arise, such as social desirability or recall biases. Social desirability bias is the tendency to overreport more socially desirable attributes and behaviors and underreport attitudes that are perceived as socially undesirable [92]. In recall bias, the study participants do not remember their previous actions related to the question that was asked [93].

A second limitation involves the use of GA to obtain the user statistics. Because GA models are calculated at the country level, Google states that metrics are not always accurate, “particularly for campaigns that target small geographical areas, such as a single city or zip code” [94]. However, one can assume that this lack of accuracy is evenly distributed across the arms of the trial.

Regarding the MCA method itself, only correlations $|r|\geq 0.50$ were considered significant, whereas the significance threshold for the coefficient correlation in exploratory and MCAs can be lowered to $|r|\geq 0.30$ and still be meaningful [95,96]. However, considering the small number of localities included in our analysis, we decided to keep the threshold at $|r|\geq 0.50$ in order to avoid any misinterpretation [97].

On a more general level, the benefits of the MCA utilization to transform qualitative data into quantitative data can also be its pitfall. Indeed, by simplifying the contained data, we can also lose the complexity of information and not fully understand the ins and outs of the situation [38,98]. This is why further analysis based on a more qualitative approach should be carried out in order to complement the methodology and deepen the findings.

Conclusions

The use and dissemination of the StopBlues eHealth tool depended heavily on the promotion that was conducted.

The urban and SES profiles of localities, along with their investment and pre-existing experience in health, appear to be critical for shaping the implementation of the promotion of the SB eHealth tool in terms of intensity and use of digital communication. The more digital channels used, the higher the utilization rates, ultimately leading to the overall success of the intervention.

Digital communication, and more specifically, social media, seem to be powerful tools for engaging hard-to-reach populations. This raises the question of the information and communication technology skills and digital literacy, along with the attitudes towards new technologies, of the professionals in charge of the promotion of eHT. They should be trained in order to be receptive and proactive in the use of eHT to the benefit of the greatest number of people.

Further, to broaden the outreach, other innovative means and promotional strategies should be considered to reinforce and back up promotion campaigns of eHT initiated at the local level. New web-based marketing techniques, where the user could be informed of useful and targeted interventions at the very moment they are “engaging in information-seeking behavior through online search inquiry” [99], could open up entirely new perspectives in the way public health interventions are built.

This is particularly the case in the field of mental health, where online information-seeking behaviors are widespread due notably to the fear of stigma [66,67]. However, further research is needed to understand how to avoid eHT becoming a source of social and health inequalities, such that it may be leveraged for social good.

Acknowledgments

The authors gratefully acknowledge the financial support of the French National Public Health Agency, which also provided the funding for the PRINTEMPS project (which this work is a part of) through a grant from the 2014 prevention call for research proposals of the French Research Institute in Public Health (grant 026/14). Additional funding was received from the European Commission through the Interreg North-West Europe. The authors also wish to express their gratitude to all the people and their organizations who participated in the conduct of the project, notably all the delegates from the localities included in the project, the French Ministries of Health and of Research, the French National Health Insurance Fund for Employees, the Arc Foundation for Cancer Research, the French National Cancer Institute, and INSERM. The authors would like to thank Clément Picot-Ngo and Jean-Baptiste Hazo for their help on the data acquisition and their overall assistance to the PRINTEMPS project. KT extends her thanks to Morgane Michel, Amélie Prigent, Peter Sobolewski, Pauline Guezennec, and Moreno Ursino, who generously gave their time and effort in support of this work. The funding bodies had no involvement in the conception and design of this work, the analysis and interpretation of data, nor the initial draft and revision of the manuscript.

Authors' Contributions

KC and AD made substantial contributions to the conception and the supervision of the work; KT participated in the design and conception of the work, contributed to the acquisition and interpretation of data, conducted the statistical analysis, and drafted the original manuscript with the support of AD; KC designed the methodology, supervised the statistical analysis and interpretation of data and contributed to the drafting of the original manuscript; and all authors critically reviewed the manuscript and approved the final submitted version. The Printemps Consortium includes Corinne Alberti, Philippe Courtet, Coralie Gandré, Bruno Giraudeau, Anaïs Le Jeannic, Jean-Luc Roelandt, Guillaume Vaiva, and Marie-Amélie Vinet.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Members of the PRINTEMPS Consortium.
[DOCX File , 13 KB - jmir_v24i4e30218_app1.docx]

Multimedia Appendix 2
Detailed characteristics of the StopBlues promotion.
[DOCX File , 19 KB - jmir_v24i4e30218_app2.docx]

Multimedia Appendix 3
The questionnaire was administered to the attendees of the first meeting at M-7 and sent by email to those who were absent. Its goal was to assess the local situation (local health system, presence of health service organization, past experience) and identify the local point person for the promotion.
[DOCX File , 15 KB - jmir_v24i4e30218_app3.docx]

Multimedia Appendix 4
The web-based questionnaire was designed and sent by email at M10 to all the delegates. The aim was to investigate how the promotion was set up and conducted in the different settings.
[DOCX File , 15 KB - jmir_v24i4e30218_app4.docx]

References

1. Jacobs RJ, Lou JQ, Ownby RL, Caballero J. A systematic review of eHealth interventions to improve health literacy. Health Informatics J 2016 Jun 10;22(2):81-98 [FREE Full text] [doi: 10.1177/1460458214534092] [Medline: 24916567]
2. Ahern DK, Kreslake JM, Phalen JM. What is eHealth (6): perspectives on the evolution of eHealth research. J Med Internet Res 2006 Mar 31;8(1):e4 [FREE Full text] [doi: 10.2196/jmir.8.1.e4] [Medline: 16585029]
3. Carlo AD, Hosseini Ghomi R, Renn BN, Areán PA. By the numbers: ratings and utilization of behavioral health mobile applications. NPJ Digit Med 2019 Jun 17;2(1):54 [FREE Full text] [doi: 10.1038/s41746-019-0129-6] [Medline: 31304400]

4. Baumel A, Muench F, Edan S, Kane JM. Objective User Engagement With Mental Health Apps: Systematic Search and Panel-Based Usage Analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
5. Albrecht U, von Jan U. Safe, sound and desirable: development of mHealth apps under the stress of rapid life cycles. Mhealth 2017;3:27 [FREE Full text] [doi: 10.21037/mhealth.2017.06.05] [Medline: 28828374]
6. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht U. mHealth 2.0: Experiences, Possibilities, and Perspectives. JMIR Mhealth Uhealth 2014 May 16;2(2):e24 [FREE Full text] [doi: 10.2196/mhealth.3328] [Medline: 25099752]
7. Almario CV. The Effect of Digital Health Technology on Patient Care and Research. Gastroenterol Hepatol (N Y) 2017 Jul;13(7):437-439 [FREE Full text] [Medline: 28867974]
8. Gorwood P. L'apport du numérique en santé mentale: beaucoup d'espoir, quelques inquiétudes. Soins Psychiatrie 2017 Jul;38(311):12-15. [doi: 10.1016/j.spsy.2017.04.006]
9. Classification of Digital Health Interventions v 1. World Health Organization. 2018. URL: https://apps.who.int/iris/bitstream/handle/10665/260480/WHO-RHR-18.06-eng.pdf [accessed 2021-10-08]
10. Digital Health Software ¨Pre-certification (Pre-Cert) Program. U.S. Food &Drug Administration. URL: https://www.fda.gov/medical-devices/digital-health-center-excellence/digital-health-software-precertification-pre-cert-program [accessed 2020-10-08]
11. Evidence standards framework for digital health technologies. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies [accessed 2020-10-08]
12. Petticrew M. When are complex interventions 'complex'? When are simple interventions 'simple'? The European Journal of Public Health 2011 Jul 18;21(4):397-398. [doi: 10.1093/eurpub/ckr084]
13. Jolley G. Evaluating complex community-based health promotion: Addressing the challenges. Evaluation and Program Planning 2014 Aug;45:71-81. [doi: 10.1016/j.evalprogplan.2014.03.006]
14. Poland B, Frohlich K, Cargo M. Context as a Fundamental Dimension of Health Promotion Program Evaluation. In: Potvin L, McQueen DV, Hall M, de Salazar L, Anderson LM, Hartz ZMA. editors. Health Promotion Evaluation Practices in the Americas Internet New York, NY: Springer New York; 2008.
15. Farquhar J, Fortmann S. Community-based Health Promotion. In: Ahrens W, Pigeot I, editors. Handbook of Epidemiology. 1st edition. Heidelberg: Springer; 2005.
16. Hegerl U, Wittenburg L, Arensman E, Van Audenhove C, Coyne JC, McDaid D, et al. Optimizing suicide prevention programs and their implementation in Europe (OSPI Europe): an evidence-based multi-level approach. BMC Public Health 2009 Nov 23;9:428 [FREE Full text] [doi: 10.1186/1471-2458-9-428] [Medline: 19930638]
17. Székely A, Konkolÿ Thege B, Mergl R, Birkás E, Rózsa S, Purebl G, et al. How to decrease suicide rates in both genders? An effectiveness study of a community-based intervention (EAAD). PLoS One 2013;8(9):e75081 [FREE Full text] [doi: 10.1371/journal.pone.0075081] [Medline: 24086443]
18. Ashton J, Grey P, Barnard K. Healthy cities — WHO's New Public Health initiative. Health Promot Int 1986;1(3):319-324. [doi: 10.1093/heapro/1.3.319]
19. Lawlor E, Cupples M, Donnelly M, Tully M. Promoting physical activity among community groups of older women in socio-economically disadvantaged areas: randomised feasibility study. Trials 2019 Apr 25;20(1):234 [FREE Full text] [doi: 10.1186/s13063-019-3312-9] [Medline: 31023329]
20. Jané-Llopis E, Barry MM. What makes mental health promotion effective? Promot Educ 2005 Nov 02;Suppl 2(2_suppl):47-55, 64, 70. [doi: 10.1177/10253823050120020108] [Medline: 15966253]
21. Bulthuis S, Kok M, Raven J, Dieleman M. Factors influencing the scale-up of public health interventions in low- and middle-income countries: a qualitative systematic literature review. Health Policy Plan 2020 Mar 01;35(2):219-234 [FREE Full text] [doi: 10.1093/heapol/czz140] [Medline: 31722382]
22. Granja C, Janssen W, Johansen MA. Factors Determining the Success and Failure of eHealth Interventions: Systematic Review of the Literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
23. Chou WS, Prestin A, Lyons C, Wen K. Web 2.0 for Health Promotion: Reviewing the Current Evidence. Am J Public Health 2013 Jan;103(1):e9-e18. [doi: 10.2105/ajph.2012.301071]
24. McAuley A. Digital health interventions: widening access or widening inequalities? Public Health 2014 Dec;128(12):1118-1120. [doi: 10.1016/j.puhe.2014.10.008] [Medline: 25458115]
25. Rowsell A, Muller I, Murray E, Little P, Byrne CD, Ganahl K, et al. Views of People With High and Low Levels of Health Literacy About a Digital Intervention to Promote Physical Activity for Diabetes: A Qualitative Study in Five Countries. J Med Internet Res 2015 Oct 12;17(10):e230 [FREE Full text] [doi: 10.2196/jmir.4999] [Medline: 26459743]
26. Teuscher D, Bukman AJ, van Baak MA, Feskens EJM, Renes RJ, Meershoek A. A lifestyle intervention study targeting individuals with low socioeconomic status of different ethnic origins: important aspects for successful implementation. BMC Public Health 2017 Jul 25;18(1):54 [FREE Full text] [doi: 10.1186/s12889-017-4592-1] [Medline: 28743281]

27. Jang K, Baek YM. How to effectively design public health interventions: Implications from the interaction effects between socioeconomic status and health locus of control beliefs on healthy dietary behaviours among US adults. Health Soc Care Community 2018 Apr 16;26(5):664-674. [doi: 10.1111/hsc.12577] [Medline: 29659076]
28. Adler NE, Newman K. Socioeconomic disparities in health: pathways and policies. Health Aff (Millwood) 2002;21(2):60-76. [doi: 10.1377/hlthaff.21.2.60] [Medline: 11900187]
29. Gandré C, Le Jeannic A, Vinet M, Turmaine K, Courtet P, Roelandt J, et al. The PRINTEMPS study: protocol of a cluster-randomized controlled trial of the local promotion of a smartphone application and associated website for the prevention of suicidal behaviors in the adult general population in France. Trials 2020 Jun 22;21(1):553 [FREE Full text] [doi: 10.1186/s13063-020-04464-2] [Medline: 32571432]
30. Glasgow RE, Lichtenstein E, Marcus AC. Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. Am J Public Health 2003 Aug;93(8):1261-1267. [doi: 10.2105/ajph.93.8.1261] [Medline: 12893608]
31. Proctor E, Luke D, Calhoun A, McMillen C, Brownson R, McCrary S, et al. Sustainability of evidence-based healthcare: research agenda, methodological advances, and infrastructure support. Implementation Sci 2015 Jun 11;10:88 [FREE Full text] [doi: 10.1186/s13012-015-0274-5] [Medline: 26062907]
32. Freudenberg N. Assessing the Public Health Impact of the mHealth App Business. Am J Public Health 2017 Nov;107(11):1694-1696. [doi: 10.2105/AJPH.2017.304083] [Medline: 29019765]
33. Singh K, Drouin K, Newmark LP, Lee J, Faxvaag A, Rozenblum R, et al. Many Mobile Health Apps Target High-Need, High-Cost Populations, But Gaps Remain. Health Affairs 2016 Dec;35(12):2310-2318. [doi: 10.1377/hlthaff.2016.0578] [Medline: 27920321]
34. Barusch A, Gringeri C, George M. Rigor in Qualitative Social Work Research: A Review of Strategies Used in Published Articles. Social Work Research 2011 Mar 01;35(1):11-19. [doi: 10.1093/swr/35.1.11]
35. Cohen DJ, Crabtree BF. Evaluative criteria for qualitative research in health care: controversies and recommendations. Ann Fam Med 2008;6(4):331-339 [FREE Full text] [doi: 10.1370/afm.818] [Medline: 18626033]
36. Escofier P, Pagès J. Analyses factorielles simples et multiples. Objectifs, méthodes et interprétation. Paris: Dunod; 2008.
37. Le Lan R. Analyse de données et classification sur données d'enquête. Journées de Méthodologie Statistique-Institut national de la statistique et des études économiques. 2005. URL: http://jms-insee.fr/2005/S12_2_ACTE_LE-LAN_JMS2005.pdf [accessed 2020-05-24]
38. Macia L. Using Clustering as a Tool: Mixed Methods in Qualitative Data Analysis. TQR 2015 Jul 20;20(7):1083-1094 [FREE Full text] [doi: 10.46743/2160-3715/2015.2201]
39. Cibois P. La Représentation factorielle des tableaux croisés et des données d'enquête : étude de méthodologie sociologique. Paris: Université Paris V René Descartes; 1980.
40. Greenacre MJ. Correspondence Analysis in Practice. New York: Chapman and Hall/CRC; Dec 01, 2016.
41. Rey G, Jougla E, Fouillet A, Hémon D. Ecological association between a deprivation index and mortality in France over the period 1997 - 2001: variations with spatial scale, degree of urbanicity, age, gender and cause of death. BMC Public Health 2009 Jan 22;9:33 [FREE Full text] [doi: 10.1186/1471-2458-9-33] [Medline: 19161613]
42. Temam S, Varraso R, Pornet C, Sanchez M, Affret A, Jacquemin B, et al. Ability of ecological deprivation indices to measure social inequalities in a French cohort. BMC Public Health 2017 Dec 15;17(1):956 [FREE Full text] [doi: 10.1186/s12889-017-4967-3] [Medline: 29246202]
43. Ghosn W. Note méthodologique : Indicateurs écologiques synthétiques du niveau socio-économique pour la recherche en Santé. Le CépiDc. 2018. URL: https://cepidc.inserm.fr/sites/default/files/2019-02/Note_indices_socioeco.pdf [accessed 2020-06-05]
44. Gorza M, Eilstein D. Outils élaborés dans la cadre du programme «Inégalités sociales de santé?»-2015. Santé Publique France. 2018. URL: https://www.santepubliquefrance.fr/docs/outils-elabores-dans-la-cadre-du-programme-inegalites-sociales-de-sante-2013-2015-fiches-variables-socioeconomiques-fiches-indices-de [accessed 2020-11-05]
45. Clanché F, Rascol O. Le découpage en unités urbaines de 2010. L'espace urbain augmente de 19 % en une décennie. Institut national de la statistique et des études économiques. 2011. URL: https://www.insee.fr/fr/statistiques/1280970 [accessed 2020-07-23]
46. Di Franco G. Multiple correspondence analysis: one only or several techniques? Qual Quant 2015 Apr 21;50(3):1299-1315. [doi: 10.1007/s11135-015-0206-0]
47. Contrats locaux de santé: Agir ensemble pour la santé des citoyens au cœur des territoires. Agence Régionale de Santé. 2012. URL: https://solidarites-sante.gouv.fr/IMG/pdf/pacte_territoire_sante_-_plaquette_-_contrats_locaux_sante.pdf [accessed 2020-07-23]
48. Cros R. La mesure de la plus-value des CLS mis en oeuvre en Bretagne. Ecole des hautes études en santé publique. 2014. URL: https://documentation.ehesp.fr/memoires/2014/iass/cros.pdf [accessed 2020-07-17]
49. Guézennec P, Roelandt J. Les conseils locaux de santé mentale en France : état des lieux en 2015. L'information psychiatrique 2015;91(7):549-556 [FREE Full text] [doi: 10.1684/ipe.2015.1372]

50. Bouhier G, Cadou-Plante B, Gallard R, Gouriff M, Gaston A, Niyomwungere C, et al. Évaluation du déploiement de la démarche d'évaluation des conseils Locaux en santé mentale. CLSM-CCOMS. 2018. URL: http://clsm-ccoms.org/wp-content/uploads/2018/02/UE09_CLSM_EHESP.pdf [accessed 2020-10-01]
51. Röthlin F, Schmied H, Dietscher C. Organizational capacities for health promotion implementation: results from an international hospital study. Health Promot Int 2015 Jun 19;30(2):369-379. [doi: 10.1093/heapro/dat048] [Medline: 23872386]
52. Santos C, Santos V, Tavares A, Varajão J. Project Management Success in Health – The Need of Additional Research in Public Health Projects. Procedia Technology 2014;16:1080-1085. [doi: 10.1016/j.protcy.2014.10.122]
53. Dako-Gyeke P, Asampong E, Afari E, Launois P, Ackumey M, Opoku-Mensah K, et al. Capacity building for implementation research: a methodology for advancing health research and practice. Health Res Policy Syst 2020 Jun 01;18(1):53 [FREE Full text] [doi: 10.1186/s12961-020-00568-y] [Medline: 32487176]
54. van der Feltz-Cornelis CM, Sarchiapone M, Postuvan V, Volker D, Roskar S, Grum AT, et al. Best practice elements of multilevel suicide prevention strategies: a review of systematic reviews. Crisis 2011 Nov 01;32(6):319-333 [FREE Full text] [doi: 10.1027/0227-5910/a000109] [Medline: 21945840]
55. Crutzen R, Roosjen JL, Poelman J. Using Google Analytics as a process evaluation method for Internet-delivered interventions: an example on sexual health. Health Promot Int 2013 Mar;28(1):36-42. [doi: 10.1093/heapro/das008] [Medline: 22377974]
56. Gordon EJ, Shand J, Black A. Google analytics of a pilot mass and social media campaign targeting Hispanics about living kidney donation. Internet Interventions 2016 Nov;6:40-49. [doi: 10.1016/j.invent.2016.09.002]
57. Clark D, Nicholas D, Jamali HR. Evaluating information seeking and use in the changing virtual world: the emerging role of Google Analytics. Learn. Pub 2014 Jul 01;27(3):185-194. [doi: 10.1087/20140304] [Medline: 20140304]
58. Mukaka MM. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. Malawi Med J 2012 Sep;24(3):69-71 [FREE Full text] [Medline: 23638278]
59. Lê S, Josse J, Husson F. FactoMineR: Package for Multivariate Analysis v 2.3. J Stat Soft 2008;25(1):1-18. [doi: 10.18637/jss.v025.i01]
60. Vaissie P, Monge A, Husson F. Factoshiny: Perform Factorial Analysis from 'FactoMineR' with a Shiny Application. 2021. URL: https://cran.r-project.org/web/packages/Factoshiny/index.html [accessed 2021-09-02]
61. Kassambara A, Mundt F. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. 2020. URL: https://cran.r-project.org/web/packages/factoextra/index.html [accessed 2021-09-02]
62. Dray S, Dufour A. The ade4 Package: Implementing the Duality Diagram for Ecologists v1.7-1. J Stat Soft 2007;22(4):1-20. [doi: 10.18637/jss.v022.i04]
63. Leviton LC, Snell E, McGinnis M. Urban issues in health promotion strategies. Am J Public Health 2000 Jun;90(6):863-866. [doi: 10.2105/ajph.90.6.863] [Medline: 10846502]
64. Lupton D. Health promotion in the digital era: a critical commentary. Health Promot Int 2015 Mar;30(1):174-183. [doi: 10.1093/heapro/dau091] [Medline: 25320120]
65. Jane M, Hagger M, Foster J, Ho S, Pal S. Social media for health promotion and weight management: a critical debate. BMC Public Health 2018 Jul 28;18(1):932 [FREE Full text] [doi: 10.1186/s12889-018-5837-3] [Medline: 30055592]
66. Powell J, Clarke A. Internet information-seeking in mental health: population survey. Br J Psychiatry 2006 Sep;189:273-277 [FREE Full text] [doi: 10.1192/bjp.bp.105.017319] [Medline: 16946364]
67. Gowen LK. Online Mental Health Information Seeking in Young Adults with Mental Health Challenges. Journal of Technology in Human Services 2013 Apr;31(2):97-111. [doi: 10.1080/15228835.2013.765533]
68. Neal DM, Campbell AJ, Williams LY, Liu Y, Nussbaumer D. “I did not realize so many options are available”: Cognitive authority, emerging adults, and e-mental health. Library & Information Science Research 2011 Jan;33(1):25-33. [doi: 10.1016/j.lisr.2010.07.015]
69. Korda H, Itani Z. Harnessing social media for health promotion and behavior change. Health Promot Pract 2013 Jan;14(1):15-23. [doi: 10.1177/1524839911405850] [Medline: 21558472]
70. Graham JE, Moore JL, Bell RC, Miller T. Digital Marketing to Promote Healthy Weight Gain Among Pregnant Women in Alberta: An Implementation Study. J Med Internet Res 2019 Feb 01;21(2):e11534 [FREE Full text] [doi: 10.2196/11534] [Medline: 30707100]
71. Mehmet M, Roberts R, Nayeem T. Using digital and social media for health promotion: A social marketing approach for addressing co‐morbid physical and mental health. Aust. J. Rural Health 2020 Jan 22;28(2):149-158. [doi: 10.1111/ajr.12589]
72. Simpson S, Reid C. Telepsychology in Australia: 2020 vision. Aust J Rural Health 2014 Dec 10;22(6):306-309. [doi: 10.1111/ajr.12103]
73. Freeman B, Potente S, Rock V, McIver J. Social media campaigns that make a difference: what can public health learn from the corporate sector and other social change marketers? Public Health Res Pract 2015 Mar 30;25(2):e2521517 [FREE Full text] [doi: 10.17061/phrp2521517] [Medline: 25848735]
74. Park A, Bowling J, Shaw G, Li C, Chen S. Adopting Social Media for Improving Health: Opportunities and Challenges. N C Med J 2019;80(4):240-243 [FREE Full text] [doi: 10.18043/ncm.80.4.240] [Medline: 31278188]

75. Hiscock R, Bauld L, Amos A, Fidler J, Munafò M. Socioeconomic status and smoking: a review. Ann N Y Acad Sci 2012 Feb;1248:107-123. [doi: 10.1111/j.1749-6632.2011.06202.x] [Medline: 22092035]
76. Laing SS, Alsayid M, Ocampo C, Baugh S. Mobile Health Technology Knowledge and Practices Among Patients of Safety-Net Health Systems in Washington State and Washington, DC. J Patient Cent Res Rev 2018;5(3):204-217 [FREE Full text] [doi: 10.17294/2330-0698.1622] [Medline: 31414005]
77. Levy H, Janke AT, Langa KM. Health literacy and the digital divide among older Americans. J Gen Intern Med 2015 Mar;30(3):284-289 [FREE Full text] [doi: 10.1007/s11606-014-3069-5] [Medline: 25387437]
78. White M, Adams J, Heywood P. How and why do interventions that increase health overall widen inequalities within populations? In: Babones S, editor. Social Inequality and Public Health. 1st edition. Bristol: Policy Press; 2009:65-82.
79. Bolin JN, Bellamy GR, Ferdinand AO, Vuong AM, Kash BA, Schulze A, et al. Rural Healthy People 2020: New Decade, Same Challenges. J Rural Health 2015;31(3):326-333. [doi: 10.1111/jrh.12116] [Medline: 25953431]
80. Umstattd Meyer MR, Perry CK, Sumrall JC, Patterson MS, Walsh SM, Clendennen SC, et al. Physical Activity-Related Policy and Environmental Strategies to Prevent Obesity in Rural Communities: A Systematic Review of the Literature, 2002-2013. Prev Chronic Dis 2016 Jan 07;13:E03 [FREE Full text] [doi: 10.5888/pcd13.150406] [Medline: 26741997]
81. White D. Development of a rural health framework: implications for program service planning and delivery. Healthc Policy 2013 Feb;8(3):27-41 [FREE Full text] [Medline: 23968625]
82. Legleye S, Rolland A. Une personne sur six n'utilise pas Internet, plus d'un usager sur trois manque de compétences numériques de base. Institut national de la statistique et des études économiques. 2019. URL: https://www.insee.fr/fr/statistiques/4241397 [accessed 2020-12-15]
83. Vogels E. Some digital divides persist between rural, urban and suburban America. Pew Research Center. 2021. This is an update of an analysis by Andrew Perrin originally accessed on 2020-12-15 URL: https://www.pewresearch.org/fact-tank/2021/08/19/some-digital-divides-persist-between-rural-urban-and-suburban-america/
84. Sabau C, Blasquiet-Revol H, Lenain MA. Les personnes âgées dans les territoires ruraux : une double opportunité pour la création d'activités et d'emplois de service. Gérontologie et société 2010;33(135):239-258 [FREE Full text] [doi: 10.3917/gs.135.0239]
85. Wenger GC. Myths and realities of ageing in rural Britain. Ageing and Society 2001 May 18;21(1):117-130. [doi: 10.1017/s0144686x01008042]
86. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006 Jun;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
87. Xie B. Effects of an eHealth literacy intervention for older adults. J Med Internet Res 2011 Nov;13(4):e90 [FREE Full text] [doi: 10.2196/jmir.1880] [Medline: 22052161]
88. 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(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
89. Frohlich KL, Potvin L. Transcending the known in public health practice: the inequality paradox: the population approach and vulnerable populations. Am J Public Health 2008 Feb;98(2):216-221. [doi: 10.2105/AJPH.2007.114777] [Medline: 18172133]
90. Leyrat C, Morgan K, Leurent B, Kahan B. Cluster randomized trials with a small number of clusters: which analyses should be used? Int J Epidemiol 2018 Feb 01;47(1):321-331. [doi: 10.1093/ije/dyx169] [Medline: 29025158]
91. Eldridge SM, Ashby D, Feder GS, Rudnicka AR, Ukoumunne OC. Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care. Clin Trials 2004 Feb;1(1):80-90. [doi: 10.1191/1740774504cn006rr] [Medline: 16281464]
92. Latkin CA, Edwards C, Davey-Rothwell MA, Tobin KE. The relationship between social desirability bias and self-reports of health, substance use, and social network factors among urban substance users in Baltimore, Maryland. Addict Behav 2017 Oct;73:133-136 [FREE Full text] [doi: 10.1016/j.addbeh.2017.05.005] [Medline: 28511097]
93. Coughlin SS. Recall bias in epidemiologic studies. J Clin Epidemiol 1990;43(1):87-91. [doi: 10.1016/0895-4356(90)90060-3] [Medline: 2319285]
94. Target geographical locations for delivery. Google Ad Manager Help. URL: https://support.google.com/admanager/answer/1260290 [accessed 2020-07-26]
95. Costa PS, Santos NC, Cunha P, Cotter J, Sousa N. The Use of Multiple Correspondence Analysis to Explore Associations between Categories of Qualitative Variables in Healthy Ageing. J Aging Res 2013;2013:302163 [FREE Full text] [doi: 10.1155/2013/302163] [Medline: 24222852]
96. Xiao C, Ye J, Esteves RM, Rong C. Using Spearman's correlation coefficients for exploratory data analysis on big dataset. Concurrency Computat: Pract Exper 2015 Dec 18;28(14):3866-3878. [doi: 10.1002/cpe.3745]
97. Krehbiel TC. Correlation Coefficient Rule of Thumb. Decision Sciences Journal of Innovative Education 2004 Jan 16;2(1):97-100. [doi: 10.1111/j.0011-7315.2004.00025.x]
98. Henry D, Dymnicki AB, Mohatt N, Allen J, Kelly JG. Clustering Methods with Qualitative Data: a Mixed-Methods Approach for Prevention Research with Small Samples. Prev Sci 2015 Oct;16(7):1007-1016 [FREE Full text] [doi: 10.1007/s11121-015-0561-z] [Medline: 25946969]

99. Birnbaum ML, Garrett C, Baumel A, Scovel M, Rizvi AF, Muscat W, et al. Using Digital Media Advertising in Early Psychosis Intervention. Psychiatr Serv 2017 Jul 17:appips201600571. [doi: 10.1176/appi.ps.201600571] [Medline: 28712355]

Abbreviations

CBHP: community-based health promotion
CRCT: cluster randomized controlled trial
eHT: eHealth technologies
FDep: French Deprivation Index
GA: Google Analytics
GP: general practitioner
INSERM: French National Institute of Health and Medical Research
LHC: local health contract
LMHC: local mental health council
MCA: multiple correspondence analysis
RHA: regional health agency
SB: StopBlues
SES: socioeconomic status
WHO: World Health Organization

Edited by R Kukafka; submitted 06.05.21; peer-reviewed by V Nagy, J Andrews; comments to author 30.07.21; revised version received 23.09.21; accepted 29.10.21; published 22.04.22.

Please cite as:
Turmaine K, Dumas A, Chevreul K, PRINTEMPS Consortium
Conditions for the Successful Integration of an eHealth Tool "StopBlues" Into Community-Based Interventions in France: Results From a Multiple Correspondence Analysis
J Med Internet Res 2022;24(4):e30218
URL: https://www.jmir.org/2022/4/e30218
doi:10.2196/30218
PMID:35451977

Original Paper

Internet-Delivered Cognitive Behavioral Therapy for Insomnia Comorbid With Chronic Pain: Randomized Controlled Trial

Tobias Wiklund[1], PhD; Peter Molander[1,2], PhD; Philip Lindner[3], PhD; Gerhard Andersson[2,3,4], PhD; Björn Gerdle[1], PhD; Elena Dragioti[1], PhD

[1]Pain and Rehabilitation Centre, and Department of Health, Medicine and Caring Sciences, Linköping University, Linkoping, Sweden

[2]Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden

[3]Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institute, & Stockholm Health Care Services, Region Stockholm, Stockholm, Sweden

[4]Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden

Corresponding Author:
Elena Dragioti, PhD
Pain and Rehabilitation Centre, and Department of Health, Medicine and Caring Sciences, Linköping University
Brigadgatan 22
Linkoping, 581 85
Sweden
Phone: 46 763251361
Email: elena.dragioti@liu.se

Abstract

Background: Patients with chronic pain often experience insomnia symptoms. Pain initiates, maintains, and exacerbates insomnia symptoms, and vice versa, indicating a complex situation with an additional burden for these patients. Hence, the evaluation of insomnia-related interventions for patients with chronic pain is important.

Objective: This randomized controlled trial examined the effectiveness of internet-based cognitive behavioral therapy for insomnia (ICBT-i) for reducing insomnia severity and other sleep- and pain-related parameters in patients with chronic pain. Participants were recruited from the Swedish Quality Registry for Pain Rehabilitation.

Methods: We included 54 patients (mean age 49.3, SD 12.3 years) who were randomly assigned to the ICBT-i condition and 24 to an active control condition (applied relaxation). Both treatment conditions were delivered via the internet. The Insomnia Severity Index (ISI), a sleep diary, and a battery of anxiety, depression, and pain-related parameter measurements were assessed at baseline, after treatment, and at a 6-month follow-up (only ISI, anxiety, depression, and pain-related parameters). For the ISI and sleep diary, we also recorded weekly measurements during the 5-week treatment. Negative effects were also monitored and reported.

Results: Results showed a significant immediate interaction effect (time by treatment) on the ISI and other sleep parameters, namely, sleep efficiency, sleep onset latency, early morning awakenings, and wake time after sleep onset. Participants in the applied relaxation group reported no significant immediate improvements, but both groups exhibited a time effect for anxiety and depression at the 6-month follow-up. No significant improvements on pain-related parameters were found. At the 6-month follow-up, both the ICBT-i and applied relaxation groups had similar sleep parameters. For both treatment arms, increased stress was the most frequently reported negative effect.

Conclusions: In patients with chronic pain, brief ICBT-i leads to a more rapid decline in insomnia symptoms than does applied relaxation. As these results are unique, further research is needed to investigate the effect of ICBT-i on a larger sample size of people with chronic pain. Using both treatments might lead to an even better outcome in patients with comorbid insomnia and chronic pain.

Trial Registration: ClinicalTrials.gov NCT03425942; https://clinicaltrials.gov/ct2/show/NCT03425942

(J Med Internet Res 2022;24(4):e29258) doi:10.2196/29258

KEYWORDS

insomnia; chronic pain; comorbid; CBT-i; RCT; web-based CBT; pain; online health; online treatment; digital health; mental health; rehabilitation

Introduction

Background

Pain conditions are the leading cause of disability and disease burden in the world [1]. In Europe, 19% of the adult population experiences moderate to severe chronic pain [2]. Acute pain is a common symptom of other diseases; however, chronic pain is widely regarded a disease or diseases [3].

Patients with chronic pain often experience insomnia. However, depending on the measures and definitions used, the prevalence rates differ considerably. An earlier study by our group [4] showed that 65% of patients in a specialized pain clinic scored >14 points on the self-report Insomnia Severity Index (ISI), which corresponds to moderate or severe insomnia [5]. Several studies have shown that sleep disturbances increase the risk of developing chronic pain conditions over time [6-8], that poor sleep increases pain levels the next day, and that pain can lead to more insomnia [9-11]. A recent study demonstrated that patients with chronic pain with atypical bedtime habits (12:01 AM to 8:59 PM) experienced higher pain levels, greater activity interference, and higher consumption of prescribed opioids than those with regular bedtime habits (9 PM to midnight) [12]. These results suggest that a misalignment of the circadian rhythm can exacerbate chronic pain.

According to a 2-process model [13], sleep is governed by the circadian rhythm and a homeostatic sleep-wake process. When these processes interact optimally, sleep occurs when we have been awake long enough and it is night (ie, the circadian drive for wakefulness is low). Sleep restriction therapy is a common component of cognitive behavioral therapy for insomnia (CBT-i) and is hypothesized to target these 2 processes [14,15]. The American Academy of Sleep Medicine recommends using CBT-i as the first-line intervention not only for chronic insomnia but also for comorbid insomnia [16]. Several studies have shown that CBT-i is efficacious for treating insomnia co-occurring with chronic pain [17,18].

Typically, CBT-i consists of several components such as a sleep diary, stimulus control, sleep restriction, and advice regarding sleep hygiene [19]. Some treatment manuals also include relaxation training and cognitive interventions to cope with worry, maladaptive thoughts, and depressive symptoms [19]. In addition, cognitive interventions such as cognitive restructuring were not included because the knowledge about what maladaptive thoughts are specific to patients with chronic pain is lacking [20-22]. No consensus exists regarding the components needed for successful treatment; however, the use of stimulus control or sleep restriction predicts outcomes [23]. As stimulus control is based on classical conditioning, patients are told that the bedroom should only be used for sleep or sex to associate the bed with sleep rather than wakefulness. Patients are also instructed to leave the bedroom if they do not fall asleep within a predefined period, usually between 15 and 20 minutes.

Because access to therapists offering CBT-i is limited, internet-delivered treatments are a promising alternative [24]. Standardized internet treatments could potentially increase the availability of nonpharmacological insomnia treatment in primary care as well as in specialized pain clinics. Open-source and evidence-based treatments will improve care by addressing a common aspect of chronic pain, that is, insomnia symptoms. In addition, the implementation of internet-based cognitive behavioral therapy for insomnia (ICBT-i) has immense potential in terms of cost-effectiveness, as the time per patient is shorter than face-to-face CBT-i [25,26]. Previous work from members of our group on persons with insomnia disorder showed that CBT-i provided via the internet (ICBT-i) has comparable outcomes to cognitive behavioral therapy (CBT) provided in a group setting [27]. ICBT-i can take various forms with respect to treatment content and therapist support. Today, there are several studies on fully atomized treatments based on artificial intelligence [28,29]. It has been demonstrated that therapist support, usually via written messages, increases the effect of CBT-i [30]. However, there is no available data on ICBT-i targeting insomnia symptoms in patients with chronic pain.

Objectives

This randomized controlled trial (RCT) investigates the acceptability of an ICBT-i treatment and assesses whether ICBT-i is more effective than internet-administrated applied relaxation (active control condition) in reducing insomnia symptoms (as measured by the ISI) comorbid to chronic pain. In addition, this RCT investigates the effects of ICBT-i on sleep diary measures, pain intensity, anxiety, depression, pain-related disability, and perceived health. Negative effects were also monitored and reported. Our primary hypothesis is that compared with an active control condition, ICBT-i would lead to greater reductions in insomnia symptoms at the end of treatment. We also hypothesized that ICBT-i would be superior to an active control condition in terms of improvement in symptom-related sleep diary measures, pain intensity, anxiety, depression, pain-related disability, and perceived health. In addition, the improvements gained from treatment are hypothesized to be maintained for 6 months after treatment.

Methods

Design

Participants and Study Procedure

This randomized controlled parallel-group study was conducted in Sweden and followed the CONSORT (Consolidated Standards of Reporting Trials) eHealth Checklist (Multimedia Appendix 1). The trial was registered at ClinicalTrials.gov (NCT03425942). Participants were selected via the Swedish Quality Registry for Pain Rehabilitation, a survey distributed during all first visits at the Pain and Rehabilitation Centre, Linköping University Hospital, Sweden. The Swedish Quality Registry for Pain Rehabilitation contains data on pain (intensity, duration, and spreading), psychological strain, and the ISI. An estimated effect size of 0.6 and an α of .05, gave a sample of at least 40 participants per study condition [27,31]. An anticipated dropout of one-third resulted in the expected need for a total sample of 120 participants.

Participants with an ISI score of >14 (moderate to severe insomnia symptoms) and aged 18-65 years were asked to participate in the trial either by their physician or via postal

invitation. Retrospective analyses were performed to identify former patients with an ISI score of >14 over the previous 2 years. If interested, participants registered on the website and completed another survey covering anxiety and depressive symptoms, demographic and physiological variables, current insomnia symptoms, pain characteristics, pain-related disability, and perceived health. Again, a cutoff score of >14 points on the ISI was used as an inclusion criterion. This measurement was considered as baseline. Participants were also asked to rate their average pain intensity during the previous week on a numeric rating scale as part of the baseline assessment. Finally, they were asked, "For how long have you had your pain problem?" The following predefined answers were provided: (1) less than 3 months, (2) 3 months to 1 year, (3) 1 to 3 years, (4) 3 to 5 years, or (5) >5 years. Participants who gave responses (2) to (5) were included in the study. When this strategy had been applied, the study sample consisted of 54 participants, and the decision was made to end recruitment and analyze that smaller sample.

When informed consent was provided, a staff member called the participant to set up a telephone interview within 5 days from registration on the website. The interview included the Mini International Neuropsychiatric Interview [32] (Swedish translation version 7.0.0) to cover psychiatric comorbidity and to identify bipolar disorder and psychoses, which were the absolute exclusion criteria. In addition, the telephone interview was used to evaluate whether the participant fulfilled the diagnostic criteria for insomnia disorder according to the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (Table S1 in Multimedia Appendix 2) [33] and to assess other inclusion and exclusion criteria. These criteria are detailed in Table S2 in Multimedia Appendix 2. All cases were evaluated by an inclusion and exclusion committee consisting of 2 experienced psychologists and 1 psychiatrist, when required.

Additional telephone interviews were conducted after treatment and after 6 months to evaluate the fulfillment of the criteria for insomnia disorder (Table S2 in Multimedia Appendix 2).

Randomization and Blinding

Included participants were randomized via random.org in blocks of 34 participants to either the ICBT-i group or the applied relaxation group. This procedure was performed to ensure equal numbers of participants in each arm after the first postal survey. Thereafter, participants were added to a randomized list consecutively (ensuring an equal sample size after 34 additional participants). Participants were blinded to the experiment or control and were told that they were assigned to 1 of the 2 treatments. Allocation was conducted after the inclusion and exclusion committee met, and participants were notified the week before treatment started. The therapists were not blinded.

Ethics Approval and Consent to Participate

The study was approved by the ethical review board in Linköping (dnr. 2014/191-31). An additional application to the ethical committee was submitted (dnr. 2017/511-32) to enable invitation of patients further back in time. The participants also provided informed consent as directed by the World Medical Association Declaration of Helsinki.

Interventions

The description of the activities per session for both interventions is provided in Table S3 in Multimedia Appendix 2. Both treatment arms were provided with weekly home assignments. The rationale for the interventions and information about sleep were provided via SMS text messaging or PowerPoint (Microsoft Corp) presentations with speaker voice; participants were free to choose modality based on preference. Treatments were always initiated on Mondays, as the participants were gaining access to a new treatment module. They were instructed to hand in their home assignments the next Sunday to receive feedback or therapist support within 48 hours. Handing in-home assignments was mandatory to gain access to the next treatment module, and no treatment lasted longer than the agreed 5 weeks.

Experimental Group

The ICBT-i treatment conducted is a novel treatment developed by TW and PM and provided via the internet platform iterapi.se [34]. The treatment is based on the most well-established CBT principles for the treatment of insomnia (ie, sleep restriction and stimulus control) [35,36]. The first week focused on a short treatment rationale and registration of current sleep patterns (sleep diary). On the basis of the sleep diary data, an individually designed sleep prescription (bedtime) was calculated and applied during the second week. The prescription was equal to the calculated average total sleep time, but no shorter than 4 hours a night. The prescribed sleep time was maintained for the rest of the treatment unless the sleep efficiency exceeded 85% or sunk below 75%; then, the prescribed sleep time was adjusted to 15 to 30 minutes a week. High sleep efficiency resulted in more time in bed and low sleep efficiency resulted in less time in bed (although not <4 hours/night). The third module taught stimulus control, which is based on classical conditioning. Participants were told to use the bedroom only for sleep and sex. That is, activities such as watching television, reading, and social media consumption were not to be conducted in the bedroom. In addition, the participants were told to go to bed only when sleepy and to get out of bed and leave the bedroom when unable to sleep. Similarly, the participants were told not to sleep in places other than the bedroom or outside the prescribed bedtimes. If stimulus control is implemented, it will restrict time in bed, similar to sleep restriction.

Week 4 was dedicated to daytime activity. Advice regarding activity balance was provided. This advice is based on the research conducted by Andrews et al [37], who found that patients with chronic pain who engage in irregular daytime activities experience poor sleep the same night. Daytime sleepiness can lead to inactivity; however, excessive activity can increase pain and hinder the ability to wind down before bedtime. The advice aims to facilitate a healthy level of daytime activity to promote sleep and create a contrast between daytime activities and nighttime rest. The last week focused on maintaining behavior changes and preventing relapse. Generally, the aim is to keep the intervention as brief as possible while maintaining the main part of the treatment effect. Therefore, some usual CBT-i components were excluded, such as

behavioral activation, scheduled worry time, and cognitive restructuring.

Applied Relaxation Control Condition

The internet-based control condition is a slightly modified version of the well-established applied relaxation techniques developed by Öst [38]. This method was chosen because it is a common treatment component that has a credible and applicable rationale for the treatment of both insomnia [39,40] and chronic pain [41,42]. Furthermore, applied relaxation is an active and rather time-consuming treatment that is supposed to control for time spent, measurement effects, and therapist support. The manual was adapted so that the length of treatment matched that of the experimental group (5 weeks). In addition to a short rationale and registration of current sleep patterns (sleep diary), the first week focused on progressive relaxation and diaphragmatic breathing. Participants were told to practice for 15 minutes twice a day and to keep a log for registration and evaluation. In the following week, time was shortened to 7 minutes twice a day. During the third week, participants were taught conditioned relaxation, and exercises were 2 to 3 minutes long. Differentiated relaxation aims only to activate the muscles required to perform a specific task (ie, other muscles can be relaxed). This was the focus of the fourth week of the study. During the fifth week, participants were taught quick relaxation. As this technique can result in relaxation in just a few seconds, it can be applied several times throughout the day and at bedtime.

Therapist Support

Both treatment arms had therapist support every week of treatment. Support was provided via written messages in the treatment platform. Therapists (master's students in psychology or senior psychologists) provided problem solving and feedback on weekly tasks and ensured the correct implementation of treatment components. Because therapist support is one of the factors that contribute to treatment outcome [30], there was no restriction of therapist support, as it did not include components from the other treatment arm. The master's students were supervised by senior psychologists trained in CBT (TW and PM). In a few cases, telephone calls were used to solve technical problems or to reach participants who did not respond to written messages. The same therapists provided both treatments, and the participants were distributed because of the randomization. Because the treatment content was standardized in both treatment arms, no measures of therapist fidelity were collected.

Demographic and Physiological Variables

Age, sex, height (in cm), weight (in kg), and educational level were recorded at baseline. The duration of pain and sleep problems were measured using two questions: How long have you had pain problems? How long have you had sleep problems? For pain problems, five predefined answers were provided: <3 months, 3 months to 1 year, 1 to 3 years, 3 to 5 years, and >5 years. For sleep problems, six predefined answers were provided: <1 month, 1 to 3 months, 3 months to 1 year, 1 to 3 years, 3 to 5 years, and >5 years. Finally, participants were asked about the debut of pain or sleep problems: the pain problems preceded the sleep problems, the sleep problems preceded the pain problems, both arose at the same time, and they did not know.

Primary Outcome: Insomnia Severity

The Swedish version of the ISI was used to quantify perceived insomnia severity. The ISI, a valid instrument with excellent internal consistency, captured the severity and impact of insomnia symptoms [5,43]. The psychometric properties of the Swedish version have also been evaluated in patients with chronic pain [44]. Cronbach α for the ISI in our sample was .67. The ISI is rated on a 5-point Likert scale (0-4) and has a maximum total score of 28. In this study, the ISI was measured at baseline, at the end of every week during treatment (T_1-T_5), and at the 6-month follow-up. Morin et al [5] suggested that a minimally important difference of >7 points be used in treatment studies.

Secondary Outcomes

Sleep Parameters Assessed by Responsive Sleep Diary

All participants were asked to complete a sleep diary throughout the treatment and for 1 week at the 6-month follow-up. These data were used to calculate the sleep-specific outcomes, that is, sleep onset latency (SOL), wake time after sleep onset (WASO), time in bed, sleep efficiency, and early morning awakenings (EMA), according to Buysse et al [45]. Every measure is based on the mean value form all registered nights that week. The patient interface presents the time in bed, total sleep time, and sleep efficiency for every night and the mean values for every week.

Sleepiness

Sleepiness was measured using the Karolinska Sleepiness Scale (KSS), which is a 9-graded scale that measures the level of sleepiness at a particular time during the day [46]. Every other step is labeled as follows: 1=very alert, 3=alert, 5=neither alert nor sleepy, 7=sleepy (but not fighting sleep), and 9=very sleepy (fighting sleep). The KSS is correlated with several electroencephalogram measures (eg, α-activity, r=0.40) and behavior variables (eg, mean reaction time on the psychomotor vigilance task, r=0.57) related to sleepiness. As test-retest reliability depends on the time of day, the KSS was sent to participants via SMS text messaging at 11 AM every day during treatment. If the participants did not answer the question within 60 minutes of receiving the SMS text message, the value was reported as missing. The KSS was added to the sleep diary to capture nonrestorative sleep and daytime symptoms.

Anxiety

The Generalized Anxiety Disorder–7-item (GAD-7) scale was used to measure anxiety symptoms [47]. The GAD-7 measures symptoms of generalized anxiety disorder but is also highly correlated with more general anxiety measures [47]. Each item has an answer response ranging from 0 to 3 and is scored with respect to the frequency of the symptom during the previous 2 weeks: not at all, several days, more than half the days, and nearly every day. This gives a total score of 0 to 21. Optimal sensitivity/specificity ratio for detecting generalized anxiety disorder is obtained with a cutoff score of ≥10. The GAD-7 has

shown excellent internal consistency (Cronbach α=.92) [47]. The Cronbach α for the GAD-7 in our sample was .88.

Depression

The Patient Health Questionnaire–9 items (PHQ-9) was used to measure depression [48].

The PHQ-9 is a 9-item self-rating scale reflecting the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, criteria for major depressive disorder [48]. Each item ranges from 0 to 3 points and is scored with respect to the frequency of the symptom during the last 2 weeks: not at all, several days, more than half the days, and nearly every day. This gives a total score of 0 to 27. The PHQ-9 has shown excellent internal consistency (Cronbach α=.86-.89) [48] and strong correlations with other well-established depression scales such as the Beck Depression Inventory II (r=0.84) and the Montgomery-Åsberg Depression Rating Scale (r=0.79) [49]. The Cronbach α for the PHQ-9 in our sample was .82. Optimal sensitivity/specificity ratio for detecting depression is obtained with a cutoff score of ≥10.

Pain Intensity

The participants were asked to rate their average pain intensity during the previous week on a numeric rating scale ranging from 0 (no pain) to 10 (worst imaginable pain). This scale has provided good validity in experimental conditions [50].

Number of Anatomical Pain Regions

The participants were presented with a list of 36 anatomical regions covering the entire body and were asked to mark all regions where they experienced pain. A total score index was calculated (total score=0-36) [51].

Pain Disability

The Pain Disability Index (PDI) was used to measure the self-reported pain-related disability [52]. The PDI has high internal consistency (Cronbach α=.85) in patients seeking specialized pain care [53]. Minimal important change depends on baseline scores [54] (eg, baseline values 28-42 require a decrease of at least 15 points). The Cronbach α for the PDI in our sample was .81.

Health-Related Quality of Life

The Health Visual Analog Scale from the European Quality of Life instrument was used to measure the current state of self-estimated health [55]. The item consists of a 100-point thermometer-like vertical scale with defined end points (worst imaginable health condition and best imaginable health condition). Higher values indicate better health, whereas lower values indicate worse health [55,56].

Potentially Adverse and Unwanted Events

The Negative Effects Questionnaire (NEQ), used to assess the side effects of psychological treatment, consists of 32 items with three parts [57]: an initial yes or no question regarding the occurrence of the negative effect in question and a rating of the impact of the negative effect using a 5-point scale (not at all to extremely). The patient also judges whether the negative effect was caused by the treatment. In this RCT, the NEQ was used to screen for and quantify the most common negative effects attributed to the treatment. There is no consensus in the scientific community on how to present results for negative effects, as definitions and measures vary. Generally, it is recommended to provide the frequencies of the negative effects that have occurred the most. The NEQ has good internal consistency (α=.95) and six factors: symptoms, quality, dependency, stigma, hopelessness, and failure [58].

The Cronbach α for the NEQ in our sample was .80.

Compliance

The number of treatment modules with submitted tasks or worksheets was reported as median values and their IQR. Modules with submitted tasks or worksheets were used as proxies for treatment compliance. This definition did not require full completion of all tasks or worksheets of a module, and any answer (in addition to a sleep diary) was interpreted as a result of working with the treatment content from that module.

Credibility

Credibility and expectations were measured several times to evaluate the design of the single-blinded control condition. At week 3, the participants answered the following question: to what extent do you think this treatment will be helpful in reducing your sleep problems? The participants answered the question on a 6-point scale with the following end points: the treatment will not help at all and the treatment will help to a large extent. At week 5 and follow-up, the participants answered the following question: how likely is it that you would recommend this treatment to a relative or friend with the same kind of problem? The participants answered the question on a 6-point scale with the following end points: not likely at all and very likely. At follow-up, the participants also answered the following question: to what extent do you think this treatment has been helpful in reducing your sleep problems? The participants answered the question on a 6-point scale with the following end points: the treatment has not helped at all and the treatment has helped to a large extent. The results were also reported as median values and IQR.

Statistics

The analysis was based on the intention-to-treat principles and the assumption that data were missing at random. Data processing and statistics were performed using the statistical package IBM SPSS Statistics (version 26.0; IBM Corp), the software R (version 4.0.3), and JAMOVI (version 1.2.27). In addition, 2-sided statistical tests were used, and a P<.05 was considered significant. Average weekly scores for the sleep diary parameters were calculated using R software. Minor corrections were applied to the sleep diary data to account for missing data and input errors: missing durations (but not timestamps) for otherwise filled entries were replaced with zeros; bed and sleep time were ordered so that the former always came before the latter; k=9 entries with negative sleep time (primarily due to mathematically impossible wake durations) were omitted; and k=8 entries obviously incorrectly inputted as timestamps were manually recoded as durations.

For descriptive analysis, we used mean values with SDs or median with IQR for continuous variables after normality testing

and number with percentage (n, %) for categorical variables. Whenever feasible, owing to the different time point measurements across outcomes, mean changes for both primary and secondary outcomes were applied before and after treatment and 6-month follow-up. For example, for the sleep diary and KSS measurements, preassessments were not available.

To examine the immediate (week 5) and long-term (6-month follow-up) treatment effects on the ISI and secondary outcomes, separate mixed models for repeated measures with treatment, time, and the interaction of treatment by time were performed with random intercept and random slope (whenever appropriate) using an unstructured covariance matrix. The examination of the variance of the slope for time was significant; therefore, a model with random slopes for time with multiple time points fitted the data, and a model with fixed slopes was more appropriate for time with 2 time points. Time was modeled as a numeric variable (0-4 and 0-1), and both linear and quadratic trajectories were considered; the latter was chosen when modeling revealed a significant trajectory. The Akaike information criterion was used to assess model fit: the smaller the Akaike information criterion value, the better the model fit [59]. All available time points were included in the models. On the basis of the principle that data were missing at random, all linear mixed models were run with restricted maximum likelihood estimation, which can produce unbiased estimators [60].

Results

Descriptive Results

Of the 677 former patients who received postal invitations, 188 (27.8%) participants were identified as eligible at their first visit. Of these 188 participants, 86 (45.7%) completed the web-based assessment, and 54 (28.7%) were included and randomized (Figure 1). Table 1 presents the characteristics of the sample. The mean age for the total sample was 49.3 (SD 12.3; range 21-67) years, 37% (20/54) had a university degree, and most participants were women (45/54, 83%). Until this point, no participant had been aware of their allocated intervention (ICBT-i or applied relaxation). Eventually, 30 participants were allocated to the ICBT-i group and 24 to the applied relaxation group. The mean age for the ICBT-i group was 48.2 (SD 11.1; range 27-66) years, and that for the applied relaxation group was 50.6 (SD 13.6; range 21-67) years. At the 6-month follow-up, 77% (23/30) of the participants in the ICBT-i group and 63% (15/24) of the participants in applied relaxation group (ie, the control group) responded.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) 2010 flow diagram. ICBT-i: internet-based cognitive behavioral therapy for insomnia; ISI: Insomnia Severity Index.

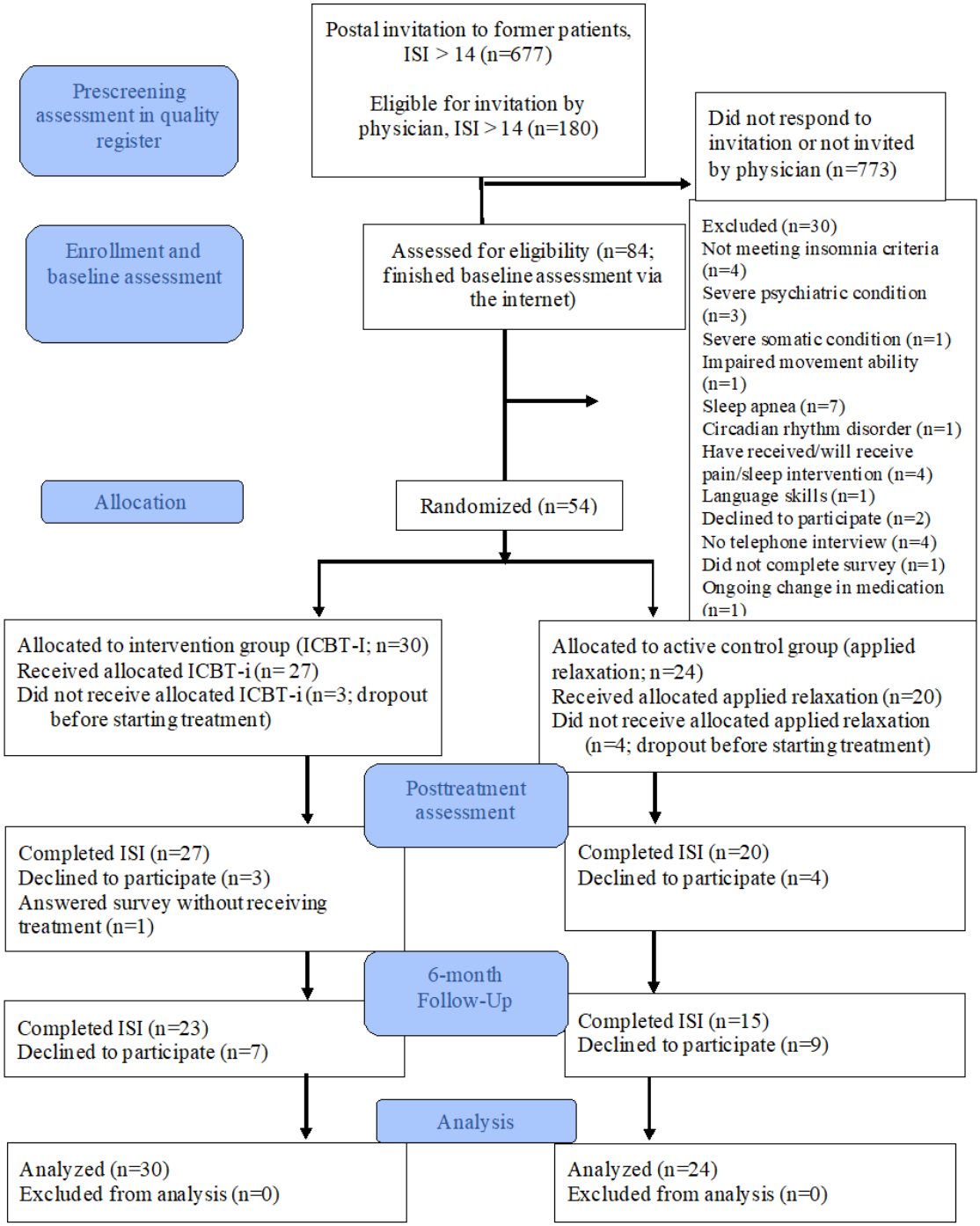


Table 1. Baseline characteristics of the study sample.

Characteristics	ICBT-i[a] (n=30)	AR[b] (n=24)	Total (N=54)
Age (years), mean (SD)	48.2 (11.1)	50.6 (13.6)	49.3 (12.3)
Female, n (%)	23 (77)	22 (92)	45 (83)
BMI, mean (SD)	27.0 (5.3)	24.5 (5.2)	25.9 (5.4)
University education, n (%)	12 (40)	8 (33)	20 (37)
ISI[c], mean (SD)	21.4 (3.3)	20.3 (2.8)	20.9 (3.1)
GAD-7[d], mean (SD)	7.4 (4.2)	8.3 (5.4)	7.8 (4.7)
PHQ-9[e], mean (SD)	13.5 (4.4)	14.4 (5.7)	13.9 (4.9)
Pain intensity NRS[f], mean (SD)	6.3 (1.6)	7.1 (1.8)	6.7 (1.7)
NPR[g], mean (SD)	13.6 (10.0)	15.7 (9.8)	14.5 (9.9)
PDI[h], mean (SD)	39.9 (11.6)	41.3 (10.3)	40.5 (10.9)
EQ5-VAS[i], mean (SD)	43.9 (17.1)	43.3 (17.7)	43.7 (17.2)
Duration of pain problems >5 years, n (%)	20 (67)	14 (58)	34 (63)
Duration of sleep problems >5 years, n (%)	23 (77)	19 (79)	42 (78)
The pain problem preceded the sleep problem, n (%)	18 (60)	14 (58)	32 (59)

[a]ICBT-i: internet-based cognitive behavioral therapy for insomnia.

[b]AR: applied relaxation.

[c]ISI: Insomnia Severity Index.

[d]GAD-7: Generalized Anxiety Disorder–7 items.

[e]PHQ-9: Patient Health Questionnaire–9 items.

[f]NRS: Numeric Rating Scale.

[g]NPR: number of pain regions.

[h]PDI: Pain Disability Index.

[i]EQ5-VAS: European Quality of Life 5-Dimension Visual Analog Scale.

Dropout Analysis

Comparisons between dropouts and completers revealed that participants who dropped out were more likely to be older, had lower education and pain disability, and had higher levels of depression and anxiety (Table S4 in Multimedia Appendix 2).

Primary Outcome

At the posttreatment phase, the mean improvement in the ICBT-I group was 8.4 (SD 4.7), and the mean improvement in the applied relaxation group was 5.0 (SD 5.4). The mixed model showed a significant immediate interaction effect (time by treatment) on the ISI, such that the ICBT-i group showed a greater rate of improvement (Figure 2 and Table 2). At follow-up for the ICBT-I group, the mean change score was 6.7 (SD 5.4); for the control group, the mean change score was 6.1 (SD 5.2). At the 6-month follow-up, there was no difference between treatments according to the ISI (Table S5 in Multimedia Appendix 2). Figure 3 shows the between-group effect sizes based on random effects (Cohen *d*) for all outcomes at week 5-and follow-up.

Figure 2. Sleep measures for internet-based cognitive behavioral therapy for insomnia (ICBT-i) and applied relaxation over time. Baseline data are presented only for Insomnia Severity Index (ISI). The active treatment period is illustrated by dots, and follow-up data are illustrated by diamonds. Note that y-axes are not illustrated from 0.

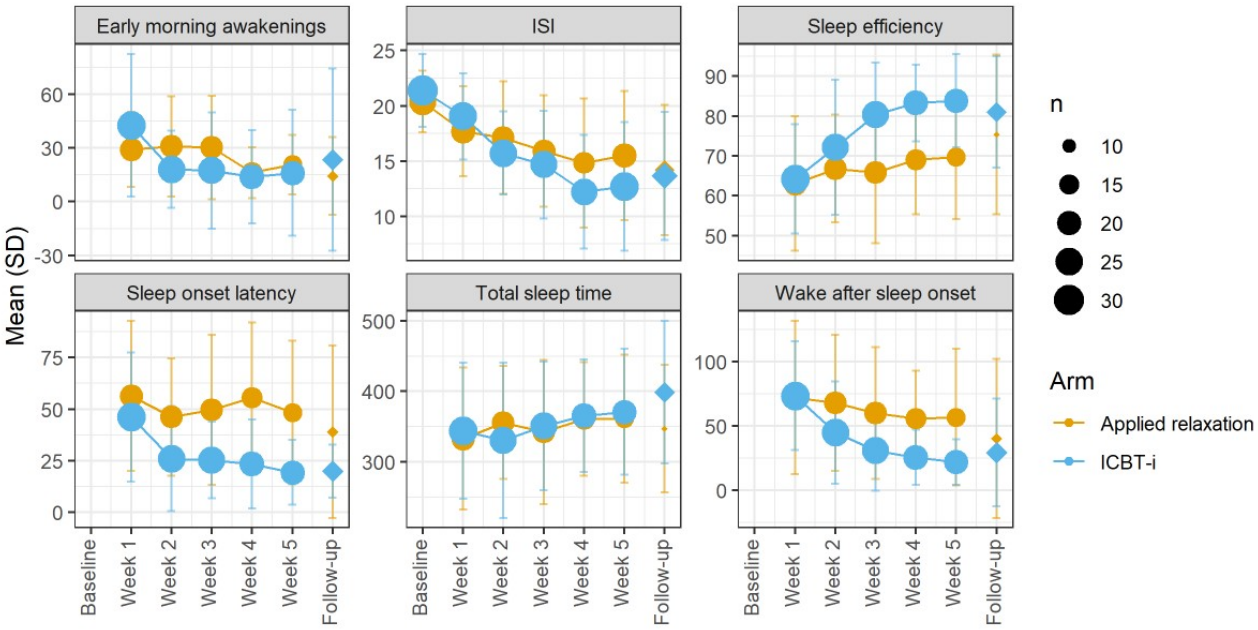


Table 2. Mixed models for ISI[a] from baseline to week 5 and for weekly sleep diary and KSS[b] scores from the beginning of the treatment (week 1) to the end of the treatment (week 5).

	Estimate (SE; 95% CI)	*P* value
ISI		
Intercept (AIC[c] 1414.2)	19.66 (0.70; 18.2 to 21.0)	*<.001* [d]
Treatment	1.14 (0.94; −0.69 to 2.98)	.22
Time (week 5)	−1.02 (0.23; −1.47 to −0.55)	*<.001*
Time (week 5)×treatment	−0.72 (0.31; −1.33 to −0.11)	*.02*
Sleep onset latency		
Intercept (AIC 2022.4)	53.03 (7.37; 38.5 to 67.4)	*<.001*
Treatment	−9.52 (9.62; −28.1 to 9.60)	.34
Time (week 5)	0.68 (5.00; −9.13 to 10.4)	.89
TimeQ[e] (week 5)	−0.58 (0.88; −2.32 to 1.14)	.50
Time (week 5)×treatment	−13.77 (6.55; −26.5 to −0.09)	*.04*
TimeQ (week 5)×treatment	2.38 (1.15; 0.13 to 4.63)	*.04*
Wake time after sleep onset		
Intercept (AIC 1957.3)	73.24 (17.59; 38.7 to 107.7)	*<.001*
Treatment	−2.27 (18.77; −39.0 to 34.5)	.90
Time (week 5)	−5.23 (4.57; −14.4 to 3.43)	.24
TimeQ (week 5)	0.29 (0.41; −0.50 to 1.09)	.47
Time (week 5)×treatment	−20.98 (5.96; −32.6 to −9.30)	*<.001*
TimeQ (week 5)×treatment	4.45 (0.53; 0.53 to 3.41)	*<.001*
Total sleep time		
Intercept (AIC 2239.3)	336.75 (20.93; 295.7 to 377.7)	*<.001*
Treatment	1.37 (27.30; −52.1 to 54.9)	.96
Time (week 5)	5.77 (3.57; −1.22 to 12.7)	.11
Time (week 5)×treatment	−0.92 (4.67; −10.1 to 8.23)	.84
Sleep efficiency		
Intercept (AIC 1500.6)	63.49 (3.71; 56.2 to 70.7)	*<.001*
Treatment	0.79 (4.46; −7.96 to 9.55)	.86
Time (week 5)	1.34 (0.90; −0.43 to 3.12)	.14
TimeQ (week 5)	0.03 (0.16; −0.27 to 0.35)	.81
Time (week 5)×treatment	8.29 (0.18; 5.97 to 10.6)	*<.001*
TimeQ (week 5)×treatment	−1.44 (0.20; −1.85 to −1.03)	*<.001*
Early morning awakenings		
Intercept (AIC 1834.3)	30.40 (6.18; 18.3 to 42.5)	*<.001*
Treatment	9.80 (8.53; −6.91 to 26.5)	.26
Time (week 5)	0.36 (2.16; −3.87 to 4.58)	.87
TimeQ (week 5)	−0.93 (0.35; −1.63 to −0.25)	*.008*
Time (week 5)×treatment	−20.65 (2.82; −26.1 to −15.2)	*<.001*
TimeQ (week 5)×treatment	4.50 (0.46; 3.60 to 5.41)	*<.001*
KSS		
Intercept (AIC 585.8)	6.15 (0.21; 5.73 to 6.56)	*<.001*
Treatment	−0.07 (0.28; −0.49 to 0.62)	.82

	Estimate (SE; 95% CI)	*P* value
Time (week 5)	−0.18 (0.06; −0.29 to −0.07)	*.002*
Time (week 5)×treatment	0.07 (0.07; −0.07 to 0.21)	.33

[a]ISI: Insomnia Severity Index.

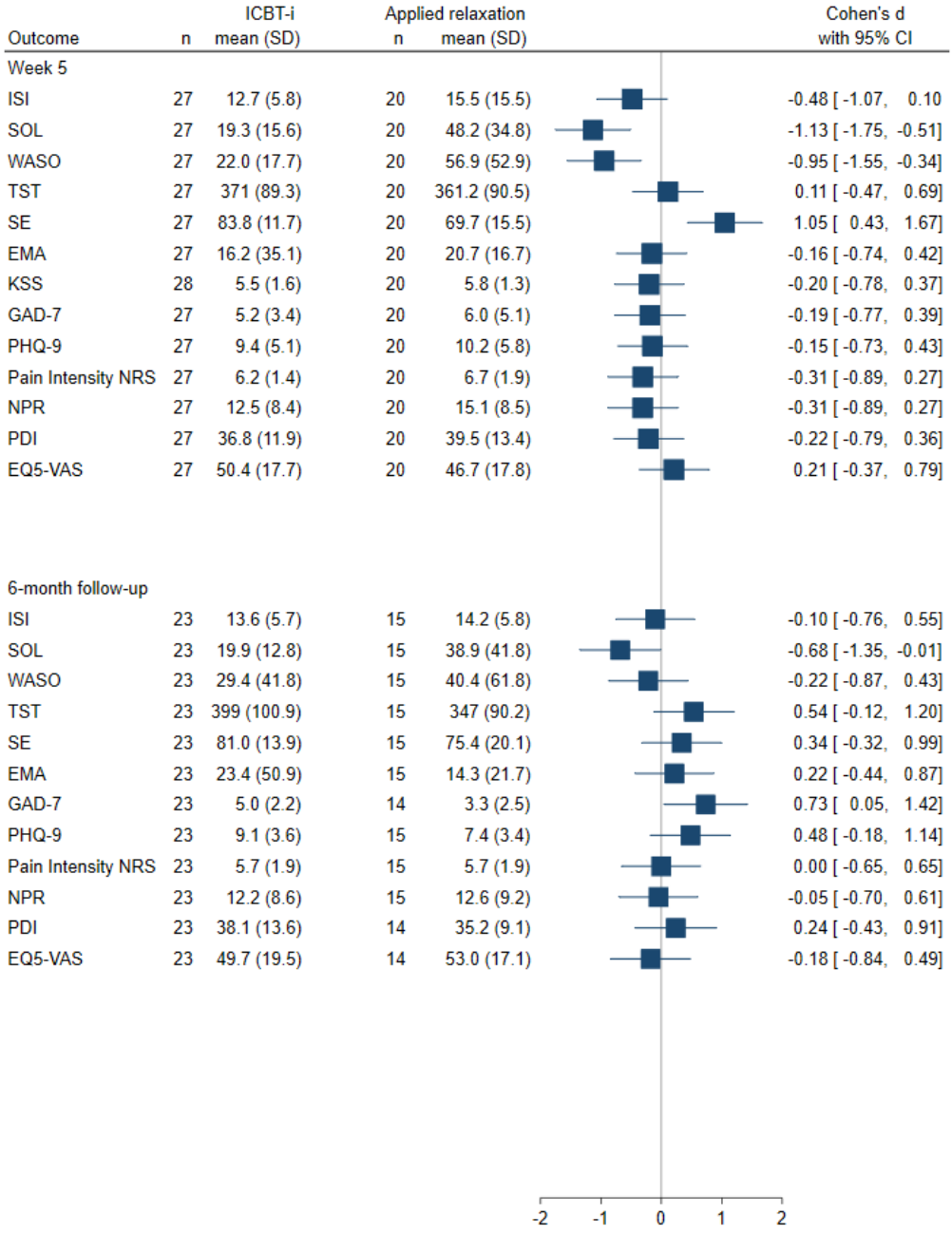
[b]KSS: Karolinska Sleepiness Scale.

[c]AIC: Akaike information criterion.

[d]Italics indicates statistically significant results.

[e]TimeQ: time quadratic.

Figure 3. Effect sizes (random effects) for internet-based cognitive behavioral therapy for insomnia (ICBT-i) and applied relaxation at week 5 and 6-month follow-up. EMA: early morning awakenings; EQ5-VAS: European Quality of Life 5-Dimension Visual Analog Scale; GAD-7: Generalized Anxiety Disorder–7 items; ISI: Insomnia Severity Index; KSS: Karolinska Sleepiness Scale; NPR: number of pain regions; NRS: Numeric Rating Scale; PDI: Pain Disability Index; PHQ-9: Patient Health Questionnaire–9 items; SE: sleep efficiency; SOL: sleep onset latency; TST: total sleep time; WASO: wake time after sleep onset.



Sleep Parameters

The mean values for sleep parameters at the posttreatment and follow-up periods are shown in Figure 2. The mixed model shows a significant time (quadratic)-by-treatment immediate effect on SOL, WASO, sleep efficiency, and EMA during the active treatment phase. Compared with the applied relaxation group, the ICBT-I group showed a more rapid increase in sleep efficiency and a more rapid decrease in EMA, SOL, and WASO (Table 2). At the 6-month follow-up, there was no statistically significant time-by-treatment effects on sleep diary measures when n dropped to 5 in the control group (Table S5 in Multimedia Appendix 2). Figure 3 shows the between-group effect sizes based on random effects (Cohen *d*) for all outcomes at week 5-and follow-up.

Daytime Sleepiness

The mean KSS values are shown in Figure S1 in Multimedia Appendix 2. The mixed models revealed either no treatment or time-by-treatment immediate effect on KSS (Table 2). That is, both treatment arms changed over time (active treatment phase), but there was no significant difference in the slope at any time point. Figure 3 shows the between-group effect sizes based on random effects (Cohen *d*) for all outcomes at week 5-and follow-up.

Other Secondary Outcomes

The mean values and their SDs at week 5-and follow-up are presented in Table S6 in Multimedia Appendix 2. No significant differences based on mixed models were found for most of the secondary outcomes with 2-time measurements (before vs after the treatment; Table 3). At follow-up, the mixed models showed significant time effects for PHQ-9 (depression) and GAD-7 (anxiety) versus those at week 5 (Table S7 in Multimedia Appendix 2). Figure 3 shows the between-group effect sizes based on random effects (Cohen *d*) for all outcomes at week 5 and follow-up.

Table 3. Mixed models for secondary outcomes from baseline to week 5.

Outcomes	Estimate (SE; 95% CI)	*P* value
GAD-7[a]		
Intercept (AIC[b] 591.2)	8.25 (0.94; 6.42 to 10.1)	*<.001* [c]
Treatment	−0.81 (1.25; −3.28 to 1.64)	.52
Time (week 5)	−1.81 (1.05; −3.87 to 0.24)	.09
Time (week 5)×treatment	−0.16 (1.38; −2.88 to 2.55)	.90
PHQ-9[d]		
Intercept (AIC 618.0)	14.17 (1.07; 12.3 to 16.5)	*<.001*
Treatment	−0.88 (1.44; −3.71 to 1.94)	.54
Time (week 5)	−3.92 (1.18; −6.23 to −1.61)	*.002*
Time (week 5)×treatment	0.15 (1.56; −2.91 to 3.22)	.92
Pain Intensity NRS[e]		
Intercept (AIC 369.4)	7.16 (0.34; 6.50 to 7.83)	*<.001*
Treatment	−0.87 (0.46; −1.76 to 0.03)	.06
Time (week 5)	−0.42 (0.30; −1.00 to 0.16)	.16
Time (week 5)×treatment	0.25 (0.39; −0.48 to 1.06)	.47
NPR[f]		
Intercept (AIC 670.7)	15.71 (1.92; 11.9 to 19.5)	*<.001*
Treatment	−2.08 (2.58; −7.14 to 2.99)	.43
Time (week 5)	−1.49 (0.95; −3.35 to 0.36)	.12
Time (week 5)×treatment	1.51 (1.25; −0.93 to 3.96)	.23
PDI[g]		
Intercept (AIC 764.6)	41.30 (2.42; 36.5 to 46.1)	*<.001*
Treatment	−1.43 (3.24; −7.79 to 4.93)	.66
Time (week 5)	−2.44 (2.09; −6.53 to 1.65)	.25
Time (week 5)×treatment	−0.85 (2.76; −6.26 to 4.56)	.76
EQ5-VAS[h]		
Intercept (AIC 856.1)	43.33 (3.60; 36.3 to 50.4)	*<.001*
Treatment	0.63 (4.84; −8.84 to 10.1)	.89
Time (week 5)	3.13 (3.60; −3.93 to 10.2)	.39
Time (week 5)×Treatment	3.32 (4.77; −6.02 to 12.7)	.49

[a]GAD-7: Generalized Anxiety Disorder–7 items.

[b]AIC: Akaike information criterion.

[c]Italics indicates statistically significant results.

[d]PHQ-9: Patient Health Questionnaire–9 items.

[e]NRS: Numeric Rating Scale.

[f]NPR: number of pain regions.

[g]PDI: Pain Disability Index.

[h]EQ5-VAS: European Quality of Life 5-Dimension Visual Analog Scale.

Treatment Credibility

Treatment credibility measures are presented in Table S8 in Multimedia Appendix 2. There were no significant group differences at week 2 (end of week 2 of treatment), week 5 (after the treatment), or follow-up.

Compliance

The median for completed modules was 4.5 (IQR 3-5) for the ICBT-i group and 4 (IQR 1.75-5) for the control group. The difference between the modules was not statistically significant (*P*=.46).

Negative Effects

The most frequently reported negative effect in both the treatment arms was increased stress. In the ICBT-i group, the second most reported negative effect was the experience that the treatment did not suit the patient. In contrast, the second most reported negative effect in the applied relaxation control condition was that participants stopped believing that there was available help. Table S9 in Multimedia Appendix 2 provides a comprehensive description of the negative effects.

Discussion

Principal Findings

In this study, we investigated the effects of ICBT-ion insomnia severity and other sleep and pain-related parameters in patients with chronic pain. Our results suggest that ICBT-i decreases insomnia severity and related nighttime symptoms. In particular, we found that the ICBT-i group, compared with the applied relaxation group, had greater immediate improvements in insomnia symptoms as measured by ISI, WASO, sleep efficiency, and EMA. The immediate effect (within-subject) of ICBT-i on ISI was clinically significant according to definition by Morin et al [5]. After 6 months, the group means approached each other, and no differences were confirmed on ISI (or other sleep parameters). Overall, our findings are consistent with previous studies suggesting that CBT-i delivers posttreatment improvements in insomnia and sleep symptoms, irrespective of the format [61-65]. However, further research is needed to investigate the effect of ICBT-i on a larger sample size of people with chronic pain.

Strengths and Limitations

The strengths of this study are its randomized design with an active control condition, low posttreatment attrition rate (13%), and clinical context. However, the attrition rate at follow-up was higher (30%). However, this study has several limitations. First, the small sample size makes it underpowered to detect small changes in pain and other symptoms and to adjust for confounders or to perform subgroup analyses, for example, by sex or educational level. The response to study invitations was surprisingly weak despite the invitations being directed to current and former patients identified with moderate to severe insomnia symptoms. This weak response might have been due to the relatively extensive treatments and commitment behavior changes require. Therefore, our results should be interpreted with caution owing to the underpowered nature of the study. This study also lacked objective sleep measures, such as accelerometer-based biosensors, a method that was excluded because of the priority to keep the treatment and assessment as brief as possible. The inclusion of these biosensors would also have made participation more complicated, with an increased risk for nonparticipation or dropout. Furthermore, our sample consisted primarily of women (45/54, 83%); therefore, a selection bias should not be ignored, regardless of the randomization procedure. This sex skew might be due to fact that insomnia symptoms are more common in women (aged >45 years) [66] and that the source of recruitment consisted of approximately 68.3% (123/180) women [4]. However, our results may still be considered ecologically valid. In addition, this study stands out with respect to the treatment length. Unpublished data from a previous treatment study by members of our group showed that SOL was the shortest after 5 weeks of treatment; therefore, 5 weeks/modules were chosen in this study. This study continued for another 2 weeks, but because an aim of this study is to develop a brief ICBT-i intervention, we decided to end the active treatment phase after 5 weeks. It can be argued that this was premature and that the treatment effect was limited by this decision. It is also possible that the effects of ICBT-i would have been better consolidated and maintained if treatment, sleep monitoring, and therapist support had lasted longer. Despite lower baseline ISI values, Jungquist et al [67] achieved a larger decrease in absolute number (13 points) on the same outcome measure after their 8-week CBT-i treatment. It may also be argued that the design of the active control condition, with a specific treatment factor (ie, applied relaxation), decreased the contrast with the experimental condition [68].

On the basis of these data, it is impossible to rule out whether the lack of time-by-treatment effect at follow-up is a result of a power issue or whether the 2 treatment conditions are equally effective in reducing insomnia symptoms in the long term. Given that 78% (42/54) of the sample had a sleeping problem duration of >5 years, explanations such as regression to the mean or spontaneous recovery are less likely. It cannot be ruled out that some overlap exists between treatment arms; therefore, the observed improvements, at least in part, may be due to common factors between the 2 arms. Both arms had access to the responsive sleep diary, calculating total sleep time, time in bed, SOL, WASO, EMA, and sleep efficiency. This could lead to unintentional behavior change in the control condition through the participants' own inferences of aggregated sleep data, although the addition of a wait-list control could have offered a valuable reference category. In contrast, a wait-list would have increased the power problem even more, given that further recruitment would not be possible. No measures of therapist fidelity were collected, but the standardized treatment content in both treatments, along with the supervision of therapists, helped maintain treatment fidelity [69]. However, differences in therapist support may exist, although the same therapists provided both treatments. Moreover, the participants in the control condition completed fewer treatment modules, which may constitute a major limitation. In addition, because blinding is problematic in this research field, it is also plausible to assume that there is a substantial risk of contamination between both treatment arms. Finally, the results from the 6-month follow-up were rather uncertain, as the response rate declined considerably. This decline is especially true for sleep parameters based on the sleep diary data.

In addition, there are some general limitations related to internet treatments. [70]. Although a therapist alliance seems possible to establish, the ability to monitor the patients' progress and

setbacks seems to be limited by this format. Consequently, the ability to make adaptations is limited, particularly when the treatment content is predefined. This study offers therapist support that likely contributes to increased adherence, which has been reported to be poor in previous studies. The exclusion of non-Swedish speakers and people lacking internet connections is another limitation related to this format.

Comparison With Previous Work

We found that immediate effects on ISI were larger in ICBT-i, but over the follow-up period, initial gains seemed to decline, but the applied control condition continued to improve. This finding was not unexpected, as results from a meta-analysis also illustrated that the long-term effects of CBT on insomnia are unclear [71]. Moreover, both treatments may be effective, but through different mechanisms and in different time perspectives. Applied relaxation targets arousal, which could hinder or interrupt sleep [38], whereas ICBT-i targets increasing the sleep homeostatic sleep load and establishing a sound circadian rhythm [35,36,72]. Therefore, we can speculate that a combination of both these treatments, which is common in CBT-i, would lead to an even better outcome. Notably, despite our intention to control for treatment extent, the applied relaxation group as a comparison arm completed a slightly less treatment content than the ICBT-i group. Furthermore, immediate time-by-treatment effects were found for the sleep parameters WASO, sleep efficiency, and EMA. SOL reached significance only at 1 time point. Changes from baseline are comparable with those of other studies, except for EMA, which previous studies failed to show any improvement or did not report [67]. The total absence of long-term time-by-treatment effects is likely, which could be explained by this study's inability to collect sleep diary data at follow-up, especially from participants in the control condition.

This study also highlighted that both web-based treatments improved psychological outcomes (depression and anxiety) over time, making these findings difficult to interpret. One may argue that both conditions improved these symptoms during the follow-up period, but unfortunately, the design of this study did not include any measure of the extent to which participants continued to apply the methods taught throughout treatment. It is also possible that medication titration might have affected these outcomes, as the medication was not specifically addressed in both conditions. Of the 30 participants in the experimental condition and 24 in the control condition, 29 (97%) and 23 (96%) at baseline reported taking prescription medication for their symptoms (including pain killers, benzodiazepine receptor agonists, or antidepressants), respectively. In addition, no exploratory analysis could be applied because changes in medication after the follow-up period were not monitored.

In contrast, none of the treatments had a significant effect on the number of pain regions or pain intensity at the posttreatment or follow-up period. Despite high initial values and clear immediate effect on insomnia severity (at least in ICBT-i), no time or time-by-treatment, effects on pain measures were found. These findings are supported by previous research in this field, which suggests that the effects of CBT-ion pain-related outcomes are highly unreliable [73]. The nonsignificant effects on pain measures may also be due to the strong correlation between insomnia and pain intensity. If the correlation is strong, a treatment effect on pain intensity is more likely than if a weak (but significant) correlation has been observed between the 2 variables [55,56]. In addition, two meta-analyses by Tang et al [17] and Selvanathan et al [74] also found overall small effects on pain intensity. Again, the lack of results in this study could be a consequence of the small sample size and low power. An alternative interpretation is that insomnia symptoms in participants with long-standing conditions can be improved without necessarily affecting pain intensity.

Finally, the most reported negative effect in both treatment arms (item 2: I felt like I was under more stress) was also the most frequent in several other CBT studies on different diagnoses [58]. The second most reported negative outcome in the active control condition (item 19: I stopped believing that there was available help) was excluded in the revised version of the NEQ because of poor goodness of fit [58]. It is also possible that this item expresses the absence of an expected treatment effect in nonresponders, rather than a negative effect of treatment. Interestingly, 22% (7/30) of the participants in the ICBT-i group and 17% (4/24) of the controls reported more problems with sleep (item 1). The item does not indicate whether this refers to long- or short-term problems; however, the impact of these negative effects seems to be limited to the means presented here.

Conclusion and Clinical Implications

Compared with applied relaxation, ICBT-i led to a greater decline in insomnia symptoms. Treatment gains were largely preserved over 6 months, but the group difference decreased as the control condition continued to improve over this period. Furthermore, applied relaxation was more successful in reducing comorbid anxiety and depressive symptoms in the long term. Further studies are needed to examine whether a combination of ICBT-i and applied relaxation leads to even larger reductions in insomnia symptoms or what necessary components of ICBT-i work best for insomnia in patients with chronic pain. Nevertheless, it was possible to achieve clinically significant improvement in insomnia symptoms via this brief 5-week internet treatment. If our finding that a brief CBT-i intervention (possibly in combination with applied relaxation) affects insomnia in patients with chronic pain can be confirmed in larger studies, this intervention may be practically applicable in the clinical setting. This could be an important advance in the treatment arsenal, as a large proportion of patients with chronic pain also experience insomnia.

Acknowledgments

Associate Professor Anna Anund at the Swedish National Road and Transport Research Institute (VTI) valuably contributed to the planning and design of this study. The authors would also like to highlight the work performed by the master's students

Anna-Maja Sundell, Johanna Höppe-Persson, Ellen French Lind, and Veronika Johansson. Recruitment, assessment, and treatment would not have been possible without their efforts. This study receives no funding.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 925 KB - jmir_v24i4e29258_app1.pdf]

Multimedia Appendix 2
Supplementary information on methods and results.
[DOCX File , 85 KB - jmir_v24i4e29258_app2.docx]

References

1. Vos T, Flaxman A, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2163-2196 [FREE Full text] [doi: 10.1016/S0140-6736(12)61729-2] [Medline: 23245607]
2. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. Eur J Pain 2006 May;10(4):287-333. [doi: 10.1016/j.ejpain.2005.06.009] [Medline: 16095934]
3. European Federation of IASP Chapters. EFIC's declaration on chronic pain as a major healthcare problem, a disease in its own right. 2001 Presented at: Presented at the European Parliament, after endorsement by 25 European Chapters of the International Society for the Study of Pain; October 9; Brussels, Belgium URL: http://www.efic.org
4. Alföldi P, Wiklund T, Gerdle B. Comorbid insomnia in patients with chronic pain: a study based on the Swedish quality registry for pain rehabilitation (SQRP). Disabil Rehabil 2014 Dec 09;36(20):1661-1669. [doi: 10.3109/09638288.2013.864712] [Medline: 24320022]
5. Morin C, Belleville G, Bélanger L, Ivers H. The Insomnia Severity Index: psychometric indicators to detect insomnia cases and evaluate treatment response. Sleep 2011 May 01;34(5):601-608 [FREE Full text] [doi: 10.1093/sleep/34.5.601] [Medline: 21532953]
6. Gupta A, Silman AJ, Ray D, Morriss R, Dickens C, MacFarlane GJ, et al. The role of psychosocial factors in predicting the onset of chronic widespread pain: results from a prospective population-based study. Rheumatology (Oxford) 2007 Apr 13;46(4):666-671. [doi: 10.1093/rheumatology/kel363] [Medline: 17085772]
7. Mork PJ, Nilsen TI. Sleep problems and risk of fibromyalgia: longitudinal data on an adult female population in Norway. Arthritis Rheum 2012 Jan 29;64(1):281-284 [FREE Full text] [doi: 10.1002/art.33346] [Medline: 22081440]
8. Generaal E, Vogelzangs N, Penninx B, Dekker J. Insomnia, sleep duration, depressive symptoms, and the onset of chronic multisite musculoskeletal pain. Sleep 2017 Jan 01;40(1):1-10. [doi: 10.1093/sleep/zsw030] [Medline: 28364456]
9. Affleck G, Urrows S, Tennen H, Higgins P, Abeles M. Sequential daily relations of sleep, pain intensity, and attention to pain among women with fibromyalgia. Pain 1996;68(2):363-368. [doi: 10.1016/S0304-3959(96)03226-5]
10. O'Brien E, Waxenberg L, Atchison J. Intraindividual variability in daily sleep and pain ratings among chronic pain patients: bidirectional association and the role of negative mood. Clin J Pain 2011;27(5):425-433. [doi: 10.1097/ajp.0b013e318208c8e4]
11. Wei Y, Blanken TF, Van Someren EJ. Insomnia really hurts: effect of a bad night's sleep on pain increases with insomnia severity. Front Psychiatry 2018 Aug 28;9:377 [FREE Full text] [doi: 10.3389/fpsyt.2018.00377] [Medline: 30210367]
12. McHugh RK, Edwards RR, Ross EL, Jamison RN. Does bedtime matter among patients with chronic pain? A longitudinal comparison study. Pain Report 2019 May;4(3):e747. [doi: 10.1097/pr9.0000000000000747]
13. Borbély AA, Daan S, Wirz-Justice A, Deboer T. The two-process model of sleep regulation: a reappraisal. J Sleep Res 2016 Apr 14;25(2):131-143 [FREE Full text] [doi: 10.1111/jsr.12371] [Medline: 26762182]
14. Maurer LF, Espie CA, Kyle SD. How does sleep restriction therapy for insomnia work? A systematic review of mechanistic evidence and the introduction of the Triple-R model. Sleep Med Rev 2018 Dec;42:127-138. [doi: 10.1016/j.smrv.2018.07.005] [Medline: 30177248]
15. Maurer L, Espie C, Omlin X, Reid MJ, Sharman R, Gavriloff D, et al. Isolating the role of time in bed restriction in the treatment of insomnia: a randomized, controlled, dismantling trial comparing sleep restriction therapy with time in bed regularization. Sleep 2020 Nov 12;43(11):zsaa096. [doi: 10.1093/sleep/zsaa096] [Medline: 32421814]
16. Schutte-Rodin S, Broch L, Buysse D, Dorsey C, Sateia M. Clinical guideline for the evaluation and management of chronic insomnia in adults. J Clin Sleep Med 2008 Oct 15;04(05):487-504. [doi: 10.5664/jcsm.27286]
17. Tang N, Lereya S, Boulton H, Miller M, Wolke D, Cappuccio F. Nonpharmacological treatments of insomnia for long-term painful conditions: a systematic review and meta-analysis of patient-reported outcomes in randomized controlled trials. Sleep 2015 Nov 01;38(11):1751-1764 [FREE Full text] [doi: 10.5665/sleep.5158] [Medline: 25902806]

18. McCrae C, Williams J, Roditi D, Anderson R, Mundt JM, Miller MB, et al. Cognitive behavioral treatments for insomnia and pain in adults with comorbid chronic insomnia and fibromyalgia: clinical outcomes from the SPIN randomized controlled trial. Sleep 2019 Mar 01;42(3):zsy234 [FREE Full text] [doi: 10.1093/sleep/zsy234] [Medline: 30496533]
19. Baglioni C, Altena E, Bjorvatn B, Blom K, Bothelius K, Devoto A, et al. The European academy for cognitive behavioural therapy for insomnia: an initiative of the European insomnia network to promote implementation and dissemination of treatment. J Sleep Res 2020 Apr;29(2):e12967 [FREE Full text] [doi: 10.1111/jsr.12967] [Medline: 31856367]
20. Stiefel F, Stagno D. Management of insomnia in patients with chronic pain conditions. CNS Drugs 2004;18(5):285-296. [doi: 10.2165/00023210-200418050-00002] [Medline: 15089114]
21. Tang N, Goodchild C, Hester J, Salkovskis P. Pain-related insomnia versus primary insomnia: a comparison study of sleep pattern, psychological characteristics, and cognitive-behavioral processes. Clinical J Pain 2012;28(5):428-436. [doi: 10.1097/ajp.0b013e31823711bc]
22. Afolalu EF, Moore C, Ramlee F, Goodchild CE, Tang NK. Development of the pain-related beliefs and attitudes about sleep (PBAS) scale for the assessment and treatment of insomnia comorbid with chronic pain. J Clin Sleep Med 2016 Sep 15;12(9):1269-1277 [FREE Full text] [doi: 10.5664/jcsm.6130] [Medline: 27448428]
23. Harvey L, Inglis SJ, Espie CA. Insomniacs' reported use of CBT components and relationship to long-term clinical outcome. Behav Res Ther 2002 Jan;40(1):75-83. [doi: 10.1016/s0005-7967(01)00004-3]
24. 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]
25. 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 09;12(6):745-764. [doi: 10.1586/erp.12.67] [Medline: 23252357]
26. 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):110. [doi: 10.1186/1471-230x-11-110]
27. Blom K, Tarkian Tillgren H, Wiklund T, Danlycke E, Forssén M, Söderström A, et al. Internet-vs. group-delivered cognitive behavior therapy for insomnia: a randomized controlled non-inferiority trial. Behav Res Ther 2015 Jul;70:47-55 [FREE Full text] [doi: 10.1016/j.brat.2015.05.002] [Medline: 25981329]
28. Hagatun S, Vedaa O, Nordgreen T, Smith OR, Pallesen S, Havik OE, et al. The short-term efficacy of an unguided internet-based cognitive-behavioral therapy for insomnia: a randomized controlled trial with a six-month nonrandomized follow-up. Behav Sleep Med 2019 Mar 27;17(2):137-155. [doi: 10.1080/15402002.2017.1301941] [Medline: 28345961]
29. Philip P, Dupuy L, Morin CM, de Sevin E, Bioulac S, Taillard J, et al. Smartphone-based virtual agents to help individuals with sleep concerns during COVID-19 confinement: feasibility study. J Med Internet Res 2020 Dec 18;22(12):e24268 [FREE Full text] [doi: 10.2196/24268] [Medline: 33264099]
30. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009 Dec 15;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
31. Sullivan G, Feinn R. Using effect size-or why the P value is not enough. J Grad Med Educ 2012;4(3):279-282. [doi: 10.4300/jgme-d-12-00156.1]
32. Sheehan D. Mini International Neuropsychiatric Interview (MINI) 7.0.2. Jacksonville, FL: Medical Outcomes System; 2015.
33. American Psychiatric Association. DSM-5 Task Force. Diagnostic And Statistical Manual Of Mental Disorders, Fifth Edition. Washington, DC: American Psychiatric Publishing; 2013.
34. Vlaescu G, Alasjö A, Miloff A, Carlbring P, Andersson G. Features and functionality of the Iterapi platform for internet-based psychological treatment. Internet Interv 2016 Nov;6:107-114 [FREE Full text] [doi: 10.1016/j.invent.2016.09.006] [Medline: 30135819]
35. Bootzin R, Epstein D. Understanding and treating insomnia. Annu Rev Clin Psychol 2011;7:435-458. [doi: 10.1146/annurev.clinpsy.3.022806.091516] [Medline: 17716026]
36. Kyle SD, Aquino MR, Miller CB, Henry AL, Crawford MR, Espie CA, et al. Towards standardisation and improved understanding of sleep restriction therapy for insomnia disorder: a systematic examination of CBT-I trial content. Sleep Med Rev 2015 Oct;23:83-88. [doi: 10.1016/j.smrv.2015.02.003] [Medline: 25771293]
37. Andrews N, Strong J, Meredith P, D'Arrigo R. Association between physical activity and sleep in adults with chronic pain: a momentary, within-person perspective. Phys Ther 2014 Apr;94(4):499-510. [doi: 10.2522/ptj.20130302] [Medline: 24231224]
38. Ost LG. Applied relaxation: description of a coping technique and review of controlled studies. Behav Res Ther 1987;25(5):397-409. [doi: 10.1016/0005-7967(87)90017-9] [Medline: 3318800]
39. Ström L, Pettersson R, Andersson G. Internet-based treatment for insomnia: a controlled evaluation. J Consult Clin Psychol 2004;72(1):113-120. [doi: 10.1037/0022-006x.72.1.113]
40. Nicassio P, Bootzin R. A comparison of progressive relaxation and autogenic training as treatments for insomnia. J Abnorm Psychol 1974 Jun;83(3):253-260. [doi: 10.1037/h0036729] [Medline: 4844912]

41. Andersson G, Lundström P, Ström L. Internet-based treatment of headache: does telephone contact add anything? Headache 2003 Apr;43(4):353-361. [doi: 10.1046/j.1526-4610.2003.03070.x] [Medline: 12656706]
42. Thorsell J, Finnes A, Dahl J. A comparative study of 2 manual-based self-help interventions, acceptance and commitment therapy and applied relaxation, for persons with chronic pain. Clin J Pain 2011;27(8):716-723. [doi: 10.1097/ajp.0b013e318219a933]
43. Bastien C, Vallieres A, Morin C. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med 2001;2(4):297-307. [doi: 10.1016/s1389-9457(00)00065-4]
44. Dragioti E, Wiklund T, Alföldi P, Gerdle B. The Swedish version of the insomnia severity index: factor structure analysis and psychometric properties in chronic pain patients. Scand J Pain 2015 Oct 01;9(1):22-27. [doi: 10.1016/j.sjpain.2015.06.001] [Medline: 29911642]
45. Buysse D, Ancoli-Israel S, Edinger J, Lichstein K, Morin C. Recommendations for a standard research assessment of insomnia. Sleep 2006 Sep;29(9):1155-1173. [doi: 10.1093/sleep/29.9.1155] [Medline: 17040003]
46. Kaida K, Takahashi M, Akerstedt T, Nakata A, Otsuka Y, Haratani T, et al. Validation of the Karolinska sleepiness scale against performance and EEG variables. Clin Neurophysiol 2006 Jul;117(7):1574-1581. [doi: 10.1016/j.clinph.2006.03.011] [Medline: 16679057]
47. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
48. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
49. Hawley CJ, Gale TM, Smith PS, Jain S, Farag A, Kondan R, et al. Equations for converting scores between depression scales (MÅDRS, SRS, PHQ-9 and BDI-II): good statistical, but weak idiographic, validity. Hum Psychopharmacol 2013 Nov 23;28(6):544-551. [doi: 10.1002/hup.2341] [Medline: 24519690]
50. Ferreira-Valente M, Pais-Ribeiro J, Jensen M. Validity of four pain intensity rating scales. Pain 2011 Oct;152(10):2399-2404. [doi: 10.1016/j.pain.2011.07.005] [Medline: 21856077]
51. Alföldi P, Dragioti E, Wiklund T, Gerdle B. Spreading of pain and insomnia in patients with chronic pain: results from a national quality registry (SQRP). J Rehabil Med 2017 Jan 19;49(1):63-70 [FREE Full text] [doi: 10.2340/16501977-2162] [Medline: 27904908]
52. Tait RC, Chibnall JT, Krause S. The Pain Disability Index: psychometric properties. Pain 1990 Feb;40(2):171-182. [doi: 10.1016/0304-3959(90)90068-O] [Medline: 2308763]
53. Chibnall JT, Tait RC. The Pain Disability Index: factor structure and normative data. Archives Physical Med Rehab 1994 Oct;75(10):1082-1086. [doi: 10.1016/0003-9993(94)90082-5]
54. Beemster T, van Bennekom C, van Velzen J, Reneman M, Frings-Dresen M. The interpretation of change score of the pain disability index after vocational rehabilitation is baseline dependent. Health Qual Life Outcomes 2018 Sep 14;16(1):182 [FREE Full text] [doi: 10.1186/s12955-018-1000-1] [Medline: 30217206]
55. Fayers P, Machin D. Quality of Life The Assessment, Analysis and Interpretation of Patient-reported Outcomes. Hoboken, NJ: John Wiley & Sons; 2000.
56. Ringqvist A, Dragioti E, Björk M, Larsson B, Gerdle B. Moderate and stable pain reductions as a result of interdisciplinary pain rehabilitation-a cohort study from the Swedish quality registry for pain rehabilitation (SQRP). J Clin Med 2019 Jun 24;8(6):905 [FREE Full text] [doi: 10.3390/jcm8060905] [Medline: 31238588]
57. Rozental A, Kottorp A, Boettcher J, Andersson G, Carlbring P. Negative effects of psychological treatments: an exploratory factor analysis of the negative effects questionnaire for monitoring and reporting adverse and unwanted events. PLoS One 2016 Jun 22;11(6):e0157503 [FREE Full text] [doi: 10.1371/journal.pone.0157503] [Medline: 27331907]
58. Rozental A, Kottorp A, Forsström D, Månsson K, Boettcher J, Andersson G, et al. The Negative Effects Questionnaire: psychometric properties of an instrument for assessing negative effects in psychological treatments. Behav Cognit Psychother 2019 Mar 15;47(5):559-572. [doi: 10.1017/s1352465819000018]
59. Hesser H. Modeling individual differences in randomized experiments using growth models: recommendations for design, statistical analysis and reporting of results of internet interventions. Internet Interv 2015 May;2(2):110-120. [doi: 10.1016/j.invent.2015.02.003]
60. Salim A, Mackinnon A, Christensen H, Griffiths K. Comparison of data analysis strategies for intent-to-treat analysis in pre-test-post-test designs with substantial dropout rates. Psychiatry Res 2008 Sep 30;160(3):335-345. [doi: 10.1016/j.psychres.2007.08.005] [Medline: 18718673]
61. Jungquist CR, O'Brien C, Matteson-Rusby S, Smith MT, Pigeon WR, Xia Y, et al. The efficacy of cognitive-behavioral therapy for insomnia in patients with chronic pain. Sleep Med 2010 Mar;11(3):302-309 [FREE Full text] [doi: 10.1016/j.sleep.2009.05.018] [Medline: 20133188]
62. Rybarczyk B, Stepanski E, Fogg L, Lopez M, Barry P, Davis A. A placebo-controlled test of cognitive-behavioral therapy for comorbid insomnia in older adults. J Consult Clin Psychol 2005 Dec;73(6):1164-1174. [doi: 10.1037/0022-006x.73.6.1164]
63. Vitiello MV, Rybarczyk B, Von Korff M, Stepanski EJ. Cognitive behavioral therapy for insomnia improves sleep and decreases pain in older adults with co-morbid insomnia and osteoarthritis. J Clin Sleep Med 2009 Aug 15;05(04):355-362. [doi: 10.5664/jcsm.27547]

64. van Straten A, van der Zweerde T, Kleiboer A, Cuijpers P, Morin CM, Lancee J. Cognitive and behavioral therapies in the treatment of insomnia: a meta-analysis. Sleep Med Rev 2018 Apr;38:3-16. [doi: 10.1016/j.smrv.2017.02.001] [Medline: 28392168]
65. Luik AI, Kyle SD, Espie CA. Digital Cognitive Behavioral Therapy (dCBT) for insomnia: a state-of-the-science review. Curr Sleep Med Rep 2017 May 8;3(2):48-56 [FREE Full text] [doi: 10.1007/s40675-017-0065-4] [Medline: 28553574]
66. Sivertsen B, Krokstad S, Øverland S, Mykletun A. The epidemiology of insomnia: associations with physical and mental health. The HUNT-2 study. J Psychosom Res 2009 Aug;67(2):109-116. [doi: 10.1016/j.jpsychores.2009.05.001] [Medline: 19616137]
67. Jungquist CR, Tra Y, Smith MT, Pigeon WR, Matteson-Rusby S, Xia Y, et al. The durability of cognitive behavioral therapy for insomnia in patients with chronic pain. Sleep Disord 2012;2012:679648-679648 [FREE Full text] [doi: 10.1155/2012/679648] [Medline: 23470897]
68. Mohr DC, Ho J, Hart TL, Baron KG, Berendsen M, Beckner V, et al. Control condition design and implementation features in controlled trials: a meta-analysis of trials evaluating psychotherapy for depression. Transl Behav Med 2014 Dec 2;4(4):407-423 [FREE Full text] [doi: 10.1007/s13142-014-0262-3] [Medline: 25584090]
69. Craske MG, Rose RD, Lang A, Welch SS, Campbell-Sills L, Sullivan G, et al. Computer-assisted delivery of cognitive behavioral therapy for anxiety disorders in primary-care settings. Depress Anxiety 2009 Mar;26(3):235-242 [FREE Full text] [doi: 10.1002/da.20542] [Medline: 19212970]
70. Gratzer D, Khalid-Khan F. Internet-delivered cognitive behavioural therapy in the treatment of psychiatric illness. CMAJ 2016 Mar 01;188(4):263-272 [FREE Full text] [doi: 10.1503/cmaj.150007] [Medline: 26527829]
71. Trauer JM, Qian MY, Doyle JS, Rajaratnam SM, Cunnington D. Cognitive behavioral therapy for chronic insomnia. Ann Intern Med 2015 Aug 04;163(3):191-204. [doi: 10.7326/m14-2841]
72. Anderson KN. Insomnia and cognitive behavioural therapy-how to assess your patient and why it should be a standard part of care. J Thorac Dis 2018 Jan;10(Suppl 1):S94-102 [FREE Full text] [doi: 10.21037/jtd.2018.01.35] [Medline: 29445533]
73. Finan PH, Buenaver LF, Coryell VT, Smith MT. Cognitive-behavioral therapy for comorbid insomnia and chronic pain. Sleep Med Clin 2014 Jun 01;9(2):261-274 [FREE Full text] [doi: 10.1016/j.jsmc.2014.02.007] [Medline: 25477769]
74. Selvanathan J, Pham C, Nagappa M, Peng PW, Englesakis M, Espie CA, et al. Cognitive behavioral therapy for insomnia in patients with chronic pain - A systematic review and meta-analysis of randomized controlled trials. Sleep Med Rev 2021 Dec;60:101460. [doi: 10.1016/j.smrv.2021.101460] [Medline: 33610967]

Abbreviations

CBT: cognitive behavioral therapy
CBT-i: cognitive behavioral therapy for insomnia
CONSORT: Consolidated Standards of Reporting Trials
GAD-7: Generalized Anxiety Disorder–7 items
ICBT-i: internet-based cognitive behavioral therapy for insomnia
ISI: Insomnia Severity Index
KSS: Karolinska Sleepiness Scale
NEQ: Negative Effects Questionnaire
PDI: Pain Disability Index
PHQ-9: Patient Health Questionnaire–9 items
RCT: randomized controlled trial
SOL: sleep onset latency
WASO: wake time after sleep onset

Edited by G Eysenbach; submitted 31.03.21; peer-reviewed by F Pallavicini, G Pramana, A Malfliet; comments to author 13.05.21; revised version received 26.07.21; accepted 09.02.22; published 29.04.22.

Please cite as:
Wiklund T, Molander P, Lindner P, Andersson G, Gerdle B, Dragioti E
Internet-Delivered Cognitive Behavioral Therapy for Insomnia Comorbid With Chronic Pain: Randomized Controlled Trial
J Med Internet Res 2022;24(4):e29258
URL: https://www.jmir.org/2022/4/e29258
doi:10.2196/29258
PMID:35486418

Original Paper

Frequency and Correlates of Online Consultations With Doctors or Therapists in Middle-Aged and Older Adults: Nationally Representative Cross-sectional Study

André Hajek[1], Prof Dr; Hans-Helmut König[1], MPH, Prof Dr

Department of Health Economics and Health Services Research, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Corresponding Author:
André Hajek, Prof Dr
Department of Health Economics and Health Services Research
University Medical Center Hamburg-Eppendorf
Martinistraße 52
Hamburg, 20246
Germany
Phone: 49 40 7410 528 77
Fax: 49 40 7410 40261
Email: a.hajek@uke.de

Abstract

Background: A few studies have identified the frequency and correlates of online consultations with doctors or therapists. However, there is a lack of studies using nationally representative data from middle-aged and older adults in Germany.

Objective: This study aims to determine the frequency and correlates of online consultations with doctors or therapists in Germany.

Methods: For this study, cross-sectional data were taken from the nationally representative German Ageing Survey (DEAS; n=3067 in the analytical sample; age range 46-98 years). As part of the DEAS, a short survey was conducted between June 8 and July 22, 2020, examining the everyday life and living conditions among these middle-aged and older individuals during the COVID-19 pandemic. The frequency of online consultations with doctors or therapists served as the dependent variable (daily, several times a week, once a week, 1-3 times a month, less often, and never). Multiple logistic regressions were performed.

Results: In sum, 10.02% (381/3806) of individuals with access to the internet had online consultations with doctors or therapists. Multiple logistic regressions showed that the likelihood of using online consultations with doctors or therapists (compared with those never using such services) was positively associated with higher education (compared with medium education; odds ratio [OR] 1.31, 95% CI 1.01-1.70), living with a partner in the same household (compared with single; OR 1.53, 95% CI 1.05-2.22), poorer self-rated health (OR 1.42, 95% CI 1.16-1.74), increased loneliness (OR 1.45, 95% CI 1.10-1.90), and increased satisfaction with life (OR 1.30, 95% CI 1.03-1.64).

Conclusions: Study findings suggest that a non-negligible proportion of about 1 out of 10 individuals aged 46 years and over had online consultations with doctors or therapists. However, compared with other countries, this proportion remains small. Knowledge about the correlates of (non)use may assist in identifying corresponding individuals. In times of reshaping the health care system, these efforts in online consultations with doctors or therapists may contribute to addressing patient needs. Moreover, increased use of such services may reduce the risk of getting infected with SARS-CoV-2 by reducing social contact.

(J Med Internet Res 2022;24(4):e29781) doi:10.2196/29781

KEYWORDS

online consultations; doctor; therapists; telehealth; COVID-19; SARS-CoV-2; digital health

Introduction

In the last 2 decades, with better access to the internet, and more high-speed connections, online consultations with doctors or therapists have grown continually [1]. Online consultations can improve patient outcomes via increased access to care and medical information [1]. Furthermore, online consultations can increase job satisfaction and work-life balance among physicians [2]. Moreover, there is an increased need for online consultations due to increased economic costs, poor access in rural areas, and

demographic aging, with an increasing number of individuals in old age suffering from various chronic illnesses.

During the COVID-19 pandemic, the use of online consultations sought to reduce unnecessary social contact by reducing the number of in-person consultations [3]. This can reduce the risk of getting infected with SARS-CoV-2, which is key among older adults with risk factors for more severe experiences of COVID-19 [4].

Online consultations with doctors or therapists can, for example, be used for issuing sick notes or care of patients after surgery. A recent systematic review recommended use of such consultations in orthopedics [5]. Additionally, it has been shown that patient satisfaction with video consultations was high to very high during the COVID-19 pandemic [6].

It should be noted that the term doctor (*arzt*) in Germany refers to physicians, while the term therapists (*therapeut*) commonly refers to psychological or medical psychotherapists, as well as other therapists (eg, occupational therapists or physical therapists). According to the German Medical Association, the total number of currently working physicians (registered with state medical associations) was around 409,000 in 2020 [7]. The physician density (inhabitants per working physician) varies between the states in Germany—from 133 (Hamburg) to 248 (Brandenburg) [7]. In Germany, around 12,000 specialists worked in psychiatry and psychotherapy in 2020 [8] and around 48,000 psychological psychotherapists worked in 2019 [9]. Moreover, around 203,000 individuals work as physical therapists in Germany (data from 2019) [10] and around 59,000 individuals work as occupational therapists (data from 2017) [11]. Regional differences are present (eg, between rural and urban areas and between certain states) [10].

It may be worth describing online consultations with doctors or therapists in Germany. Certain video consultations have been included in the outpatient medical fee schedule (EBM) applied for patients who are members of a statutory health insurance (approximately 90% of the population) in Germany since April 2017. Moreover, members of a private health insurance (around 10% of the population) can use online consultation as it is in accordance with the medical fee schedule used for privately insured patients (GOÄ). According to the National Association of Statutory Health Insurance Physicians (KBV) [12], around 25% of doctors' offices offered video consultations in Germany in March 2020, whereas only 2% offered such services in 2017. Furthermore, around 3000 hours of online consultations were conducted in 2019, whereas 1.4 million hours of online consultations were conducted in the first half of 2020 [12]. Besides, 3 out of 4 online consultations were conducted in the field of psychotherapy in the first half of 2020. Additionally, around 12% of the general practitioners (GPs) offered online consultations in the second quarter of 2020 [12]. Online consultations are valued by GPs, medical specialists, and hospital doctors in Germany [13]. However, it is agreed that they cannot substitute care in various specialties, such as hand surgery, but rather complement care [14].

In Germany, there is still a general lack of data on the frequency of online consultations with doctors or therapists based on *nationally representative samples* of middle-aged and older adults. Because of this lack in knowledge, our aim was to determine the frequency and correlates of online consultations with doctors or therapists *based on nationally representative data from Germany*. We would like to note that identifying the correlates of such online consultations was an explorative aim of this study (without any prespecified hypotheses). Knowledge about the use of online consultations is important for policy makers. For example, based on this knowledge, strategic planning could be established and implemented to identify and support nonadopters of online consultations. This could assist in improving the uptake of online consultations.

Methods

Sample

Data for this study were taken from the German Ageing Survey (DEAS; a nationally representative sample of individuals aged 40 years in Germany). As part of the DEAS, a short survey was conducted between June 8 and July 22, 2020, examining the everyday life and living conditions among these middle-aged and older individuals during the COVID-19 pandemic. This survey was addressed to all panel participants who had taken part in previous DEAS waves at least once. Thus, the basis for this survey was all willing panel participants who could still be reached from the baseline samples 1996 to 2014 (from 1996: 539 individuals; from 2002: 525 individuals; from 2008: 1549 individuals; from 2014: 2210 individuals). Thus, individuals were at least 46 years old in the survey conducted in 2020.

In sum, 4823 individuals who had already participated in 1 or more former waves were included in this short survey (paper and pencil interview) with a response rate of 56.5%, which is comparable to other survey studies in Germany [15]. Of the included sample, 3806 individuals had access to the internet and filled out the dependent variable (analytical sample in regressions analysis: n=3067; reduction in sample size can be explained by missing data in the independent variables, as listwise deletion was used in regression analysis [16]). Further details regarding the DEAS study are given elsewhere [17].

Dependent Variables

Individuals with access to the internet (ie, individuals who reported "yes" when asked whether he or she has access to the internet) were asked to report the frequency of "consultations with doctors or therapists via an online platform" (daily, several times a week, once a week, 1-3 times a month, less often, and never). It was dichotomized (0=never; 1=otherwise including daily, several times a week, once a week, 1-3 times a month, and less often). This variable served as the dependent variable.

Independent Variables

We included socioeconomic, lifestyle-related, health-related, COVID-19–related factors, and psychosocial factors as independent variables. With regard to socioeconomic factors, we included age, sex, educational level (International Standard Classification of Education 97 [ISCED-97] [18]: low, medium, or high education), employment status (employed, retired, non-employed), living situation (single, with partner in household, with partner without a common household), having at least one child (no or yes), migration background (no or yes),

monthly household net income (in Euros), region (West Germany or East Germany), and type of district (large cities, urban cities, urban-rural districts, rural districts).

With regard to lifestyle factors, we included engagement in physical activities and the frequency of walks (in both cases: daily, several times a week, once a week, 1-3 times a month, less often, never). With regard to health-related factors, we included self-rated health (from 1=very good to 5=very bad) and depressive symptoms (10-item Center for Epidemiological Studies-Depression [CES-D] scale [19] score ranging from 0 to 30, with higher scores reflecting more depressive symptoms). In this study, Cronbach alpha for the CES-D was .85. The CES-D has favorable psychometric characteristics [20].

With regard to COVID-19–related factors, we included the following variables: feeling that the COVID-19 crisis posed a threat to oneself (from 0=not at all a threat for me to 10=extreme threat for me); infection among people in one's own personal environment with the coronavirus (yes, no, don't know); one's own infection with the coronavirus (yes, no, don't know); and the feeling that one can influence an infection with the coronavirus (from 1=not at all to 7=entirely). With regard to psychosocial factors, we included life satisfaction (Satisfaction with Life Scale [SWLS] by Diener et al [21], which includes 5 items, scored from 1 to 5, where higher values correspond to greater satisfaction with life) and loneliness (6-item De Jong Gierveld Loneliness Scale [22]; scored from 1 to 4, with higher values reflecting higher loneliness levels). In this study, Cronbach alpha for the SWLS was .87 and that for the De Jong Gierveld Loneliness Scale was .79. Both tools (loneliness [23] and life satisfaction [24]) have favorable psychometric properties.

Statistical Analysis

In a first step, sample characteristics were computed, stratified by the use of online consultations with doctors or therapists (unpaired *t* tests and chi-square tests were used, as appropriate). Subsequently, multiple logistic regressions were conducted to identify the correlates of use of online consultations with doctors or therapists. Statistical significance was set at $P<.05$. Stata 16.1 (StataCorp) was used to perform statistical analyses.

Ethics Approval

Written informed consent was obtained from all individual participants included in the study. An ethics vote was not deemed necessary according to criteria for the need of an ethical statement. This is in line with the German Research Foundation guidelines. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Results

Description of the Sample

Stratified by the use of online consultations with doctors or therapists, sample characteristics are presented in Table 1. In the total sample, the average age was 67.6 years (SD 9.7; range 46-98 years) and 49.31% (1877/3806) of the individuals were female. Overall, 10.02% (381/3806) of the individuals with access to the internet had online consultations with doctors or therapists.

In bivariate analysis, individuals using online consultations with doctors or therapists significantly (at least $P<.05$) differed from those who had never had an online consultation in terms of sex, living situation, presence of children, self-rated health, depressive symptoms, feeling that the COVID-19 crisis poses a threat to oneself, and loneliness. Further details are given in Table 1.

Table 1. Sample characteristics stratified by the use of online consultations with doctors or therapists (German Ageing Survey, fifth wave, n=3806).

Characteristics	Total (n=3806)[a]	Never had online consultations with doctors or therapists (n=3425)[a]	Have had online consultations with doctors or therapists (n=381)[a]	*P* value[b]
Sex				.03
Men	1929	1716 (88.96)	213 (11.04)	
Women	1877	1709 (91.05)	168 (8.95)	
Age (years), mean (SD)	67.6 (9.7)	67.6 (9.6)	68.0 (10.2)	.37
Educational level (ISCED-97[c] classification)				.10
Low (ISCED 0-2)	104	90 (86.54)	14 (13.46)	
Medium (ISCED 3-4)	1687	1536 (91.05)	151 (8.95)	
High (ISCED 5-6)	2014	1798 (89.28)	216 (10.72)	
Employment status				.78
Employed	1196	1083 (90.55)	113 (9.45)	
Retired	2412	2166 (89.80)	246 (10.20)	
Non-employed	143	129 (90.21)	14 (9.79)	
Living situation				.01
Single	737	685 (92.94)	52 (7.06)	
With partner in the same household	2851	2548 (89.37)	303 (10.63)	
With partner without a common household	168	149 (88.69)	19 (11.31)	
Having at least one child				.006
No	321	303 (94.39)	18 (5.61)	
Yes	3311	2964 (89.52)	347 (10.48)	
Migration background				.11
No	3623	3267 (90.17)	356 (9.83)	
Yes	177	153 (86.44)	24 (13.56)	
Household net income, mean (SD)	4175.8 (12,363.9)	4152.6 (12,778.0)	4384.9 (7690.7)	.74
Region				.21
West Germany	2742	2457 (89.61)	285 (10.39)	
East Germany	1064	968 (90.98)	96 (9.02)	
Type of district				.92
Large cities	1058	954 (90.17)	104 (9.83)	
Urban cities	1438	1293 (89.92)	145 (10.08)	
Urban-rural cities	770	696 (90.39)	74 (9.61)	
Rural districts	540	482 (89.26)	58 (10.74)	
Engagement in physical activities				.41
Daily	454	407 (89.65)	47 (10.35)	
Several times a week	1395	1244 (89.18)	151 (10.82)	
Once a week	703	629 (89.47)	74 (10.53)	
1-3 times a month	253	230 (90.91)	23 (9.09)	
Less often	676	613 (90.68)	63 (9.32)	
Never	303	282 (93.07)	21 (6.93)	
Frequency of walks				.33
Daily	712	634 (89.04)	78 (10.96)	

Characteristics	Total (n=3806)[a]	Never had online consultations with doctors or therapists (n=3425)[a]	Have had online consultations with doctors or therapists (n=381)[a]	*P* value[b]
Several times a week	1447	1292 (89.29)	155 (10.71)	
Once a week	651	598 (91.86)	53 (8.14)	
1-3 times a month	277	251 (90.61)	26 (9.39)	
Less often	572	514 (89.86)	58 (10.14)	
Never	122	114 (93.44)	8 (6.56)	
Self-rated health (scored from 1=very good to 5=very bad), mean (SD)	2.4 (0.8)	2.4 (0.8)	2.6 (0.8)	<.001
Depressive symptoms (10-item Center for Epidemiological Studies-Depression scale; scored from 0 to 30, with higher values reflecting more depressive symptoms), mean (SD)	8.2 (4.7)	8.1 (4.7)	9.2 (5.1)	<.001
Feeling that the COVID-19 crisis is a threat for oneself (scored from 0=not at all a threat for me to 10=extreme threat for me), mean (SD)	4.0 (2.1)	4.0 (2.1)	4.3 (2.3)	.02
Infection among people in one's own personal environment with coronavirus				.19
Yes	307	272 (88.60)	35 (11.40)	
No	3345	3022 (90.34)	323 (9.66)	
Don't know	129	111 (86.05)	18 (13.95)	
Own infection with the coronavirus				.79
Yes	17	15 (88.24)	2 (11.76)	
No	3547	3195 (90.08)	352 (9.92)	
Don't know	213	189 (88.73)	24 (11.27)	
Feeling that I can influence an infection with the coronavirus (scored from 1=not at all to 7=entirely), mean (SD)	4.6 (1.4)	4.6 (1.4)	4.5 (1.5)	.31
Loneliness (6-item De Jong Gierveld loneliness scale; scored from 1 to 4, with higher values reflecting higher loneliness levels), mean (SD)	1.9 (0.5)	1.9 (0.5)	2.0 (0.5)	<.001
Life satisfaction (Satisfaction with Life Scale; scored from 1 to 5, with higher values corresponding to greater satisfaction with life), mean (SD)	3.9 (0.7)	3.9 (0.7)	3.9 (0.7)	.67

[a]Data are presented as n or n (%) unless stated otherwise.

[b]*P* values were based on independent unpaired *t* tests and chi-square tests, as appropriate.

[c]ISCED: International Standard Classification of Education 97.

Regression Analysis

The findings of multiple logistic regressions with experience of an online consultation(s) with doctors or therapists as the outcome measure are shown in Table 2. Multiple logistic regressions showed that the likelihood of having online consultations with doctors or therapists (compared with never using such services) was positively associated with higher education (compared with medium education; odds ratio [OR] 1.31, 95% CI 1.01-1.70), living with a partner in the same household (compared with single; OR 1.53, 95% CI 1.05-2.22), poorer self-rated health (OR 1.42, 95% CI 1.16-1.74), increased loneliness (OR 1.45, 95% CI 1.10-1.90), and increased satisfaction with life (OR 1.30, 95% CI 1.03-1.64). Other socioeconomic, lifestyle-related, health-related, and COVID-19–related factors did not achieve statistical significance (*P*>.05).

Table 2. Determinants of online consultations with doctors or therapists. Results of multiple logistic regression analysis (German Ageing Survey, short survey).[a]

Independent variables	Online consultations with doctors or therapists, OR (95% CI)
Sex: Women (reference: men)	0.86 (0.66-1.11)
Age	1.01 (0.99-1.03)
Educational level (ISCED-97[b] classification)	
Low education (reference: medium education)	1.39 (0.66-2.92)
High education	1.31[c] (1.01-1.70)
Employment status	
Retired (reference: employed)	0.83 (0.55-1.24)
Non-employed	0.88 (0.44-1.76)
Living situation	
With partner in the same household (reference: single)	1.53[c] (1.05-2.22)
With partner without a common household	1.52 (0.79-2.93)
Having at least one child: Yes (reference: no)	1.58[d] (0.95-2.63)
Migration background: Yes (reference: no)	1.11 (0.63-1.95)
Household net income (in €1000)[e]	1.00 (0.99-1.01)
Region: East Germany (reference: West Germany)	0.88 (0.64-1.21)
Type of district	
Large cities (reference: rural districts)	0.84 (0.56-1.26)
Urban cities	0.77 (0.51-1.16)
Urban-rural cities	0.79 (0.52-1.20)
Engagement in physical activities	
Several times a week (reference: daily)	1.22 (0.79-1.87)
Once a week	1.17 (0.72-1.88)
1-3 times a month	0.87 (0.47-1.64)
Less often	0.89 (0.54-1.48)
Never	0.59 (0.30-1.18)
Frequency of walks	
Several times a week (reference: daily)	1.01 (0.71-1.44)
Once a week	0.81 (0.52-1.25)
1-3 times a month	0.87 (0.50-1.52)
Less often	1.12 (0.72-1.73)
Never	0.73 (0.29-1.82)
Self-rated health (from 1=very good to 5=very bad)	1.42[f] (1.16-1.74)
Depressive symptoms (10-item Center for Epidemiological Studies-Depression scale, from 0 to 30, with higher values reflecting more depressive symptoms)	1.02 (0.99-1.06)
Feeling that the COVID-19 crisis poses a threat to oneself (from 0=not at all a threat for me to 10=extreme threat for me)	1.03 (0.97-1.09)
Infection among people in one's own personal environment with coronavirus	
No (reference: yes)	0.86 (0.56-1.31)
Don't know	1.57 (0.78-3.18)
Personal experience of infection with the coronavirus	

Independent variables	Online consultations with doctors or therapists, OR (95% CI)
No (reference: yes)	0.08 (0.00-1.73)
Don't know	0.07[d] (0.00-1.59)
Feeling that one can influence an infection with the coronavirus (from 1=not at all to 7=entirely)	0.97 (0.89-1.07)
Loneliness (6-item De Jong Gierveld loneliness scale; from 1 to 4, with higher values reflecting higher loneliness levels)	1.45[g] (1.10-1.90)
Life satisfaction (Satisfaction with Life Scale; from 1 to 5, with higher values corresponding to greater satisfaction with life)	1.30[c] (1.03-1.64)
Constant	0.04[d] (0.00-1.52)
Observations	3067
Pseudo R^2	0.04

[a]Missing values were handled using listwise deletion. Outcome: 0=never using online consultations with doctors or therapists; 1=using online consultations with doctors or therapists.

[b]ISCED-97: International Standard Classification of Education 97.

[c]*P*<.05.

[d]*P*<.10.

[e]€1 = US $1.10.

[f]*P*<.001.

[g]*P*<.01.

Discussion

Principal Findings

In sum, 10.02% (381/3806) of the individuals with access to the internet had used online consultations with doctors or therapists. Regressions showed that the likelihood of having used online consultations with doctors or therapists (compared with never using such services) was positively associated with higher education, living with partner in the same household (compared with single), poorer self-rated health, increased loneliness, and increased satisfaction with life.

Previous Research and Possible Explanations

To date, only a few international studies have reported the frequency of online consultations with doctors or therapists *based on nationally representative samples*. For example, 46.1% of GP services were provided using such consultations (video and telephone) in Australia in early May 2020 [25]. Moreover, 44% of respondents supported using online consultations for medication abortion among adults in the United States during the COVID-19 pandemic [26]. In general, the quite high proportion of individuals using online consultations with doctors or therapists (compared with the time prior to the pandemic in Germany) found in our study is in accordance with recently conducted research covering the adult population in Germany during the pandemic [27]. Because of the high proportion of online consultations, we assume that, in most cases, online consultations replaced physical meetings. However, because these data were not available in the data set used in our study, future research is required to clearly distinguish between different types of online consultations with doctors or therapists (ie, replacing or complementary to in-person consultations).

This high proportion supports the conclusion made by Wosik et al [28], who stressed the rise of virtual care. In international comparison (eg, compared with other European countries), Germany still lags behind [29]. By contrast, consumer-enabled and connected health technologies are widespread in the Netherland and Nordic countries.

To date, some studies have described the frequency and correlates of online consultations with doctors or therapists. For example, it has recently been shown that older, female, or poorer patients had used video consultations less frequently [30] in the United States. A greater reluctance to use online consultations with doctors was also reported among individuals with a lower educational attainment in Denmark [31].

Given that individuals filled out the questionnaire in June or July 2020, it appears plausible that those with poorer self-rated health had a greater likelihood of using online consultations, compared with those who did not use online consultations, as these individuals were in need of care and around 1.4 million hours of online consultations were made in Germany during the first half of 2020 [12].

Compared with single individuals, individuals living with a partner in the same household had a greater likelihood of using online consultations. This might be explained by the fact that partners may urge the individuals to use such services [32] when individuals are in need of care. Moreover, they could assist when technical difficulties arise. Furthermore, they may seek to protect their partner from possible infection with SARS-CoV-2.

The association between higher education and an increased likelihood of using online consultations appears to be plausible. This is due to the fact that higher education is often associated with lower computer anxiety and better computer skills [33].

With regard to other correlates, it was particularly surprising that COVID-19–related factors were not significantly associated with the outcome measure. It seems that other factors (ie, education, health, and psychosocial factors) are important for the outcome measure. Future research is required to clarify the association between COVID-19–related factors and the use of online consultations with doctors or therapists.

At first glance, it may seem contradictory that both higher loneliness and greater life satisfaction are associated with a greater likelihood of using online consultations. However, previous research has also demonstrated a link between increased loneliness and an increased number of GP visits [34]—for example, to address social needs in higher age [35]. Furthermore, the association between life satisfaction and online consultations may be explained by the fact that life satisfaction is positively associated with an increased use of preventive health care services (eg, in women: higher likelihood of obtaining a mammogram, x-ray, or pap smear; in men: higher likelihood of obtaining a prostate examination) [36]. Additionally, higher life satisfaction is associated with higher meaning in life (ie, a sense of comprehension and significance in life [37]) [38]. In turn, a higher meaning in life [39] is associated with more frequent health care use (GP and specialist visits) and an increased use of preventive health care services [40]. Individuals with high life satisfaction and high meaning in life may particularly value their lives and may use health care services (curative and preventive) to stay healthy for as long as possible [40].

Strengths and Limitations

Some strengths and limitations of this study should be considered. This is the first study identifying the frequency and correlates of online consultations with doctors or therapists among individuals in the second half of life in Germany. Additionally, a nationally representative sample (data collection during the pandemic) was used. While the outcome measure had a high face validity, future studies are needed to distinguish between online consultations with doctors (including the medical specialty) or therapists (eg, psychotherapists, occupational therapists, or physical therapists). For example, it may be the case that online consultations with doctors differ from online consultation with therapists (eg, in terms of the reason for the consultation, the objective, and the average duration). Online consultations may also differ between, for example, physicians and psychological or medical psychotherapists. Differences may also exist between occupational therapists and physical therapists regarding online consultations. Moreover, other domains of health care use (eg, outpatient physician visits or hospital stays) and other health-related factors (eg, chronic conditions) were not assessed in this short survey. A small sample selection bias has been identified in the DEAS study [41]. However, the distribution of family status, family composition, labor force participation, and educational level is very close to the distribution in the German population [41]. Furthermore, this study examined patient-related factors, whereas more studies are required to investigate physician-related factors such as practice size [42] or region (rural vs urban [43]) when identifying the correlates of online consultations with doctors or therapists.

Conclusions

Findings of this study suggest that a non-negligible proportion of around 1 in 10 individuals aged 46 years and over had online consultations with doctors or therapists. However, compared with other countries, there is still room for improvement (regarding the proportion of online consultations by community-dwelling individuals aged 46 years and over in Germany). Examination and comparison of the characteristics of adopters and nonadopters of online consultations could assist in strategic planning and improve uptake of online consultation. Future research in other countries is required.

With regard to Germany, it may be beneficial to strongly intensify efforts linked to broadband infrastructure in Germany, which has clear potential for improvement and still lags behind other countries. A good quality and stable connection, as well as perhaps monetary incentives for both patients (eg, cost savings) and doctors/therapists (eg, remuneration incentives) may assist in increasing the proportion of online consultations. Furthermore, strategies to reduce computer anxiety, particularly among the oldest old, may assist in this.

Authors' Contributions

AH and H-HK contributed to conceptualization of the study. AH performed data curation, project administration, and formal analysis; and contributed to methodology and wrote the first draft of the manuscript. H-HK was responsible for resources and supervision. Both authors contributed to manuscript revision and editing, read, and approved the submitted version.

Conflicts of Interest

None declared.

References

1. Hyder MA, Razzak J. Telemedicine in the United States: An Introduction for Students and Residents. J Med Internet Res 2020 Nov 24;22(11):e20839 [FREE Full text] [doi: 10.2196/20839] [Medline: 33215999]
2. Zaresani A, Scott A. Does digital health technology improve physicians' job satisfaction and work-life balance? A cross-sectional national survey and regression analysis using an instrumental variable. BMJ Open 2020 Dec 12;10(12):e041690 [FREE Full text] [doi: 10.1136/bmjopen-2020-041690] [Medline: 33310807]
3. Elsner P. Teledermatology in the times of COVID-19 - a systematic review. J Dtsch Dermatol Ges 2020 Aug;18(8):841-845. [doi: 10.1111/ddg.14180] [Medline: 33448667]

4. Kayser M, Valtin C, Greer M, Karow B, Fuge J, Gottlieb J. Video Consultation During the COVID-19 Pandemic: A Single Center's Experience with Lung Transplant Recipients. Telemed J E Health 2021;27(7):807-815. [doi: 10.1101/2020.05.12.20093799]
5. Petersen W, Karpinski K, Backhaus L, Bierke S, Häner M. A systematic review about telemedicine in orthopedics. Arch Orthop Trauma Surg 2021 Oct 26;141(10):1731-1739 [FREE Full text] [doi: 10.1007/s00402-021-03788-1] [Medline: 33635399]
6. Gerbutavicius R, Brandlhuber U, Glück S, Kortüm GF, Kortüm I, Navarrete Orozco R, et al. Evaluation of patient satisfaction with an ophthalmology video consultation during the COVID-19 pandemic. Ophthalmologe 2021 Jan 10;118(Suppl 1):89-95 [FREE Full text] [doi: 10.1007/s00347-020-01286-0] [Medline: 33301067]
7. Gesamtzahl der Ärzte. Bundesärztekammer. URL: https://www.bundesaerztekammer.de/ueber-uns/aerztestatistik/gesamtzahl-der-aerzte/ [accessed 2022-03-28]
8. Bundesärztekammer. Berufstätige Ärzte. URL: https://www.bundesaerztekammer.de/ueber-uns/aerztestatistik/berufstaetige-aerzte/ [accessed 2022-03-28]
9. Statistisches Bundesamt. Number of psychotherapists up 19% between 2015 and 2019. Statistisches Bundesamt. URL: https://www.destatis.de/EN/Press/2021/03/PE21_N022_23.html;jsessionid=FD8FDDB80A72F708482756082CE7053B.live742 [accessed 2022-03-28]
10. PhysioDeutschland. Zahlen, Daten, Fakten zur Physiotherapie. PhysioDeutschland. URL: https://tinyurl.com/3fk5nua3 [accessed 2022-03-28]
11. Deutscher Verband der Ergotherapeuten e.V.. Die Ergotherapie in Deutschland. Karlsbad, Germany: Deutscher Verband der Ergotherapeuten e. V; 2017.
12. Kassenärztliche Bundesvereinigung. Immer mehr Praxen greifen zur Kamera - Zahl der Videosprechstunden auf über eine Million gestiegen. Kassenärztliche Bundesvereinigung. URL: https://www.kbv.de/html/1150_50419.php [accessed 2022-03-28]
13. Allensbach ifD. Ergebnisse einer Repräsentativbefragung von niedergelassenen und Krankenhausärzten im April/Mai 2010. In. In: Allensbach am Bodensee. Allensbach, Germany: Institut für Demoskopie Allensbach; 2010.
14. Welle K, Täger S, Hackenberg RK, Markowetz A, Schildberg FA, Burger C, et al. Examining the Hand in the Video Consultation. Z Orthop Unfall 2021 Apr 21;159(2):202-208 [FREE Full text] [doi: 10.1055/a-1304-3593] [Medline: 33477179]
15. Neller K. Kooperation und verweigerung: eine non-response-studie [Co-operation and refusal: a non-response study]. ZUMA Nachrichten 2005;29(57):9-36 [FREE Full text]
16. Pepinsky TB. A Note on Listwise Deletion versus Multiple Imputation. Polit. Anal 2018 Aug 03;26(4):480-488. [doi: 10.1017/pan.2018.18]
17. Klaus D. Daten und Methoden des Deutschen Alterssurveys. In: Mahne K, Wolff JK, Simonson J, Tesch-Römer C, editors. Altern im Wandel: Zwei Jahrzehnte Deutscher Alterssurvey (DEAS). Berlin, Germany: DZA; 2016:25-42.
18. UNESCO. International Standard Classification of Education. In: ISCED 1997, Re-edition. Paris, France: UNESCO; 2006.
19. Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for Depression in Well Older Adults: Evaluation of a Short Form of the CES-D. American Journal of Preventive Medicine 1994 Mar;10(2):77-84. [doi: 10.1016/s0749-3797(18)30622-6]
20. Mohebbi M, Nguyen V, McNeil JJ, Woods RL, Nelson MR, Shah RC, ASPREE Investigator Group. Psychometric properties of a short form of the Center for Epidemiologic Studies Depression (CES-D-10) scale for screening depressive symptoms in healthy community dwelling older adults. Gen Hosp Psychiatry 2018 Mar;51:118-125 [FREE Full text] [doi: 10.1016/j.genhosppsych.2017.08.002] [Medline: 28890280]
21. Diener E, Emmons RA, Larsen RJ, Griffin S. The Satisfaction With Life Scale. J Pers Assess 1985 Feb;49(1):71-75. [doi: 10.1207/s15327752jpa4901_13] [Medline: 16367493]
22. Gierveld JDJ, Tilburg TV. A 6-Item Scale for Overall, Emotional, and Social Loneliness. Res Aging 2016 Aug 18;28(5):582-598. [doi: 10.1177/0164027506289723]
23. De Jong Gierveld J, Van Tilburg T. The De Jong Gierveld short scales for emotional and social loneliness: tested on data from 7 countries in the UN generations and gender surveys. Eur J Ageing 2010 Jun 9;7(2):121-130 [FREE Full text] [doi: 10.1007/s10433-010-0144-6] [Medline: 20730083]
24. Pavot W, Diener E, Colvin CR, Sandvik E. Further validation of the Satisfaction with Life Scale: evidence for the cross-method convergence of well-being measures. J Pers Assess 1991 Aug;57(1):149-161. [doi: 10.1207/s15327752jpa5701_17] [Medline: 1920028]
25. Scott A, Bai T, Zhang Y. Association between telehealth use and general practitioner characteristics during COVID-19: findings from a nationally representative survey of Australian doctors. BMJ Open 2021 Mar 24;11(3):e046857 [FREE Full text] [doi: 10.1136/bmjopen-2020-046857] [Medline: 33762248]
26. LaRoche KJ, Jozkowski KN, Crawford BL, Haus KR. Attitudes of US adults toward using telemedicine to prescribe medication abortion during COVID-19: A mixed methods study. Contraception 2021 Jul;104(1):104-110 [FREE Full text] [doi: 10.1016/j.contraception.2021.04.001] [Medline: 33848466]

27. Reitzle L, Schmidt C, Färber F, Huebl L, Wieler LH, Ziese T, et al. Perceived Access to Health Care Services and Relevance of Telemedicine during the COVID-19 Pandemic in Germany. Int J Environ Res Public Health 2021 Jul 19;18(14):7661 [FREE Full text] [doi: 10.3390/ijerph18147661] [Medline: 34300110]
28. Wosik J, Fudim M, Cameron B, Gellad Z, Cho A, Phinney D, et al. Telehealth transformation: COVID-19 and the rise of virtual care. J Am Med Inform Assoc 2020 Jun 01;27(6):957-962 [FREE Full text] [doi: 10.1093/jamia/ocaa067] [Medline: 32311034]
29. HIMSS. e-Health Trendbarometer - Consumer-enabled and Connected Health in Europe. Chicago, IL: HIMSS; 2021.
30. Eberly LA, Kallan MJ, Julien HM, Haynes N, Khatana SAM, Nathan AS, et al. Patient Characteristics Associated With Telemedicine Access for Primary and Specialty Ambulatory Care During the COVID-19 Pandemic. JAMA Netw Open 2020 Dec 01;3(12):e2031640 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.31640] [Medline: 33372974]
31. Sørensen JFL. Attitudes toward telehealth use among rural residents: a Danish survey. J Rural Health 2008;24(3):330-335. [doi: 10.1111/j.1748-0361.2008.00178.x] [Medline: 18643814]
32. Hajek A, Bock J, König HH. The role of personality in health care use: Results of a population-based longitudinal study in Germany. PLoS One 2017 Jul 26;12(7):e0181716 [FREE Full text] [doi: 10.1371/journal.pone.0181716] [Medline: 28746388]
33. Agarwal R, Prasad J. Are Individual Differences Germane to the Acceptance of New Information Technologies? Decision Sciences 1999 Mar;30(2):361-391. [doi: 10.1111/j.1540-5915.1999.tb01614.x]
34. Burns A, Leavey G, Ward M, O'Sullivan R. The impact of loneliness on healthcare use in older people: evidence from a nationally representative cohort. J Public Health (Berl.) 2020 Jun 22;30(3):675-684. [doi: 10.1007/s10389-020-01338-4]
35. Hajek A, Brettschneider C, Eisele M, Lühmann D, Mamone S, Wiese B, AgeCoDeAgeQualiDe Study Group. Does transpersonal trust moderate the association between chronic conditions and general practitioner visits in the oldest old? Results of the AgeCoDe and AgeQualiDe study. Geriatr Gerontol Int 2019 Aug 24;19(8):705-710. [doi: 10.1111/ggi.13693] [Medline: 31237101]
36. Kim ES, Kubzansky LD, Smith J. Life satisfaction and use of preventive health care services. Health Psychol 2015 Jul;34(7):779-782 [FREE Full text] [doi: 10.1037/hea0000174] [Medline: 25420064]
37. George LS, Park CL. Are meaning and purpose distinct? An examination of correlates and predictors. The Journal of Positive Psychology 2013 Sep;8(5):365-375. [doi: 10.1080/17439760.2013.805801]
38. Karataş Z, Uzun K, Tagay. Relationships Between the Life Satisfaction, Meaning in Life, Hope and COVID-19 Fear for Turkish Adults During the COVID-19 Outbreak. Front Psychol 2021 Mar 11;12:633384 [FREE Full text] [doi: 10.3389/fpsyg.2021.633384] [Medline: 33776856]
39. Hajek A, König HH. Meaning in life and health care use: findings from a nationally representative study of older adults in Germany. BMC Geriatr 2019 Dec 23;19(1):368 [FREE Full text] [doi: 10.1186/s12877-019-1389-3] [Medline: 31870312]
40. Kim ES, Strecher VJ, Ryff CD. Purpose in life and use of preventive health care services. Proc Natl Acad Sci U S A 2014 Nov 18;111(46):16331-16336 [FREE Full text] [doi: 10.1073/pnas.1414826111] [Medline: 25368165]
41. Klaus D, Engstler H, Mahne K, Wolff J, Simonson J, Wurm S, et al. Cohort Profile: The German Ageing Survey (DEAS). Int J Epidemiol 2017 Aug 01;46(4):1105-1105g [FREE Full text] [doi: 10.1093/ije/dyw326] [Medline: 28180273]
42. Kane CK, Gillis K. The Use Of Telemedicine By Physicians: Still The Exception Rather Than The Rule. Health Aff (Millwood) 2018 Dec;37(12):1923-1930. [doi: 10.1377/hlthaff.2018.05077] [Medline: 30633670]
43. Jetty A, Moore MA, Coffman M, Petterson S, Bazemore A. Rural Family Physicians Are Twice as Likely to Use Telehealth as Urban Family Physicians. Telemed J E Health 2018 Apr;24(4):268-276. [doi: 10.1089/tmj.2017.0161] [Medline: 28805545]

Abbreviations

CES-D: Center for Epidemiological Studies-Depression
DEAS: German Ageing Survey
EBM: outpatient medical fee schedule
GOÄ: medical fee schedule used for privately insured patients
GP: general practitioner
ISCED-97: International Standard Classification of Education 97
KBV: National Association of Statutory Health Insurance Physicians
OR: odds ratio
SWLS: Satisfaction with Life Scale

Edited by A Mavragani; submitted 20.04.21; peer-reviewed by S Hajesmaeel Gohari, R Halkes, KC Wong; comments to author 30.07.21; revised version received 18.11.21; accepted 24.02.22; published 07.04.22.

Please cite as:
Hajek A, König HH
Frequency and Correlates of Online Consultations With Doctors or Therapists in Middle-Aged and Older Adults: Nationally Representative Cross-sectional Study
J Med Internet Res 2022;24(4):e29781
URL: https://www.jmir.org/2022/4/e29781
doi:10.2196/29781
PMID:35389360

Original Paper

Gap in Willingness and Access to Video Visit Use Among Older High-risk Veterans: Cross-sectional Study

Stuti Dang[1,2,3], MPH, MD; Kiranmayee Muralidhar[4], MBBS, MPH; Shirley Li[5], BS; Fei Tang[1], PhD; Michael Mintzer[1], MD; Jorge Ruiz[1,2], MD; Willy Marcos Valencia[6], MSc, MD

[1]Geriatric Research, Education and Clinical Center, Miami Veterans Affairs Healthcare System, Miami, FL, United States

[2]Division of Geriatrics and Palliative Care, Miller School of Medicine, University of Miami, Miami, FL, United States

[3]The Elizabeth Dole Center of Excellence for Veteran and Caregiver Research, Miami, FL, United States

[4]Department of Epidemiology, Miller School of Medicine, University of Miami, Miami, FL, United States

[5]Miller School of Medicine, University of Miami, Miami, FL, United States

[6]Division of Endocrinology, Diabetes and Metabolic Diseases, Department of Medicine, Medical University of South Carolina, Charleston, SC, United States

Corresponding Author:
Stuti Dang, MPH, MD
Geriatric Research, Education and Clinical Center
Miami Veterans Affairs Healthcare System
1201 NW 16th St
11 GRC
Miami, FL, 33125
United States
Phone: 1 305 575 3388
Email: stuti.dang@va.gov

Abstract

Background: The recent shift to video care has exacerbated disparities in health care access, especially among *high-need, high-risk (HNHR)* adults. Developing data-driven approaches to improve access to care necessitates a deeper understanding of HNHR adults' attitudes toward telemedicine and technology access.

Objective: This study aims to identify the willingness, access, and ability of HNHR veterans to use telemedicine for health care.

Methods: WWe designed a questionnaire conducted via mail or telephone or in person. Among HNHR veterans who were identified using predictive modeling with national Veterans Affairs data, we assessed willingness to use video visits for health care, access to necessary equipment, and comfort with using technology. We evaluated physical health, including frailty, physical function, performance of activities of daily living (ADL) and instrumental ADL (IADL); mental health; and social needs, including Area Deprivation Index, transportation, social support, and social isolation.

Results: The average age of the 602 HNHR veteran respondents was 70.6 (SD 9.2; range 39-100) years; 99.7% (600/602) of the respondents were male, 61% (367/602) were White, 36% (217/602) were African American, 17.3% (104/602) were Hispanic, 31.2% (188/602) held at least an associate degree, and 48.2% (290/602) were confident filling medical forms. Of the 602 respondents, 327 (54.3%) reported willingness for video visits, whereas 275 (45.7%) were unwilling. Willing veterans were younger (P<.001) and more likely to have an associate degree (P=.002), be health literate (P<.001), live in socioeconomically advantaged neighborhoods (P=.048), be independent in IADLs (P=.02), and be in better physical health (P=.04). A higher number of those willing were able to use the internet and email (P<.001). Of the willing veterans, 75.8% (248/327) had a video-capable device. Those with video-capable technology were younger (P=.004), had higher health literacy (P=.01), were less likely to be African American (P=.007), were more independent in ADLs (P=.005) and IADLs (P=.04), and were more adept at using the internet and email than those without the needed technology (P<.001). Age, confidence in filling forms, general health, and internet use were significantly associated with willingness to use video visits.

Conclusions: Approximately half of the HNHR respondents were unwilling for video visits and a quarter of those willing lacked requisite technology. The gap between those willing and without requisite technology is greater among older, less health literate, African American veterans; those with worse physical health; and those living in more socioeconomically disadvantaged

neighborhoods. Our study highlights that HNHR veterans have complex needs, which risk being exacerbated by the video care shift. Although technology holds vast potential to improve health care access, certain vulnerable populations are less likely to engage, or have access to, technology. Therefore, targeted interventions are needed to address this inequity, especially among HNHR older adults.

(J Med Internet Res 2022;24(4):e32570) doi:10.2196/32570

KEYWORDS

high-risk veterans; older adults; telemedicine; video visits; health disparities; Area Deprivation Index; mobile phone

Introduction

Background

The onset of the COVID-19 pandemic led to sudden and dramatic changes in the delivery of health care in the context of social distancing and lockdown decisions. Telemedicine has emerged as a solution to caring for patients who are medically complex during the pandemic [1]. Institutions have diverted resources toward purchasing necessary telemedicine equipment and expansion of technological infrastructure and hastily implemented telemedicine training sessions for providers [2,3]. Telemedicine reimbursement models also saw formula adjustments. For example, the Centers for Medicare and Medicaid Services (CMS) insurance models changed in March 2020 to reduce the costs of telemedicine [4], and the CMS issued waivers that allowed providers to care for patients remotely without financial penalties [5]. These factors have contributed to the accelerated implementation of telemedicine across health care systems [2,3].

The Veterans Affairs (VA) has been a leader in integrating the use of technology into health care. The implementation of telemedicine technologies and new programs at the VA has accelerated in recent years to expand access to more veterans. Since 2018, the VA's *Anywhere to Anywhere* initiative expanded the scope of telehealth so that care can be delivered via telehealth across state borders and even in the veterans' homes [6]. During the COVID-19 pandemic, similar to other health care systems, the VA moved rapidly to leverage its telemedicine capabilities to provide needed care to veterans at home [2,7]. A major pivot by the VA during the COVID-19 pandemic was the rapid adoption and use of the VA's telemedicine platform, Veteran Video Connect (VVC), which allowed most visits to be done via telemedicine at home. VVC is a videoconferencing application for veterans and their providers. It securely connects veterans to their health care team from any internet-enabled computer, tablet, or mobile device. In the face of this public health emergency, the VA also suspended previous Health Insurance Portability and Accountability Act compliance requirements to allow providers to connect with patients on non–public-facing technology if VVC was not working or at overcapacity [8].

Nevertheless, despite the rapid pivot to telemedicine, there have been valid concerns regarding patient-level challenges to wider implementation and integration of technology into health care. Using 2018 data from the National Health and Aging Trends Study of community-dwelling adults, Lam et al [9] estimated that approximately one-third of the older adults in the United States were not ready for video visits, which is largely attributed to inexperience with technology. Individuals who face barriers to accessing care in person are also likely the same individuals who face challenges accessing telemedicine and include those who are older and minority; have lower educational attainment, lower income, and self-reported poor health status [9-11]; and live in rural areas [9,12-14]. Therefore, disparities in health care access risk exacerbation by the ongoing shift to adopt telemedicine [9,11], especially among the highest risk patients with the most complex clinical scenarios [9,12].

Objective

To develop data-driven approaches and understand how best to deploy telemedicine to increase access to care for older adults who are complex and frail, it would be beneficial to form a deeper understanding of their attitude toward using video visits for receiving health care. Using a population health approach, the VA identifies a subgroup of veterans called *high-need, high-risk (HNHR)* veterans, who represent the VA population that would qualify for Medicare's demonstration of home-based primary care (HBPC; ie, independence at home) [15]. The primary aim of this study is to evaluate HNHR older veterans' willingness, access, and ability to use video visits for health care purposes. Our secondary aim is to characterize the willingness for telemedicine in the context of their physical, emotional, and social determinants. Our hypothesis is that among HNHR older adults, the access and ability to use video visits would be lower than that shown previously among community-dwelling adults [9].

Ultimately, this paper seeks to add to the ongoing efforts to provide actionable data that may help health care systems leverage telemedicine as a means of increasing access to health care. We can expect the increased reliance on telemedicine to be sustained, and increasing our understanding of the factors contributing to digital disparities will help identify targeted interventions to address the identified challenges to telemedicine for HNHR patients, who are also the patients most likely in need of support.

Methods

Overview

This cross-sectional observational study was part of a larger quality improvement study to better define the needs of HNHR veterans in the Miami VA Healthcare System. Here, we analyzed the willingness, technology access, and ability to use video visits in the HNHR veteran group.

Study Population

The VA Geriatrics and Extended Care Data Analysis Center uses population health VA data to identify HNHR veterans who are medically complex and functionally impaired and at the highest risk for hospitalization and long-term institutionalization and, therefore, eligible for HBPC. The criteria for the Geriatrics and Extended Care Data Analysis Center HNHR designation include hospitalization in the prior 12 months and medical complexity measures that include the 13-condition JEN Frailty Index (JFI) [16] score ≥6, suggesting dependency in ≥2 activities of daily living (ADL), and NOSOS (VA version of the CMS measure to project cost). Patients were excluded if they had end-stage renal disease; were enrolled in HBPC or medical foster home; had received hospice, palliative care, or nursing home care in the past 12 months; or lived >60 minutes away from the closest VA primary care site as VA HBPC programs were less likely to be available at this distance [15].

Over a 1-year period that extended from October 2017 to September 2018, 2543 Miami VA Healthcare System veterans were listed as HNHR. Of those 2543 veterans, 1300 (51.12%) were randomly selected and sent a questionnaire via the US Postal Service. The mailings were sent in two waves: May 2018 and November 2018. The questionnaires were conducted by mail only once, with no reminders to improve the response rate. An additional group of 173 HNHR veterans scheduled for a geriatric frailty clinic appointment completed the questionnaire.

Questionnaire Design and Variables

We designed a questionnaire to assess physical health, including frailty—with the Fatigue, Resistance, Ambulation, Illnesses, and Loss of Weight scale [17]—physical function, mobility, ADL [18], instrumental ADL (IADL) [19], and homebound status [20]; assess mental health using the Patient Health Questionnaire [21] for depression screening and perception of aging [22]; and assess social support, social isolation [23], and transportation. We assessed for willingness to use video visits for VA health care; among those willing to use video visits, we asked about access to the video-capable technology. Furthermore, we assessed the ability to use technology by asking about comfort in performing an internet search and using email. We also asked about My Health*e*Vet use and access and the desired mode of communication with VA. The used questions were either study specific, validated, or modified from validated questions. The details of the questionnaire are presented in Table 1. We have tried to segment and label our variables into those that relate to the level of the patient's need for telehealth versus barriers and facilitators that we can do something about, although this distinction is somewhat arbitrary and case dependent, as only some of the factors are addressable some of the time. Physical and mental health characteristics may often relate to the level of patient need for telehealth but may also present a barrier, whereas the social and technology characteristics are the surrounding factors that act as facilitators or barriers, depending on the situation.

Table 1. Survey components.

Indicator	Source	Details
Demographics		
Education	Study specific	Highest level of education completed
Health literacy [24]	Question to identify patients with inadequate health literacy	Confidence filling medical forms; score ranged from 1 to 5, with a higher score indicating more confidence; a score of 5 was considered health literate
Physical health (need or barrier)		
Frailty [17]	5-item FRAIL[a] scale	The 5-item FRAIL scale includes fatigue, resistance, ambulation, illness, and weight loss. The final score ranges from 0 to 5 and represents frail (score 3-5), prefrail (score 1-2), and robust (score 0) health status. A score of 3 to 5 was considered a positive screen.
General health [25]	Modified from the Stanford Chronic Disease Self-Management Program Questionnaire	Self-rated general health; scores ranged from 1 to 5, with a higher score indicating better self-rated general health
Self-rated physical status	Self-rated physical status	Scores for self-rated physical status ranged from 1 to 10, with a higher score indicating better physical status
Walking, falls, and exercise	Study specific	Issues with walking, stepping, and balance; assistive devices used; number of falls in the past year; barriers to exercise; pedometer use
ADL[b] [18]	Barthel index for ADL	Barthel ADL score (range 0-100), with a higher score indicating greater independence
IADL[c] [19]	Lawton score for IADL	Lawton IADL score (range 0-8), with a higher score indicating greater independence
Homebound status [20]	Determining homebound status as part of a mobility questionnaire using validated questions from the National Health and Aging Trends Study	Individuals were categorized as homebound, semihomebound, and not homebound based on their responses to how often they left their home, how much help they had in leaving their home, and how much difficulty they had in leaving their home in the previous month, similar to the reference study.
Mental health (need or barrier)		
Depression screen [21]	PHQ-2[d]	PHQ-2 scores ranged from 0 to 6; a score ≥3 is considered positive for the likelihood of depression
Self-perception of aging [22]	Attitude Toward Own Aging subscale of the Philadelphia Geriatric Center Morale Scale	The 5-question scale (range 0-5) was treated as a binary variable. For the first (*feeling worse as I get older*) and third (*feeling useless as I get older*) questions on the scale, the responses *strongly disagree, disagree, somewhat disagree* were scored as 0, whereas the responses *somewhat agree, agree, strongly agree* were scored as 1. The responses to the second (*as much pep as last year*), fourth (*as happy as when I was younger*), and fifth (*things are better than I thought it would be*) questions were scored in a reverse manner. A higher score indicated a negative perception of aging.
Social characteristics (facilitator or barrier)		
Social support	Study specific	Having a formal or informal caregiver; caregiver's distance from home
Social isolation [23]	Berkman–Syme Social Network Index	Scoring was performed as the following: married (no=0; yes=1), meeting and talking to close friends and relatives (<3 times a week=0; ≥3 times a week=1), participation in religious meetings or services (<4 times a year=0; ≥4 times a year=1), and attend meetings of the clubs or organizations (never or does not belong=0, all the responses=1). Scores were summed: 0 or 1 being the most isolated category, and 2, 3, or 4 formed the other 3 categories of increasing social integration.
Transportation [26]	Questions assessing transportation barriers	Trouble with transportation, delayed physicians' appointments because of transportation troubles, and travel time from home to their physician
Technology (facilitator or barrier)		
Technology willingness, access, and ability	Study specific	Willingness to use video visits with VA[e] providers; access to video-capable equipment among those willing to use video visits; ability to do an internet search and use email; My HealtheVet enrollment and use; preferred mode of contact

[a]FRAIL: Fatigue, Resistance, Ambulation, Illnesses, and Loss of Weight.

[b]ADL: activities of daily living.

[c]IADL: instrumental activities of daily living.

[d]PHQ-2: Patient Health Questionnaire-2.

[e]VA: Veterans Affairs.

Additional measures obtained from VA records included the Care Assessment Needs score (VA measure for hospitalization and mortality risk) [27] and the Hierarchical Condition Categories score [28]. We also obtained the Area Deprivation Index (ADI), an established measure of socioeconomic disadvantage at the census tract level, from the Neighborhood Atlas [29].

Statistical Analysis

Descriptive characteristics were presented as frequency (percentage) for categorical variables and as mean (SD) for continuous variables. We compared the characteristics of respondents who were willing to use video visits with those who were not; among those willing to use video visits, we further compared those with and without self-reported access to video-capable technology. The chi-square test was used for comparing categorical variables, and the 2-tailed *t* test was used for comparing continuous variables. We reported all *P* values and considered them to be significant when <.05. Multivariable logistic regression was conducted to identify predictors for willingness to use video visits. All statistical analyses were performed using SAS (version 9.4; SAS Institute, Inc).

Ethical Considerations

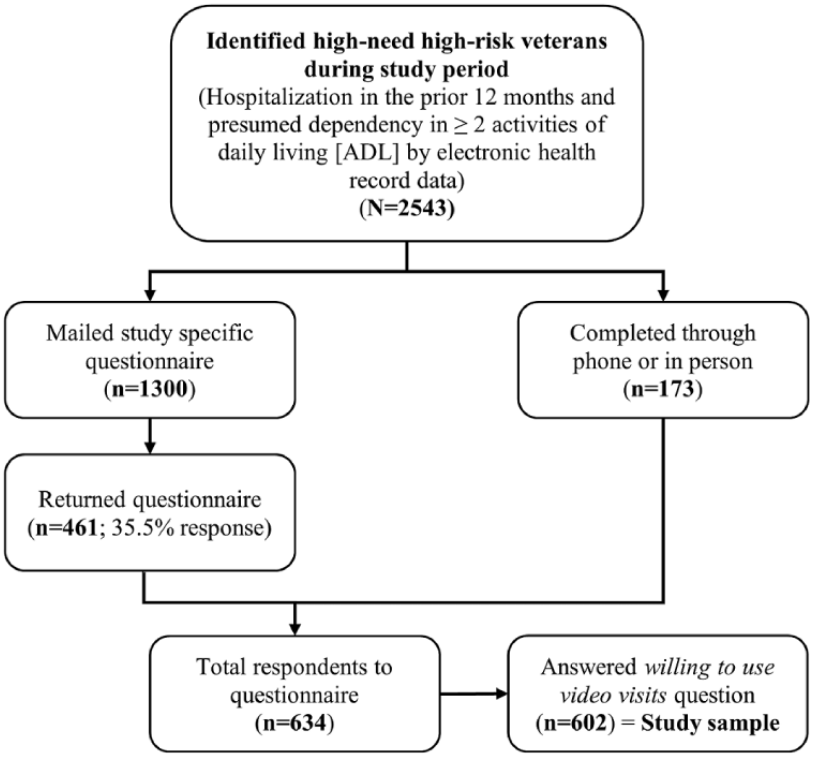
The Miami VA institutional review board granted this study a waiver and deemed it as a quality improvement study (reference number 1360043-3).

Results

Survey Respondents

A total of 1300 HNHR veterans were mailed the questionnaire, of which 461 (35.46%) were returned. In addition, 102 veterans filled the questionnaire over the phone and 71 in person in the frailty clinic, for a total of 634 respondents. Of the 634 individuals returning the survey, 602 (94.9%) respondents answered the *willing to use video visits* question. These 602 respondents represent the main focus of our study (Figure 1). When asked about their willingness to use video visits with their VA care team, 54.3% (327/602) reported their willingness, henceforth labeled as *willing*, whereas 45.7% (275/602) were not willing to use video visits, henceforth labeled as *unwilling*.

Figure 1. Flowchart showing completed questionnaires.



The average age of our 602 respondents was 70.6 (SD 9.2; range 39-100) years. Among them, 20.3% (122/602) were aged <65 years, 25.4% (153/602) were aged 65 to 69 years, 25.7% (155/602) were aged 70 to 75 years, 13.8% (83/602) were aged 75 to 79 years, and 14.8% (89/602) were aged >80 years. Only 0.3% (2/602) of the respondents were female. Approximately 61% (367/602) of the respondents were White, 36% (217/602) respondents were African American, and 17.3% (104/602) were Hispanic. Among the 602 responders, 290 (48.2%) were confident filling medical forms by themselves, and 188 (31.2%) had at least an associate degree.

Difference Between Respondents by Mail versus In Person and Telephone

Individuals completing the survey via mail were significantly more confident filling out medical forms (224/440, 50.9% vs 66/162, 40.7%; *P*=.03); in significantly worse physical health, as measured by their JFI (mean 7.2, SD 1.2 vs mean 6.8, SD 1.2; *P*<.001), Care Assessment Needs scores (mean 94.2, SD 6.8 vs mean 91.6, SD 8.1; *P*<.001), and the total number of Hierarchal Condition Categories conditions (mean 5.7, SD 2.4 vs mean 4.9, SD 1.9; *P*<.001); and significantly more socially

isolated (Social Networking Index of 1.5, SD 1.1 vs 1.7, SD 1.1; *P*=.049). There were no other differences between those veterans who completed the mailed survey versus those completing the survey by phone or in person.

Furthermore, we compared the willingness to use video visits between veterans who finished the survey in person and those who did not, and the difference was not significant (*P*=.13). Although more veterans reported no trouble for transportation in those who filled out the survey in person (52/80, 65%) than those who did not (321/531, 60.5%), the difference was not statistically significant (*P*=.05). Similarly, the difference in the percentage of veterans who missed an appointment owing to transportation between those who filled out the survey in person and not in person was not significant (*P*=.28).

Difference Between Respondents Who Were Willing Versus Unwilling to Use Video Visits

We characterized the differences between 54.3% (327/602) patients *willing* (to use video visits) versus 45.7% (275/602) patients *unwilling* (to use video visits), as shown in Table 2. Those who were willing were significantly younger (average age 68.9, SD 8.8 years) than those unwilling (average age 72.5, SD 9.1 years; *P*<.001). There appears to be a sharp drop in willingness after the age of 75 years.

They were also more likely to have at least an associate educational degree (120/327, 36.7% vs 68/275, 24.7%; *P*=.002) and be more health literate (180/327, 55% vs 110/275, 40%; *P*<.001). Those who were willing were more likely to not use assistive devices for walking (137/327, 41.9% vs 80/275, 29.1%; *P*=.002) and less dependent in their IADL (mean 1.8, SD 2.0 vs mean 2.2, SD 2.2; *P*=.02). Willing veterans reported worse self-rated general health compared with those of unwilling veterans (mean 2.8, SD 0.9 vs mean 3.0, SD 1.0; *P*=.01) and worse physical status (mean 5.2, SD 2.0 vs mean 5.7, SD 2.2; *P*=.004). Willing veterans were also less likely to live in disadvantaged areas (*P*=.048).

When asked about their ability to use technology, a significantly higher number of those willing were able to perform an internet search if given access to a computer (242/327, 74% vs 109/275, 39.6%; *P*<.001); were using email (226/327, 69.1% vs 88/275, 32%; *P*<.001); and were enrolled in the VA's patient portal, My Health*e*Vet (199/327, 60.9% vs 76/275, 27.6%; *P*<.001). The willing and the unwilling to use video visits groups differed regarding the preferred modes of contact (*P*=.003). Compared with those unwilling to use video visits, willing veterans were more likely to prefer contact by the VA via cell phone (189/327, 57.8% vs 129/275, 46.9%) or via My Health*e*Vet secure message (24/327, 7.3% vs 10/275, 3.6%) and less likely to prefer contact by *landline home phone* (67/327, 20.5% vs 73/275, 26.5%) or mail (44/327, 13.5% vs 57/275, 20.7%).

Table 2. Patient characteristics of those willing to use video visits versus not willing to use video visits with their Veterans Affairs care team (N=602).

Characteristics	All completed surveys for study	Willing to use video visits (n=327)	Not willing to use video visits (n=275)	*P* value
Demographics				
Age (years)[a]				
Values, mean (SD; range)	70.6 (9.2; 39-100)	68.9 (8.8; 39-95)	72.5 (9.1; 42-100)	<.001
Age group, n (%)[a]				<.001
<65	122 (20.3)	83 (25.4)	39 (14.2)	
65-69	153 (25.4)	83 (25.4)	70 (25.5)	
70-75	155 (25.7)	88 (26.9)	67 (24.4)	
75-79	83 (13.8)	44 (13.5)	39 (14.2)	
≥80	89 (14.8)	29 (8.9)	60 (21.8)	
White, n (%)	367 (61)	207 (63.3)	160 (58.2)	.23
African American, n (%)	217 (36)	110 (33.6)	107 (38.9)	.21
Hispanic, n (%)	104 (17.3)	56 (17.1)	48 (17.5)	.99
Education (at least associate degree)[a], n (%)	188 (31.2)	120 (36.7)	68 (24.7)	.002
Confident filling out medical forms[a], n (%)	290 (48.2)	180 (55)	110 (40)	<.001
Physical health				
JEN Frailty Index[a,b], mean (SD)	7.1 (1.2)	7.0 (1.2)	7.2 (1.2)	.04
Care Assessment Needs score[b], mean (SD)	93.5 (7.3)	93.1 (7.5)	93.9 (7.7)	.20
Total number of Hierarchical Condition Categories[b], mean (SD)	5.5 (2.3)	5.4 (2.3)	5.6 (2.3)	.29
FRAIL[b,c] scale screen positive (score ≥3), n (%)	253 (42)	132 (40.4)	121 (44)	.41
Self-rated physical status score[a,d], mean (SD)	5.4 (2.1)	5.2 (2.0)	5.7 (2.2)	.004
Issue with walking, stepping, and balance, n (%)	444 (73.8)	237 (72.5)	207 (75.3)	.49
No prosthetic use[a], n (%)	217 (36)	137 (41.9)	80 (29.1)	.002
General health score[a,d], mean (SD)	2.9 (0.9)	2.8 (0.9)	3.0 (1.0)	.01
ADL[e] score[d], mean (SD)	84.3 (20.1)	84.9 (19.4)	83.1 (21.0)	.28
ADL deficits[b], mean (SD)	2.3 (2.8)	2.2 (2.6)	2.4 (2.9)	.38
IADL[f] score[a,d], mean (SD)	6.0 (2.1)	6.2 (2.0)	5.8 (2.2)	.02
IADL deficits[a,b], mean (SD)	2.0 (2.1)	1.8 (2.0)	2.2 (2.2)	.02
Homebound or semihomebound, n (%)	169 (28.1)	91 (27.8)	78 (28.4)	.96
Mental health				
PHQ-2[g] depression screen positive (score ≥3), n (%)	196 (32.6)	117 (35.8)	79 (28.77)	.08
Self-perception of aging score [b], mean (SD)	3.2 (1.5)	3.3 (1.5)	3.1 (1.5)	.10
Social characteristics				

Characteristics	All completed surveys for study	Willing to use video visits (n=327)	Not willing to use video visits (n=275)	*P* value
Area Deprivation Index Score[a,b], n (%)				.048
1-25	113 (18.8)	66 (20.2)	47 (17.1)	
26-50	155 (25.7)	95 (29.1)	60 (21.8)	
51-75	192 (31.9)	99 (30.3)	93 (33.8)	
76-100	138 (22.9)	64 (19.6)	74 (26.9)	
Have a caregiver, n (%)	204 (33.9)	106 (31)	98 (35.6)	.46
Social Networking Index[d], mean (SD)	1.5 (1.1)	1.6 (1.1)	1.5 (1.1)	.27
Having no trouble in transportation, n (%)	373 (62)	198(60.6)	175(63.3)	.49
Travel time to physician >60 minutes, n (%)	177 (29.4)	107 (32.7)	70 (25.5)	.06
Have delayed physicians' appointments owing to transportation troubles, n (%)[a]	136 (22.6)	82 (25.1)	54 (19.6)	.14
Technology ability[a]—facilitator, n (%)				
Use email[a]	314 (52.2)	226 (69.1)	88 (32)	<.001
Able to do an internet search[a]	351 (58.3)	242 (74)	109 (39.6)	<.001
Use email and internet search	296 (49.2)	214 (65.4)	82 (29.8)	<.001
Enrolled in My Health*e*Vet (MHV)[a]	275 (45.7)	199 (60.9)	76 (27.6)	<.001
Preferred mode of contact[a]				.003
By home phone	140 (23.3)	67 (20.5)	73 (26.5)	
By cell phone	318 (52.8)	189 (57.8)	129 (46.3)	
By MHV secure message	34 (5.6)	24 (7.3)	10 (3.6)	
By email	101 (16.8)	44 (13.5)	57 (20.7)	

[a]*P*<.05 defined statistical significance.

[b]Lower score is better.

[c]FRAIL: Fatigue, Resistance, Ambulation, Illnesses, and Loss of Weight.

[d]Higher score is better.

[e]ADL: activities of daily living.

[f]IADL: instrumental activities of daily living.

[g]PHQ-2: Patient Health Questionnaire-2.

Differences Between Willing Respondents With and Those Without Access to Video-Capable Technology

Upon being asked about their access to technology, of the 327 veterans who were willing to use video visits, 248 (75.8%) had a smartphone or computer with a camera, whereas 69 (21.1%) did not. The characteristics of these subgroups are presented in Table 3. Patients with access to the necessary devices were younger (mean 68.3, SD 8.9 vs mean 71.8, SD 8.6; *P*=.004), more health literate (144/248, 58.1% vs 28/69, 41%; *P*=.01), and less likely to be African American (73/248, 29.4% vs 33/69, 48%; *P*=.007) than those without technology access. Veterans with video-capable technology were more functionally independent in their ADL (Barthel ADL score: mean 86.4, SD 17.8 vs mean 77.3, SD 24.3, *P*=.005; number of ADL deficits: mean 2.0, SD 2.5 vs mean 3.2, SD 3.2, *P*=.005) and IADL (Lawton IADL score 6.3, SD 1.9 vs 5.7, SD 2.2, *P*=.04; and number of IADL deficits 1.7, SD 1.9 vs 2.3, SD 2.2 and *P*=.04). They were less likely to report issues with walking, stepping, or balance (173/248, 69.8% vs 58/69, 84%; *P*=.03) and more likely to not use assistive devices for walking (115/248, 46.4% vs 18/69, 26%; *P*=.004). They were less likely to live in disadvantaged areas (*P*=.049). They were also less likely to have trouble with transportation (167/248, 67.3% vs 25/69, 36%; *P*<.001) and less likely to have delayed their physicians' appointments because of transportation troubles (54/248, 21.8% vs 26/69, 38%; *P*=.01). Veterans with access to a video-capable device were more likely to be able to use the internet (204/248, 82.3% vs 28/69, 41%; *P*<.001), use email (196/248, 79% vs 20/69, 29%; *P*<.001), and be enrolled in My Health*e*Vet (173/248, 69.8% vs 17/69, 25%; *P*<.001).

Table 3. Patient characteristics by access to a video-capable technology of those willing to use video visits who answered both questions (N=317).

Characteristics	Access to a video-capable device (n=248)	No access to a video-capable device (n=69)	*P* value
Demographics			
Age (years)			
Values, mean (SD; range)	68.3 (8.9; 39-95)	71.8 (8.6; 55-94)	.004
Age group, n (%)[a]			.02
≤64	68 (27.4)	12 (17.4)	
65-69	64 (25.8)	14 (20)	
70-74	66 (26.6)	21 (30)	
75-79	31 (12.5)	12 (17)	
≥80	19 (7.7)	10 (14)	
White, n (%)[a]	168 (68)	34 (49)	.007
African American, n (%)[a]	73 (29.4)	33 (48)	.007
Hispanic, n (%)	46 (18.5)	9 (13)	.37
Education (at least associate degree), n (%)	91 (36.7)	22 (32)	.55
Confident filling out medical forms, n (%)[a]	144 (58.1)	28 (41)	.01
Physical health			
JEN Frailty Index[b], mean (SD)	7.0 (1.1)	7.1 (1.2)	.53
Care Assessment Needs score[b], mean (SD)	92.9 (7.0)	93.4 (6.7)	.59
Total number of Hierarchical Condition Categories[b], mean (SD)	5.4 (2.5)	5.5 (1.9)	.72
FRAIL[c] scale screen positive (score ≥3), n (%)	100 (40.3)	28 (41)	.99
Physical status score[d], mean (SD)	5.3 (2.0)	5.0 (1.8)	.23
Issue with walking, stepping, balance, n (%)[a]	173 (69.8)	58 (84)	.03
No prosthetic use, n (%)[a]	115 (46.4)	18 (26)	.004
General health score[a,d], mean (SD)	2.8 (0.9)	2.6 (0.8)	.08
ADL[e] score[a,d], mean (SD)	86.4 (17.8)	77.3 (24.3)	.005
ADL deficits[a,b], mean (SD)	2.0 (2.5)	3.2 (3.2)	.005
IADL[f] score[a,d], mean (SD)	6.3 (1.9)	5.7 (2.2)	.04
IADL deficits[a,b], mean (SD)	1.7 (1.9)	2.3 (2.2)	.04
Homebound or semihomebound, n (%)	70 (28.2)	17 (25)	.66
Mental health			
PHQ-2[g] screen positive (score ≥3), n (%)	86 (34.7)	26 (38)	.75
Self-perception of aging score[b], mean (SD)	3.2 (1.6)	3.5 (1.3)	.11
Social characteristics			
Area Deprivation Index Score[a,b], n (%)			.49
1-25	54 (21.8)	11 (15.9)	
26-50	76 (30.6)	19 (28)	
51-75	69 (27.8)	25 (36)	
76-100	46 (18.5)	14 (20)	

Characteristics	Access to a video-capable device (n=248)	No access to a video-capable device (n=69)	*P* value
Have a caregiver, n (%)	81 (32.7)	25 (36)	.68
Social Networking Index[d], mean (SD)	1.6 (1.1)	1.5 (1.1)	.51
Have no trouble with transportation, n (%)[a]	167 (67.3)	25 (36)	<.001
Travel time to physician >60 minutes, n (%)	83 (33.5)	23 (33)	.99
Have delayed physicians' appointments owing to transportation troubles, n (%)[a]	54 (21.8)	26 (38)	.01
Technology ability, n (%)[a]			
Use of email[a]	196 (79.0)	20 (29)	<.001
Able to do an internet search[a]	204 (82.3)	28 (41)	<.001
Use email and internet search	187 (75.4)	17 (25)	<.001
Enrolled in My Health*e*Vet (MHV)[a]	173 (69.8)	17 (25)	<.001
Preferred mode of contact[a]			.03
By home phone	46 (18.6)	20 (29)	
By cell phone	144 (58.1)	39 (57)	
By MHV secure message	22 (8.9)	0 (0)	
By email	34 (13.7)	9 (13)	

[a]*P*<.05 defined statistical significance.

[b]Lower score is better.

[c]FRAIL: Fatigue, Resistance, Ambulation, Illnesses, and Loss of Weight.

[d]Higher score is better.

[e]ADL: activities of daily living.

[f]IADL: instrumental activities of daily living.

[g]PHQ-2: Patient Health Questionnaire-2.

Number of Willing Respondents With Access and Ability to Use Video Visits

In our HNHR group, 54.3% (327/602) were willing to receive care from their VA health care team via video visits (Table 2), and of those, 78.2% (248/317) had access to video-capable technology (Table 3). Therefore, 41.2% (248/602) participants were willing and had the technology for a video visit. Among the willing 248 patients with access to a video-capable device, only 204 (82.3%) were likely to be comfortable using technology when factoring in previous use of the internet or email (Table 3). Therefore, the percentage of HNHR veterans with access and ability likely decreases to approximately 33.9% (204/602).

Multivariable Logistic Regression

Multivariable logistic regression was conducted to give a sense of the relative importance of different predictors of willingness. The odds ratios for willingness estimated for age, degree, confidence in filling out forms, JFI score, self-perception of health, prosthetics use, general health, IADL score, ADI, use of email, use of the internet, and My Health*e*Vet use are presented in Table 4. As shown in Table 4, age, confidence in filling out forms, prosthetics use, general health, and use of the internet were significantly associated with willingness of video visit use in the multivariable analysis, indicating that they are the strongest predictors compared with others that were only significant in the univariate analysis.

Table 4. Odds ratio for predictors of willingness to use video visits in multivariable logistic regression.

Characteristics	Odds ratio (95% CI)	*P* value
Age	0.97 (0.95-0.995)	.02
Education (at least associate degree)	1.39 (0.94-2.07)	.10
Confidence in filling medical forms	1.47 (1.01-2.14)	.046
JEN Frailty Index score	1.03 (0.89-1.20)	.70
Self-perception of aging	0.92 (0.82-1.03)	.15
Prosthetics use	1.85 (1.23-2.80)	.003
General health	0.72 (0.56-0.92)	.01
Instrumental activities of daily living score	1.03 (0.93-1.14)	.60
Area Deprivation Index score	0.99 (0.99-1.00)	.06
Use of email	0.98 (0.67-1.43)	.91
Use of the internet	2.34 (1.65-3.34)	<.001
My Health*e*Vet use	1.29 (0.90-1.85)	.17

Discussion

Principal Findings

Our study aimed to identify the readiness of using video visits for health care by assessing willingness, access, and ability in older HNHR patients with complex needs, functional limitations, and a variety of chronic conditions [30]. A little over half were willing to use video visits, three quarters of those had access, and only 80% of them were comfortable with technology. Overall, we believe that only one-third of the HNHR veterans had the willingness, access, and ability to use video visits for health care. Therefore, data from our project suggest that among vulnerable HNHR older adults, the proportion not ready for video visits may be much higher than the one-third previously reported for a cross-section of community-dwelling older adults [9] and likely is approximately two-thirds of the HNHR veterans.

The access gap between those willing yet without technology was larger among those who were older, less health literate, or African American or lived in disadvantaged areas. Veterans who did not have a device were less healthy, more likely to be dependent and have transportation challenges, and less well-versed with using the internet and email. In contrast, veterans who were willing to use video visits were younger, more literate, more adept at using technology, more functionally independent in their IADL, and less likely to live in disadvantaged areas but had worse self-rated health. Age, confidence in filling out forms, prosthetic use, general health, and internet use were significantly associated with willingness to use video visits in the multivariable analysis. Age is a strong predictor, and there appears to be a sharp drop in willingness after the age of 75 years. Moreover, there was a very strong correlation of both technology access and digital skills on willingness.

In addition, willingness was correlated with a previous history of having missed their in person physicians' appointments because of issues involving transportation. Although the HNHR population's willingness to use video visits represents an opportunity to address critical access barriers often seen in this population, these inequities in access to video visits and their lack of prior technology use warrant further attention, as reliance on telemedicine visits could exacerbate the gap in access to care for vulnerable populations. Although there were no differences in the willingness to use video visits and insignificant differences in transportation barriers between veterans who finished the survey in person and those who did not, individuals completing the survey via mail were more confident filling out medical forms but in worse physical health and more socially isolated than those who completed by phone or in person. Although not significant, these results may be an indirect reflection regarding the availability of resources for attending in person appointments and need further inspection.

Owing to the unprecedented challenges to health care during the COVID-19 pandemic, there has been a substantial increase in patients' willingness to use technology to reduce in person appointments to safeguard against COVID-19 [2]. However, even as telemedicine willingness increases, not only is it necessary to address the lack of access to technology in and of itself but also other strategies to address telemedicine unreadiness are needed. Some ways of addressing technology access challenges may be providing necessary equipment and bandwidth via the health care system [31] or helping patients acquire affordable devices and broadband internet [11]. In August 2020, the Assistant Under Secretary for Health for Clinical Services submitted a memorandum for expanding access to telehealth for veterans through a digital divide consult. This consult is available to veterans who do not have a video-capable device or connectivity for eligibility in participating in the Lifeline program to receive a loaned device (eg, iPads or iPhones) for accessing telemedicine in their home or location of choice. The VA offers tablets and data plans to veterans who qualify using a digital divide consult and has simplified the use of technology for video visits by configuring VA-loaned tablets to allow for a single-use mode [31]. The single-use mode replaces the complexity of multiple VA functions, features, and apps on the device with a VVC icon that readily connects the veteran to a telemedicine medical room [31].

Strategies are needed to address technology literacy and offer necessary education and support so that patients may engage successfully in video visits. Specific outreach efforts need to target communities that have been found to be less ready for video visits, including African Americans and those with high area deprivation scores. More systems need to implement initiatives that enable trained staff or even volunteers to help patients navigate the complexities of devices and applications [4] and programs that enhance self-efficacy, which have proven successful in the adoption of technology [32]. Other potential approaches include offering technology education and support, using nonmedical staff to conduct a mock visit before the actual visit to train older adults in navigating the technology, using trained peers or community health workers to provide in-home training or act as telepresenters for in-home video visits with high-risk older adults, and encouraging family caregivers and friends to participate during telemedicine encounters.

Moreover, the presentation of video versus in person visits is somewhat of a false dichotomy. Video visits may have more capacity to address multi-morbid diseases, as indicated by longer visit durations and a larger number of visit diagnoses than those of telephone visits [33]. However, there is a population that has significant barriers to both physical (transportation) and video (digital literacy) interactions. For this group, telephone visits may be more accessible than video or in person visits and can potentially be another means of increasing care. For a few patients, neither telemedicine (telephone or video) nor in person may be feasible, and home care models such as Medicare's Independence At Home and VA's HBPC may be necessary.

This study has several strengths. A strength of our study is that it specifically assesses an older, functionally dependent, HNHR population with complex needs and social isolation. We used a novel VA set of HNHR older adults and surveyed them about their attitudes toward telemedicine and their physical, emotional, and social determinants. In addition to characterizing the willingness, access, and ability to use video visits for health care, in the context of their physical, emotional, and social characteristics, as has previously been done [9,11-13], we correlated it to frailty status and the neighborhood they reside in.

However, this study does have several limitations. One of the limitations is that technology access was only asked for those who were willing. Had we surveyed our total study sample regarding access, the proportion of those lacking access would likely be higher, given the lower use of email and internet and lower education level and health literacy among the unwilling veterans. Moreover, we did not explore the reasons driving the unwillingness to use video visits, explore the subgroup that has the technology but is unwilling, or include an *uncertain* response category for willingness in our survey. Understanding their barriers and facilitators might provide important insights beyond affordable access to devices and connectivity and digital skills [34]. Previous reports suggest that in addition to poor technology access and literacy, technology unwillingness may be driven by several other factors, including sensory or memory impairment [9,12], which we did not assess. The ADI does not explicitly incorporate neighborhood availability of affordable broadband, which may be a big factor in whether or not people use it; however, it may reflect digital redlining [34].

We also did not ask about or compare willingness among those who had versus did not have prior telehealth visits. Some of the constructs are somewhat narrowly assessed: specifically, the social support measure that assesses caregiver presence with an unvalidated question. However, this was supplemented by the Berkman–Syme Social Network Index, which takes into account marital status, frequency of meeting and talking to close friends and relatives, and participating in religious and club meetings. Similarly, mental health is assessed with a validated 2-item depression scale and is therefore supplemented by the 5-item Self-Perception of Aging scale. Another limitation is that our population was US veterans and overwhelmingly male. The gender demographics here reflect that of the VA, where 89.6% of all veterans are male [35], and not of the general older adult population. Older female HNHR patients may have different needs and access challenges than those described in this study. Moreover, our study was urban and limited geographically to the Miami area and, thus, may not represent regional variations. In addition, we did not assess the availability of the caregivers who may be willing and able to help with the video visit and may have access to the needed devices. Adjustment for multiple comparisons tends to increase type II error [36,37]; therefore, we did not adjust for multiple comparisons. Other limitations include a relatively low survey response rate. The survey was also conducted for patients in an integrated health care system, which may make the findings less generalizable to patients from other types of systems.

Conclusions

Our results underscore the well-recognized fact that older adults, a group that uses health care at one of the highest rates, face significant barriers to accessing needed care, whether it be in person or telemedicine. Certain characteristics put individuals within this group at an even higher risk for barriers to care. Future research is needed to urgently explore ways of mitigating the identified obstacles to telemedicine among HNHR patients at a system level and study and address potential barriers such as concerns about care quality and relationships with physicians at the patient–provider level [38,39]. Programs for HNHR patients should address the specific factors identified here to pave the way for equitable access to health care among high-risk patients. It is recognized that individuals' characteristics, as well as the surrounding social and health care system, are the most important factors that affect telemedicine adoption [40], and some may also serve as barriers. Thus, it was difficult to make a distinction. However, it is important to recognize that only some of the factors are modifiable; thus, the need to make a distinction may be less pertinent. These respondents completed the survey before the COVID-19 pandemic, and it is possible that the COVID-19 pandemic may have significantly changed patients' video acceptance and technology availability as they may have adopted video for personal and health reasons [2]. Thus, the development of innovative, sustainable strategies to support and improve care access for this vulnerable population will help during the COVID-19 pandemic; however, it will also help better manage HNHR patients and keep them healthy in

their homes for as long as possible after the COVID-19 pandemic.

Acknowledgments

This work was supported in part by funding from the Health Services Research & Development Service of the US Department of Veterans Affairs (VA; SDR 18-313 under Grant HX-18-015), the VA Office of Geriatrics and Extended Care (GEC; VA GEC SP3E-GREA160) and by the Veterans Integrated Service Network 8 GEC Office. The authors also gratefully acknowledge the support of the sponsors and of the Miami VA Geriatric Research, Education, and Clinical Center; Bruce W Carter, Department of VA Medical Center; and the Miami VA Research. The authors also wish to thank the project staff for their help with various aspects of the project. This quality improvement project registration number is ClinicalTrials.gov NCT04846049.

The data presented here were presented partly in a poster at the Gerontological Society of America Annual conference in 2019, in Austin, Texas, titled, *Age- and Ethnicity-Related Disparities in Technology Use Among High-Risk Veterans*.

Authors' Contributions

All authors contributed to the concept, preparation, and revision of the manuscript and approved its final version. The investigators retained full independence in the conduct of this research. The views expressed in this paper are those of the authors and do not reflect the position or policy of the Department of Veterans Affairs or the US government. The authors assume full responsibility for the ideas presented.

Conflicts of Interest

None declared.

References

1. Using telehealth to expand access to essential health services during the COVID-19 pandemic. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/telehealth.html [accessed 2018-05-31]
2. Heyworth L, Kirsh S, Zulman D, Ferguson J, Kizer K. Expanding access through virtual care: the VA's early experience with COVID-19. NEJM Catalyst Innovations in Care Delivery. 2020. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.20.0327 [accessed 2022-02-17]
3. Vazquez J, Islam T, Gursky J, Beller J, Correa DJ. Access to care matters: remote health care needs during COVID-19. Telemed J E Health 2021 Apr 01;27(4):468-471. [doi: 10.1089/tmj.2020.0371] [Medline: 33017270]
4. Triana AJ, Gusdorf RE, Shah KP, Horst SN. Technology literacy as a barrier to telehealth during COVID-19. Telemed J E Health 2020 Sep 01;26(9):1118-1119. [doi: 10.1089/tmj.2020.0155] [Medline: 32429770]
5. Coronavirus waivers and flexibilities. Centers for Medicare & Medicaid Services. 2020. URL: https://www.cms.gov/about-cms/emergency-preparedness-response-operations/current-emergencies/coronavirus-waivers [accessed 2022-02-17]
6. VA expands telehealth by allowing health care providers to treat patients across state lines. US Department of Veterans Affairs. 2018. URL: https://tinyurl.com/yxvsbpho [accessed 2018-05-11]
7. Connolly SL, Stolzmann KL, Heyworth L, Weaver KR, Bauer MS, Miller CJ. Rapid increase in telemental health within the department of veterans affairs during the COVID-19 pandemic. Telemed J E Health 2021 Apr 01;27(4):454-458. [doi: 10.1089/tmj.2020.0233] [Medline: 32926664]
8. COVID-19: clinical trainees leveraging video telehealth from home. VHA Office of Connected Care. 2020. URL: https://www.gerischolars.org/mod/page/view.php?id=1508 [accessed 2022-02-18]
9. Lam K, Lu AD, Shi Y, Covinsky KE. Assessing telemedicine unreadiness among older adults in the United States during the COVID-19 pandemic. JAMA Intern Med 2020 Oct 01;180(10):1389-1391 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2671] [Medline: 32744593]
10. Dorsey ER, Topol EJ. State of telehealth. N Engl J Med 2016 Jul 14;375(2):154-161. [doi: 10.1056/nejmra1601705]
11. Nouri S, Khoong E, Lyles C, Karliner L. Addressing equity in telemedicine for chronic disease management during the COVID-19 pandemic. NEJM Catalyst Innovations in Care Delivery. 2020. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.20.0123 [accessed 2022-02-18]
12. Roberts ET, Mehrotra A. Assessment of disparities in digital access among medicare beneficiaries and implications for telemedicine. JAMA Intern Med 2020 Oct 01;180(10):1386-1389 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2666] [Medline: 32744601]
13. Eberly L, Kallan MJ, Julien HM, Haynes N, Khatana SA, Nathan AS, et al. Patient characteristics associated with telemedicine access for primary and specialty ambulatory care during the COVID-19 pandemic. JAMA Netw Open 2020 Dec 01;3(12):e2031640 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.31640] [Medline: 33372974]
14. Ferguson J, Jacobs J, Yefimova M, Greene L, Heyworth L, Zulman DM. Virtual care expansion in the Veterans Health Administration during the COVID-19 pandemic: clinical services and patient characteristics associated with utilization. J Am Med Inform Assoc 2021 Mar 01;28(3):453-462 [FREE Full text] [doi: 10.1093/jamia/ocaa284] [Medline: 33125032]

15. Edes T, Dally S, Davis D, Intrator O, Phibbs C, Kinosian K. Focusing Independence at Home Qualifying Criteria (IAH-Q) to identify High Need High Risk (HNHR) Veterans who benefit from Home Based Primary Care (HBPC). In: J Am Geriatr Society. 2019 Presented at: American Geriatrics Society 2019 Annual Scientific Meeting; May 2 - 4, 2019; Portland, Oregon p. 170-171 URL: https://meeting.americangeriatrics.org/sites/default/files/inline-files/2019%20JAGS%20abstract%20supplement.pdf
16. Kinosian B, Wieland D, Gu X, Stallard E, Phibbs CS, Intrator O. Validation of the JEN frailty index in the National Long-Term Care Survey community population: identifying functionally impaired older adults from claims data. BMC Health Serv Res 2018 Nov 29;18(1):908 [FREE Full text] [doi: 10.1186/s12913-018-3689-2] [Medline: 30497450]
17. Morley JE, Malmstrom TK, Miller DK. A simple frailty questionnaire (FRAIL) predicts outcomes in middle aged African Americans. J Nutr Health Aging 2012 Jul 11;16(7):601-608 [FREE Full text] [doi: 10.1007/s12603-012-0084-2] [Medline: 22836700]
18. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. Maryland State Med J 1965;14:61-65. [doi: 10.1037/t02366-000]
19. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist 1969 Sep 01;9(3 Part 1):179-186. [doi: 10.1093/geront/9.3_part_1.179]
20. Ornstein KA, Leff B, Covinsky KE, Ritchie CS, Federman AD, Roberts L, et al. Epidemiology of the homebound population in the United States. JAMA Intern Med 2015 Jul 01;175(7):1180-1186 [FREE Full text] [doi: 10.1001/jamainternmed.2015.1849] [Medline: 26010119]
21. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. Med Care 2003;41(11):1284-1292. [doi: 10.1097/01.MLR.0000093487.78664.3c]
22. Lawton MP. The Philadelphia Geriatric Center Morale Scale: a revision. J Gerontol 1975 Jan 01;30(1):85-89. [doi: 10.1093/geronj/30.1.85] [Medline: 1109399]
23. Berkman L, Syme SL. Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents. Am J Epidemiol 1979 Feb;109(2):186-204. [doi: 10.1093/oxfordjournals.aje.a112674] [Medline: 425958]
24. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. Fam Med 2004 Sep;36(8):588-594 [FREE Full text] [Medline: 15343421]
25. Idler EL, Angel RJ. Self-rated health and mortality in the NHANES-I Epidemiologic Follow-up Study. Am J Public Health 1990 Apr;80(4):446-452. [doi: 10.2105/ajph.80.4.446] [Medline: 2316767]
26. Locatelli SM, Sharp LK, Syed ST, Bhansari S, Gerber BS. Measuring health-related transportation barriers in urban settings. J Appl Meas 2017;18(2):178-193 [FREE Full text] [Medline: 28961153]
27. Wang L, Porter B, Maynard C, Evans G, Bryson C, Sun H, et al. Predicting risk of hospitalization or death among patients receiving primary care in the Veterans Health Administration. Med Care 2013 Apr;51(4):368-373. [doi: 10.1097/MLR.0b013e31827da95a] [Medline: 23269113]
28. Dang S, Gomez-Orozco C, Sorial A, Ruiz D, Siddharthan K, Rodriguez R, et al. Integration of MyHealtheVet (MHV) into the management of Heart Failure (HF) patients in the Patient Aligned Care Teams (PACTs). In: Poster Presentation at the 2015 HSR&D/QUERI National Conference Abstract. 2015 Presented at: HSR&D/QUERI National Conference; July 8-10, 2015; Philadelphia, PA p. 8-10 URL: https://www.hsrd.research.va.gov/meetings/2015/abstract-display.cfm?RecordID=696
29. Kind AJ, Buckingham WR. Making neighborhood-disadvantage metrics accessible — the neighborhood atlas. N Engl J Med 2018 Jun 28;378(26):2456-2458. [doi: 10.1056/nejmp1802313]
30. Pritchard D, Petrilla A, Hallinan S, Taylor DH, Schabert VF, Dubois RW. What contributes most to high health care costs? Health care spending in high resource patients. J Manag Care Spec Pharm 2016 Feb;22(2):102-109. [doi: 10.18553/jmcp.2016.22.2.102] [Medline: 27015249]
31. VA issued devices Standard Operating Procedure (SOP). Department of Veterans Affairs Office of Connected Care. 2020. URL: https://www.va.gov/covidtraining/docs/04-07-2020-SOP-Acute-Medical-Management.pdf [accessed 2022-02-18]
32. Gatti FM, Brivio E, Galimberti C. “The future is ours too”: a training process to enable the learning perception and increase self-efficacy in the use of tablets in the elderly. Edu Gerontol 2017 Jan 09;43(4):209-224. [doi: 10.1080/03601277.2017.1279952]
33. Schifeling CH, Shanbhag P, Johnson A, Atwater RC, Koljack C, Parnes BL, et al. Disparities in video and telephone visits among older adults during the COVID-19 pandemic: cross-sectional analysis. JMIR Aging 2020 Nov 10;3(2):e23176 [FREE Full text] [doi: 10.2196/23176] [Medline: 33048821]
34. Digital inclusion. National Digital Inclusion Alliance. 2021. URL: https://www.digitalinclusion.org/definitions/ [accessed 2022-02-18]
35. Department of Veterans Affairs statistics at a glance. National Center for Veterans Analysis and Statistics. 2021. URL: https://www.va.gov/vetdata/docs/Quickfacts/Stats_at_a_glance_9_30_21.PDF [accessed 2022-02-18]
36. Thomas D, Siemiatycki J, Dewar R, Robins J, Goldberg M, Armstrong BG. The problem of multiple inference in studies designed to generate hypotheses. Am J Epidemiol 1985 Dec;122(6):1080-1095. [doi: 10.1093/oxfordjournals.aje.a114189] [Medline: 4061442]
37. Perneger TV. What's wrong with Bonferroni adjustments. Br Med J 1998 Apr 18;316(7139):1236-1238 [FREE Full text] [doi: 10.1136/bmj.316.7139.1236] [Medline: 9553006]

38. Kaplan B. Revisiting health information technology ethical, legal, and social issues and evaluation: telehealth/telemedicine and COVID-19. Int J Med Inform 2020 Nov;143:104239 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104239] [Medline: 33152653]
39. McConnochie KM. Webside manner: a key to high-quality primary care telemedicine for all. Telemed J E Health 2019 Nov 01;25(11):1007-1011. [doi: 10.1089/tmj.2018.0274] [Medline: 30648924]
40. Harst L, Timpel P, Otto L, Richter P, Wollschlaeger B, Winkler K, et al. Identifying barriers in telemedicine-supported integrated care research: scoping reviews and qualitative content analysis. J Public Health (Berl.) 2019 Apr 09;28(5):583-594. [doi: 10.1007/s10389-019-01065-5]

Abbreviations

ADI: Area Deprivation Index
ADL: activities of daily living
CMS: Centers for Medicare and Medicaid Services
HBPC: home-based primary care
HNHR: high-need, high-risk
IADL: instrumental activities of daily living
JFI: JEN Frailty Index
VA: Veterans Affairs
VVC: Veteran Video Connect

Edited by A Mavragani; submitted 03.08.21; peer-reviewed by K Possemato, P Yellowlees, T Jamieson, A Sheon; comments to author 01.10.21; revised version received 20.12.21; accepted 25.01.22; published 08.04.22.

Please cite as:
Dang S, Muralidhar K, Li S, Tang F, Mintzer M, Ruiz J, Valencia WM
Gap in Willingness and Access to Video Visit Use Among Older High-risk Veterans: Cross-sectional Study
J Med Internet Res 2022;24(4):e32570
URL: https://www.jmir.org/2022/4/e32570
doi:10.2196/32570
PMID:35394440

Original Paper

Designing a Framework for Remote Cancer Care Through Community Co-design: Participatory Development Study

Eliah Aronoff-Spencer[1,2*], MD, PhD; Melanie McComsey[1*], PhD; Ming-Yuan Chih[3*], PhD; Alexandra Hubenko[4], MBA; Corey Baker[5], PhD; John Kim[3], Ed, MPH; David K Ahern[6], PhD; Michael Christopher Gibbons[7], MPH, MD; Joseph A Cafazzo[8], PENG; Pia Nyakairu[8], BID; Robin C Vanderpool[9], DrPH; Timothy W Mullett[3], MBA, MD; Bradford W Hesse[10], PhD

[1]Design Lab, University of California San Diego, La Jolla, CA, United States
[2]Division of Infectious Diseases and Global Public Health, Department of Medicine, UC San Diego School of Medicine, La Jolla, CA, United States
[3]Department of Health & Clinical Sciences, College of Health Sciences, University of Kentuck, Lexington, CA, United States
[4]Qualcomm Institute, University of California San Diego, La Jolla, CA, United States
[5]Department of Computer Science, College of Engineering, University of Kentucky, Lexington, KY, United States
[6]Department of Psychiatry, Brigham and Women's Hospital, Boston, MA, United States
[7]Greystone Group, Inc, Washington, DC, United States
[8]University Health Network, Toronto, ON, Canada
[9]National Cancer Institute, Bethesda, CA, United States
[10]National Cancer Institute (Retired), Kona, HI, United States
[*]these authors contributed equally

Corresponding Author:
Timothy W Mullett, MBA, MD
Department of Health & Clinical Sciences
College of Health Sciences
University of Kentuck
Lexington, CA
United States
Phone: 1 8592297665
Email: timothy.mullett@uky.edu

Abstract

Background: Recent shifts to telemedicine and remote patient monitoring demonstrate the potential for new technology to transform health systems; yet, methods to design for inclusion and resilience are lacking.

Objective: The aim of this study is to design and implement a participatory framework to produce effective health care solutions through co-design with diverse stakeholders.

Methods: We developed a design framework to cocreate solutions to locally prioritized health and communication problems focused on cancer care. The framework is premised on the framing and discovery of problems through community engagement and lead-user innovation with the hypothesis that diversity and inclusion in the co-design process generate more innovative and resilient solutions. Discovery, design, and development were implemented through structured phases with *design studios* at various locations in urban and rural Kentucky, including Appalachia, each building from prior work. In the final design studio, working prototypes were developed and tested. Outputs were assessed using the System Usability Scale as well as semistructured user feedback.

Results: We co-designed, developed, and tested a mobile app (myPath) and service model for distress surveillance and cancer care coordination following the *LAUNCH (Linking and Amplifying User-Centered Networks through Connected Health)* framework. The problem of awareness, navigation, and communication through cancer care was selected by the community after framing areas for opportunity based on significant geographic disparities in cancer and health burden resource and broadband access. The codeveloped digital myPath app showed the highest perceived combined usability (mean 81.9, SD 15.2) compared with the current gold standard of distress management for patients with cancer, the paper-based National Comprehensive Cancer Network Distress Thermometer (mean 74.2, SD 15.8). Testing of the System Usability Scale subscales showed that the myPath app had significantly

better usability than the paper Distress Thermometer (t_{63}=2.611; *P*=.01), whereas learnability did not differ between the instruments (t_{63}=–0.311; *P*=.76). Notable differences by patient and provider scoring and feedback were found.

Conclusions: Participatory problem definition and community-based co-design, *design-with* methods, may produce more acceptable and effective solutions than traditional *design-for* approaches.

(J Med Internet Res 2022;24(4):e29492) doi:10.2196/29492

KEYWORDS

cancer care; distress screening; human-centered design; participatory design; Appalachia; mobile phone

Introduction

Background

The COVID-19 pandemic has upended the US health care delivery system, putting a spotlight on long-standing health and economic inequities, gaps in care, uneven quality, and fragmentation in service models [1]. The rapid implementation and uptake of telemedicine and virtual care in response to the pandemic has been a positive shift [2-4]; however, existing flaws in the system remain, and new challenges to reliable system integration with virtual care are likely, especially in rural regions [5]. The impact of the pandemic on cancer care is particularly worrisome because screening and diagnostic testing have been significantly curtailed or delayed, resulting in projections of a substantial rise in excess deaths from cancer [6]. Of particular concern, patients have been reluctant to return to health care facilities out of concern for exposure to the novel coronavirus.

These compounding challenges have exposed an imperative to redesign systems of cancer care that are “anti-fragile” [7]: resilient, flexible, and democratic. In particular, techniques associated with human-centered design (HCD), such as participatory design; community engagement; and the iterative, collaborative development of sustainable solutions, have been considered essential in re-establishing trust in a health care system challenged during the pandemic [8]. In the tradition of HCD proponents such as von Hippel [9], we believe that innovation to solve problems requires working with the very people, in their own context, whom the problem affects. These are the people who understand the problem best, for whom the solutions must work, and who possess critical knowledge about local resources or potential obstacles to success.

This Study

Drawing on recommendations from a 2016 report issued by the legislatively mandated *President’s Cancer Panel* [10] and in the spirit of the *Beau Biden Cancer Moonshot*, this study describes a human-centered participatory design approach to engage patient, caregiver, and community stakeholders in solving a community-defined problem: here, the problem of serving rural patients experiencing distress during cancer treatment. The effort represents the work of an interagency, public–private partnership called the *LAUNCH (Linking and Amplifying User-Centered Networks through Connected Health)* initiative [11]. As a demonstration project, this study illustrates not only the richness and creativity that can come from co-design methods, but also their potential efficacy: our co-designed intervention outperformed the standard of care in usability testing.

For the demonstration project, Appalachian Kentucky was identified as a region that could benefit most from this kind of participatory design because of both positive and negative attributes: high cancer burdens, connectivity challenges related to rural geography, higher poverty rates, increased social capital, and a historical tradition of community engagement and resource sharing [12]. According to data from the National Program of Cancer Registries, cancer incidence [13] and mortality [14] in Appalachia are among the highest in the nation, and differences between counties and some unincorporated areas can show even starker disparities. Appalachia is also home to some of the most rural and difficult-to-access communities in the United States. Although these communities may benefit most from telemedicine and remote care, broadband access and adoption remain among the lowest in the country [15,16].

During the ethnographic research conducted before this co-design project, we developed relationships with community leaders in Appalachian Kentucky and identified communities where similar initiatives have had success because of high interest and engagement. Our ethnographic work demonstrated that community service and collaborative problem solving are a regular part of the social fabric of many Appalachian communities [12]. Our ethnographic work also demonstrated that more top-down approaches to systemic innovation were likely to fail in the region because of historic exploitation of the people and the land by outsiders [17].

In this context, we adopted a formal design process with roots in HCD, which is used to innovate in complex sociotechnical systems such as health care [18]. This approach may also be referred to as DesignX [19], co-design [20], or participatory design [21], each of which has been shown to improve the efficacy, adoption, and trust of new health services [8]. This approach is consistent with, and an extension of, scenario-based design proposed by Carroll [22] and borrows heavily from rapid contextual design tools that are well reviewed by Holtzblatt et al [23]. Throughout this paper, we treat co-design and cocreation as related but distinct concepts. Co-design is a set of techniques that engage all relevant stakeholders in the formulation, design, and iterative testing of ≥1 solutions to a given problem. Cocreation is the overall process of bringing together stakeholder communities to enable development of co-designed solutions, which are implemented within, and used by, these stakeholder communities.

In what follows, we describe the participatory design, development, and usability testing of a mobile app initially

framed as a system to facilitate communication about cancer-related distress among patients, caregivers, and providers. We situate these processes within a novel framework we refer to as the LAUNCH Roadmap. We then present results from formative evaluation of the app and service model.

Methods

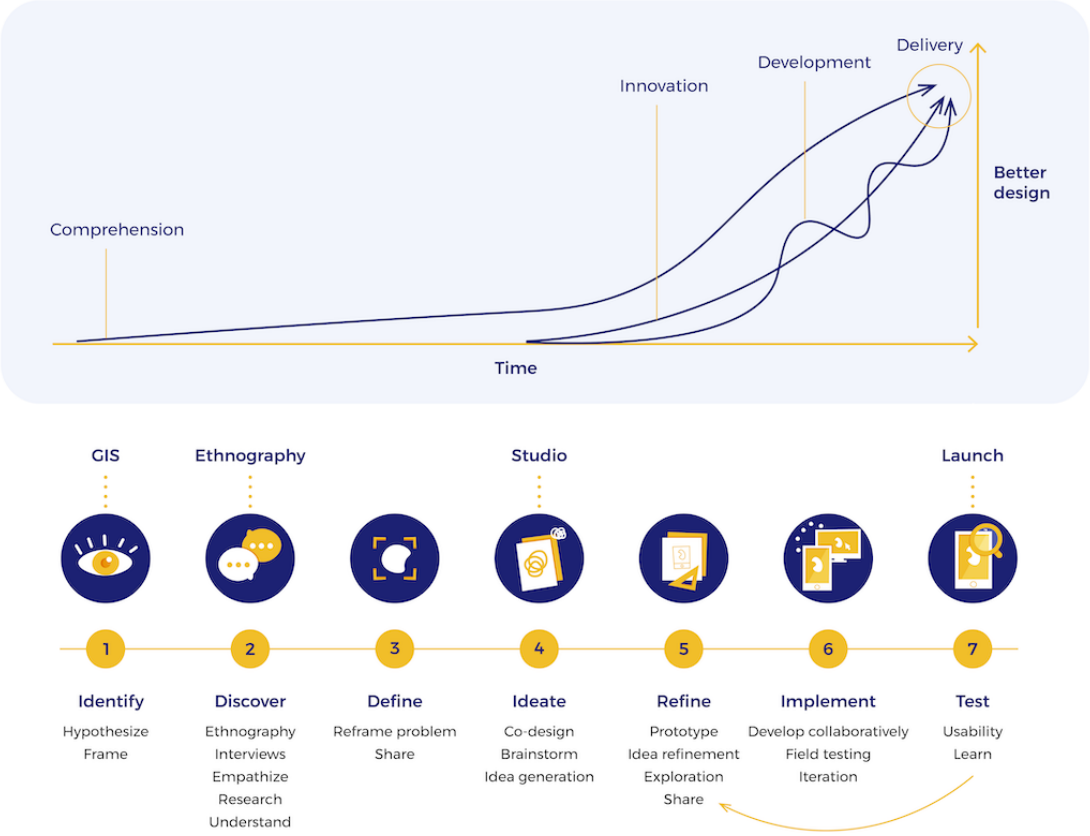
People and Context

We first engaged stakeholders in Kentucky through ethnographic methods described in the study by McComsey et al [12]. These included semistructured interviews and in situ observations with patients with cancer and cancer survivors; family and caregivers; and providers, payers, technologists, and broadband providers. The purpose of the ethnographic research was to understand the experience of having cancer in Appalachia, to take inventory of the health care and connectivity resources supporting cancer care in rural Kentucky, and to develop a network of local champions to continue similar projects of inquiry and co-design.

The Participatory Design Process

Our participatory design process included deliberate opportunities for innovation and iteration of the process itself. This report details the participatory design process as it emerged during collaborative problem solving. The *LAUNCH Roadmap* combines both novel and proven approaches to participatory design. This framework is detailed in prior work by the LAUNCH initiative [11,12] and summarized in this paper in Figure 1.

Figure 1. LAUNCH (Linking and Amplifying User-Centered Networks through Connected Health) Roadmap. Visual journey of the co-design process: From problem identification to delivery: Above: Co-design assumes that problem comprehension, innovation, and development codevelop through a process of participatory iteration. Below: The 7 recursive stages that begin with problem identification, discovery, and definition, followed by ideation, prototyping, and refinement through deployment.



According to the LAUNCH Roadmap, the innovation process begins with (1) problem identification, which is an iterative process requiring (2) contextual discovery and (3) collaborative problem defining. Next, (4) an ideation step generates potential solutions, which are further (5) refined and ultimately (6) implemented and (7) tested (Figure 1). Our process culminated in a systematic comparison of our outcomes with standard of care outcomes, a project that is underway at the time of this writing.

The LAUNCH Roadmap itself was emergent throughout our process, rather than being an existing framework to be applied (Table 1). We see the LAUNCH Roadmap as a blueprint that can be adapted to different contexts and that may emerge in different forms when iterated upon within the course of any community co-design project. As design processes must be tailored to local contexts, frameworks for implementation must be flexible and follow a bottom-up trajectory of emergence.

Table 1. LAUNCH (Linking and Amplifying User-Centered Networks through Connected Health) Roadmap activities (identify, discover, define, ideate, refine, implement, and test).

Purpose	Location (date)	Activity	Outcome
Identify			
• Identify the intersection between connectivity and cancer	• Lexington, Kentucky (November 4, 2017) • La Jolla, California (March 3, 2018)	• LAUNCH kickoff • Quantitative and GIS[a] presentation on the double burden of cancer and connectivity	• Formalized interagency partnership; executed memorandum of understanding between FCC[b] and NCI[c] to share complementary technical and policy expertise through C2H[d]
• Engage local stakeholders	• Kentucky (June-September 2018)	• Local stakeholder meetings	• Recruited coalition of local champions, connectivity stakeholders, and friends of LAUNCH
• Engage national-level stakeholders	• Washington, DC (May 5, 2019)	• Information gathering from industry, government, and academic experts	• Graphical documentation of stakeholder data and perspectives • Meeting summary: L.A.U.N.C.H. Senior Leadership Think Tank: Exploring the Future of Connected Cancer Care in Rural America and Beyond
Discover			
• Understand the experiences of patients with cancer, cancer survivors, caregivers, and health and broadband providers. Document the cancer care and connectivity infrastructure available in eastern Kentucky	• Kentucky (June 2018 to September 2018)	• Ethnographic interviews • Contextual observation • Ongoing coalition building	• Studies and reports: Experiencing Cancer in Appalachian Kentucky and other investigations on cancer care and connectivity
Define			
• Define or redefine the problem to be solved. Align with all stakeholders on problem prioritization	• La Jolla, California (November 13, 2018)	• Tactical meeting dedicated to reviewing year 1 progress • Concretize objectives	• Blueprint for year 2 of LAUNCH
Ideate			
• Generate ideas from diverse stakeholders to improve cancer experiences in rural Kentucky	• Pikeville, Kentucky (August 27, 2018) • Lexington, Kentucky (February 9, 2019) • McKee, Kentucky (June 17, 2019)	• Connected health community ideation studio at SOAR[e] Summit • Ideation studio at Markey Cancer Center • Ideation studio at People's Rural Telephone Cooperative	• Impromptu video pitches by summit attendees for ideas to *help local patients with cancer* • Four co-designed concepts with diverse stakeholders: a portable cancer resource hub, a digital patient navigator, a community sourcing tool, and a wraparound support ecosystem • Seven co-designed innovation recipes for helping rural patients with cancer
Refine			

Purpose	Location (date)	Activity	Outcome
• Select best concepts for prototyping. Co-design specific concepts	• Lexington, Kentucky (October 7-8, 2019)	• Co-design studios with diverse stakeholders	• Artistic representations and video pitches of co-designed categories, questions, scoring system, and communication methods for a new monitoring tool • Artistic representations and video pitches of co-designed electronic interfaces and service models for the new cancer symptom monitoring tool
Implement			
• Develop working prototypes of co-designed concepts • Iterate on prototypes	• Toronto, Canada and La Jolla, California (October 7, 2019) • Berlin, Germany (October 8, 2019)	• Remote, real-time development of prototypes by expert designers and developers	• Prototype of a paper-based distress monitoring tool: *You and Your Well-being* • • Prototype of an electronic distress monitoring tool: *myPath* • Prototype of a provider dashboard and service design for these new tools
Test			
• Test the usability of the prototypes • Collect feedback on the prototypes	• Lexington, Kentucky (October 9-10, 2019)	• User-feedback booths in cancer center lobby and at Markey Cancer Center Affiliate Network Annual Meeting	• Usability surveys for paper-based tool, electronic tool, and provider dashboard • Informal conversations and feedback about prototypes

[a]GIS: geographic information system.

[b]FCC: Federal Communications Commission.

[c]NCI: National Cancer Institute.

[d]C2H: Connect2HealthFCC Task Force.

[e]SOAR: Shaping Our Appalachian Region.

Identify, Discover, Define

The Changing Face of Cancer Care

Cancer treatments are under constant innovation. Some examples include the recent development of minimally invasive surgical techniques, stereotactic radiation therapy, evolving immunotherapies, and the expanding field of precision oncology. However, although innovation in cancer treatment has drawn attention and resources, mechanisms for delivery across the continuum of diagnosis, staging, treatment, survivorship, and end of life have not kept pace [24,25]. Care delivery remains rooted in outpatient clinic visits to assess symptoms and provide results, requiring patients and caregivers to travel and take time off from work and family schedules. The logic of current care delivery is to move patients to information rather than information to patients. This can cause delays because patients must learn to navigate a complex care environment to obtain test results, fill prescriptions, and coordinate referrals to allied health services.

Symptom management, as well, has not kept pace. As advances in cancer treatment over the last decade have produced significant improvements in survival rates and significant reduction of morbidity [26,27], people with cancer now face the challenge of managing their disease as a chronic condition, often with distressing symptoms secondary to the disease process or treatment effects. Much of the burden of this symptom management falls on the person with illness and caregivers outside of the health care system [28].

Furthermore, the burdens of seeking care and managing symptoms are not equally distributed. People in traditionally underserved and rural communities are especially affected by limited access to essential resources. Geographic isolation, poverty, and transportation challenges, coupled with limited or no internet connectivity, can impede access to necessary support services [11,29,30]. Even in cases where broadband and internet connectivity are available, the challenges of rugged geography or long distances can cause connections to be unreliable and unstable. In these cases, novel, innovative methods may be required to ensure connectivity for critical health-related functions [31,32].

It is in this context of the changing landscape of cancer care and emerging technological possibilities that we identified an opportunity at the intersection of health and connectivity. Discovery proceeded with 6 weeks of ethnographic fieldwork conducted in communities and health care facilities across Appalachian Kentucky [12]. We sought to understand the

experiences of local patients with cancer, cancer survivors, caregivers, and health care and broadband providers to inform problem definition and co-design strategies. We further sought to document the cancer care and connectivity infrastructure available in Appalachian Kentucky to build upon and integrate with existing resources as well as to learn from the shortcomings of previous solutions.

Definitions of Distress

In particular, our ethnographic work helped us to critically examine our approach to *distress*, a term common in the cancer symptom management clinical lexicon but which proved uncommon among those who had actually experienced cancer. However, our participants did speak extensively about the physical and psychosocial impacts associated with cancer and pointed to ways they coped with them.

According to the National Comprehensive Cancer Network (NCCN) Guidelines for Distress Management, *distress* is defined as a multifactorial unpleasant experience of a psychological (ie, cognitive, behavioral, and emotional), social, spiritual, and/or physical nature that may interfere with the ability to cope effectively with cancer, its physical symptoms, and its treatment [33]. Identification and treatment of distress have been shown to be critically important to improving health outcomes, quality of life, and adherence to recommended treatments. [34]. For this reason, distress screening at the time of cancer diagnosis is the recommended standard of care, and most cancer centers have shown improvements in routine screening [35].

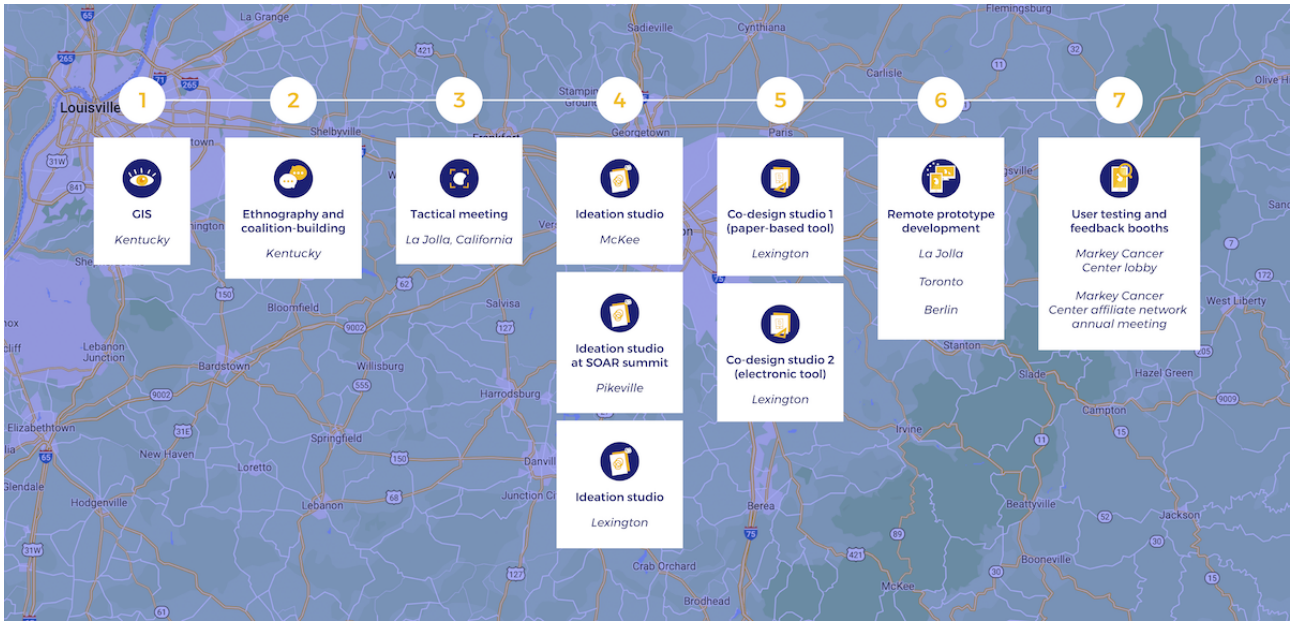
Barriers to collecting these data have typically been related to workflow in the outpatient clinical setting. In addition, communicating results of patient-reported outcomes to providers has been a challenge, decreasing the impact of the information. Finally, collecting this sensitive and timely information from patients in the setting of the waiting room or during triage for a clinic visit may leave patients reluctant to share their actual feelings or symptoms, both past and present. In spite of these obstacles, the Commission on Cancer continues to emphasize the importance of distress screening as an important facet of patient care [36]. Commentators have been advocating for the use of implementation science methods, including those associated with HCD, to improve the effectiveness, reliability, and equity of distress screening efforts [37].

Our ethnographic findings indicated that people in Appalachian Kentucky ascribe different values to personal versus social suffering. Although personal suffering was considered taboo because it could cause others to suffer, suffering on behalf of another person was a source of pride because it indicated the strength of familial and community relationships. We also found that participants preferred to speak about this suffering in colloquial rather than clinical language, both to minimize the taboo of their personal suffering and to capture some nuance lacking in clinical terminology. As the LAUNCH stakeholders began to align on tackling the specific problem of distress monitoring and symptom management, we also began to align on more flexible definitions of key terminology and on a method for integrating the diverse perspectives of patients, caregivers, and clinicians in problem definition.

Ideate, Refine, and Implement

Methods for ideation, co-design, and development of low-fidelity prototypes were adapted from the IDEO Human Centered Design Toolkit [38] and used in LAUNCH Roadmap activities 4-6 (Figures 1 and 2 and Table 1). Ideation (activity 4) occurred over the course of 3 events in Kentucky described below. Refinement of concepts and development of prototypes (activities 5 and 6) were conducted over the course of a 4-day sprint in Lexington. Prototypes were first described verbally, then *developed* with low-fidelity methods such as Post-it Notes (3M) as well as whiteboards and corkboards. In parallel, a professional designer (in Toronto, Ontario, Canada), a mobile app developer (in Berlin, Germany), and a design support group (in La Jolla, California) worked as a *pair design* [39] distributed group to translate the low-fidelity concepts into working prototypes in 72 hours. These working prototypes were then iterated upon, in real time, by participants in Kentucky using web-based collaborative software. Further material can be viewed in the Multimedia Appendices 1-5.

Figure 2. LAUNCH (Linking and Amplifying User-Centered Networks through Connected Health) co-design approach in rural Kentucky. The co-design, cocreation, and testing of the myPath system visualized as staged studios in Kentucky. (1) The initial problem statement counties with double burden of high cancer rates and low connectivity derived from geographic information system (GIS) observations across the United States. (2) Contextual inquiry and coalition building in the region was performed to obtain context, refine the problem statement, and develop a network of stakeholders. (3) Design studio 1 in Lexington focused on problem refinement and early discovery. (4) These findings were presented to a new group in McKee to ideate and generate low-fidelity concepts (solutions). (5) These concepts were refined in studio 3 to generate a single service design and midfidelity prototype. (6) In studio 5, high-fidelity, working prototypes were developed in a 2-day sprint. (7) These prototypes were tested in a concurrent, multisite 2-day usability pilot with diverse stakeholders. GIS: geographic information system; SOAR: Shaping Our Appalachian Region.



Test

Overview

Usability testing (activity 7) was conducted over the course of 2 days at booths at 2 separate locations. A booth was set up at the outpatient clinic area of University of Kentucky Markey Cancer Center (MCC) to collect feedback from a convenience sample, including patients, caregivers, and providers. Another booth was set up during the MCC Affiliate Network (MCCAN) Annual Meeting at the Lexington Convention Center held on October 9-10, 2019. The MCCAN Annual Meeting was attended by more than 300 cancer providers and researchers from across Kentucky and provided an excellent venue to collect feedback from a knowledgeable pool of attendees. Participants who came to these 2 booths were introduced to the paper versions of the instruments first, followed by the app versions. Providers and researchers were also invited to evaluate a digital dashboard developed specifically for providers receiving myPath data. Participants were given an opportunity to use these prototypes, and at the end, to complete a usability questionnaire to evaluate the respective prototypes against the gold standard, the paper-based NCCN Distress Thermometer (DT).

Survey Instrument

In all, 3 usability surveys were deployed in the usability testing booths. The first survey evaluated the paper versions of the two instruments (NCCN DT vs myPath); the second survey evaluated the digital versions of the 2 instruments; and the third survey evaluated the web-based myPath provider dashboard. In addition to collecting basic demographic data, we used the System Usability Scale (SUS) [40,41] to measure perceived usability and learnability of the standard instrument and co-designed prototypes.

Data Analysis

Descriptive statistics were reported. Independent 2-tailed *t* tests were used to compare the app version of myPath with the paper version of the NCCN DT and test the app version of either instrument versus the paper version of the same instrument. Paired *t* tests were used to compare the SUS scores of the NCCN DT versus myPath in the same format (either the paper or app version). A 1-way analysis of variance was conducted to compare mean SUS scores among different types of users. Results in multiple post hoc comparisons were adjusted using the Šidák method [42]. Analyses were performed in SPSS software (version 27; IBM Corp).

Ethics Approval

This study received ethics approval from the University of California, San Diego institutional review board (record number 180589).

Results

Kentucky Co-design

We planned and carried out a sequence of community-based, participatory activities in Kentucky (Figure 2). (1-3) Framing and contextual inquiry set the stage for participatory design work. (4) Ideation began in Pikeville, then coalesced at an ideation studio in Lexington, which engaged patients, providers, and caregivers to frame problems in cancer care and conceptualize new strategies to solve these problems. This work contextualized the next design studio in an Appalachian county

(Jackson) where community members, patients, and caregivers were encouraged to brainstorm granular solutions to problems generated in the Lexington studio or to make practical those solutions that had been considered. (5) These concepts were presented at a subsequent co-design sprint in Lexington to select a lead candidate. (6) We then prototyped and refined this concept and rapidly developed working interventions to *test* in the real world with patients, providers, caregivers, and others. (7) The outcome of this work was a new software application and service model developed with participants and delivered to stakeholders for implementation and formal testing.

Outcomes

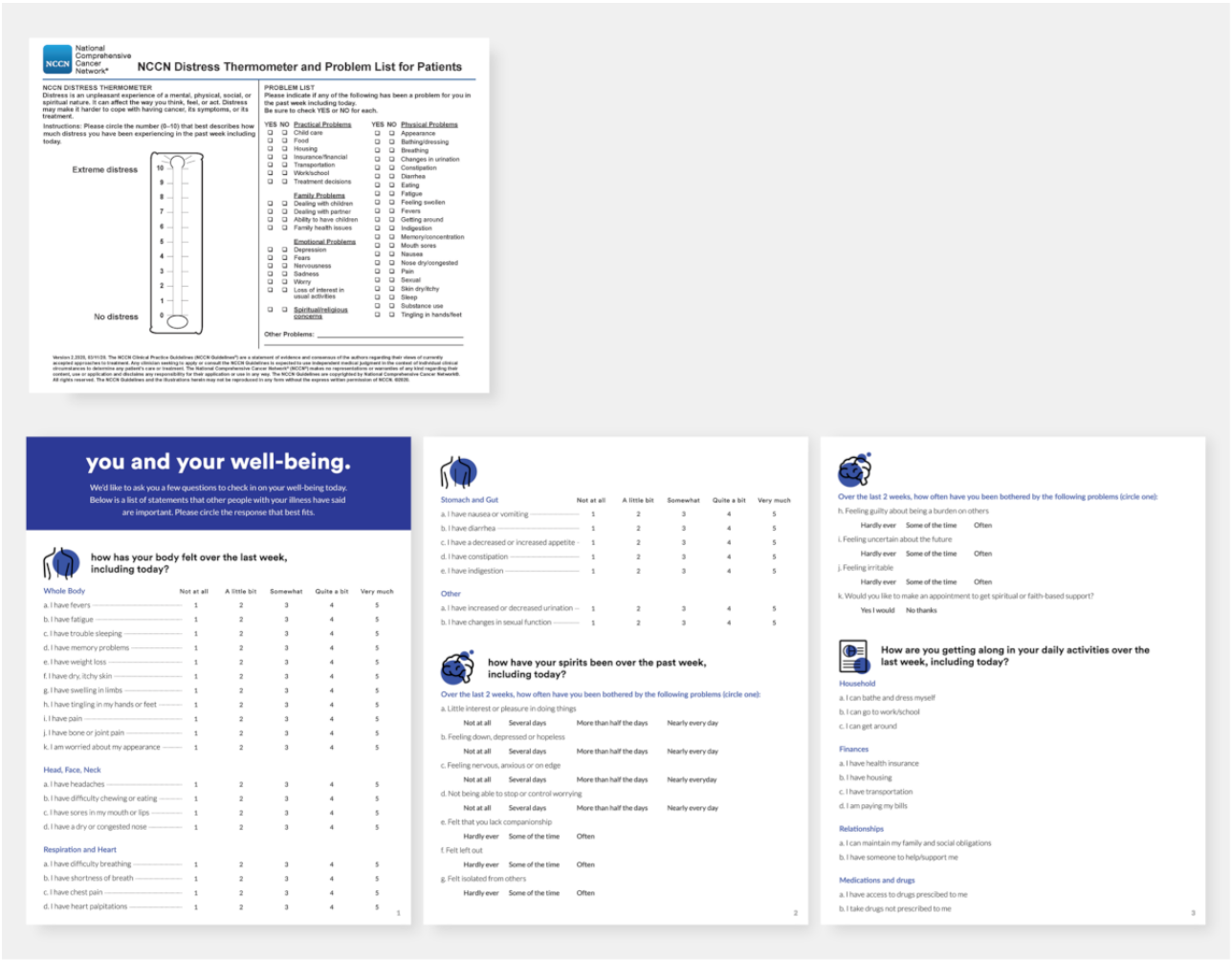
Overview

We have summarized the key outcomes of our design process in Table 1, showing the progression of the co-design activities. In the following sections, we detail the key findings of the refine, implement, and test activities in which we developed paper-based and digital prototypes and conducted usability testing of the 2 tools.

Paper-Based Prototype Development

As the existing NCCN tool is paper-based, at our design studio on October 7, 2019, we first developed a comparable paper tool inspired by co-design and cocreation work from prior studios. Key findings from these sessions (Table 1) produced the following requirements (and potential pitfalls of this approach): (1) compelling user experience, (2) framing the problem from patient perspective (*my wellness* and *my path* vs *Distress Thermometer*), (3) framing questions positively, (4) adding patient-centered questions, and (5) giving feedback and actionable information upon completion of the survey. The finalized design of the newly developed instrument is shown in comparison with the current accepted approach (NCCN DT) in Figure 3.

Figure 3. Comparison of paper prototypes. The gold standard National Comprehensive Cancer Network (NCCN) Distress Thermometer is shown (top) with the newly designed paper instrument developed to track patient progress and well-being.



Digital Prototype Development

Certain feedback from the design studios could not be implemented using paper-based methods alone and required digitization, for example, to get feedback based on answers and to communicate patient distress with providers in a timely manner. In the Lexington co-design studio on October 8, 2019, we worked with participants to co-design a digital version of the new paper tool (named myPath by the participants) as well as a digital version of the NCCN DT (Digital DT [DDT]) not only to assess incremental improvement from added functionality, but also to discern the role of digitization alone. Highlighted user requirements for myPath included compelling colors, content matching the new paper tool, new summary

screen giving immediate feedback upon completion, and actionable insights with instructions and connection to providers. If patient-reported symptoms or needs through the myPath mobile app are over a predetermined threshold, an email alert will be sent to the care team and patients' reports will be highlighted on the dashboard for providers to review. The care team will then decide the best approach to intervene. The developed DDT and myPath app are shown in Figure 4.

Recognizing the other side of the patient–provider dyad, we prototyped a provider-centered dashboard meant to integrate with user-generated data in the myPath app. The key requirements for the dashboard as developed by the participating providers were as follows: (1) make it easy and clear to read, (2) only show critical information, (3) make it clear what has been done or what needs to be done, and (4) minimize disruption to provider workflows. The developed dashboard is shown in Figure 5.

Figure 4. Comparison of *digital* distress thermometer (left) and myPath app (right).

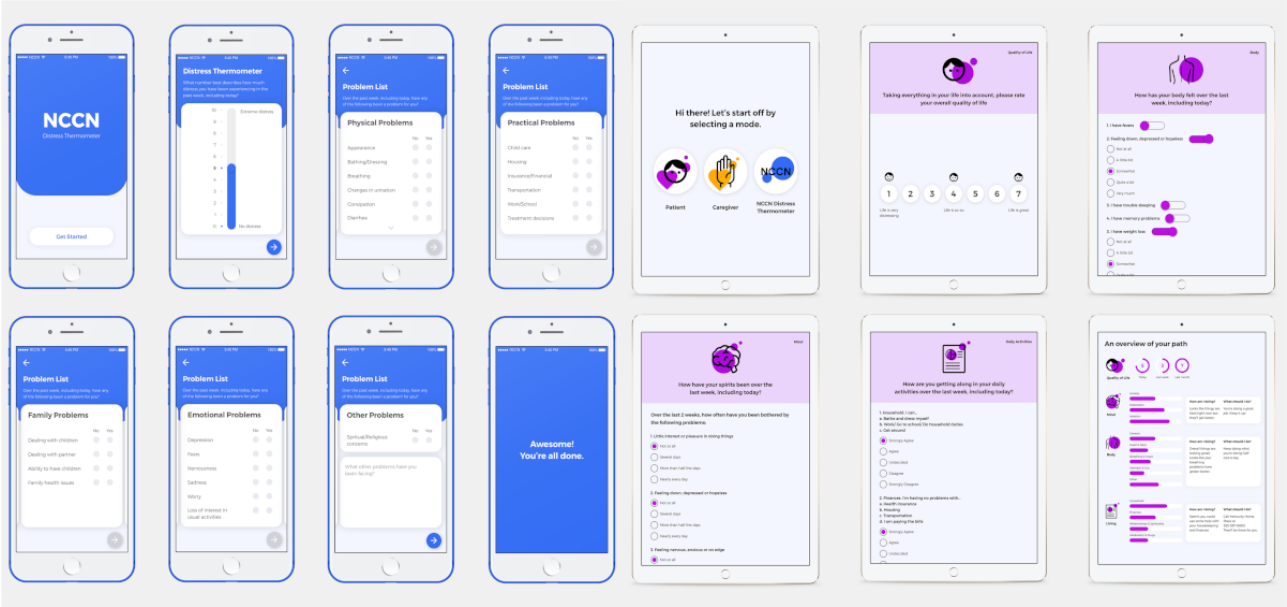
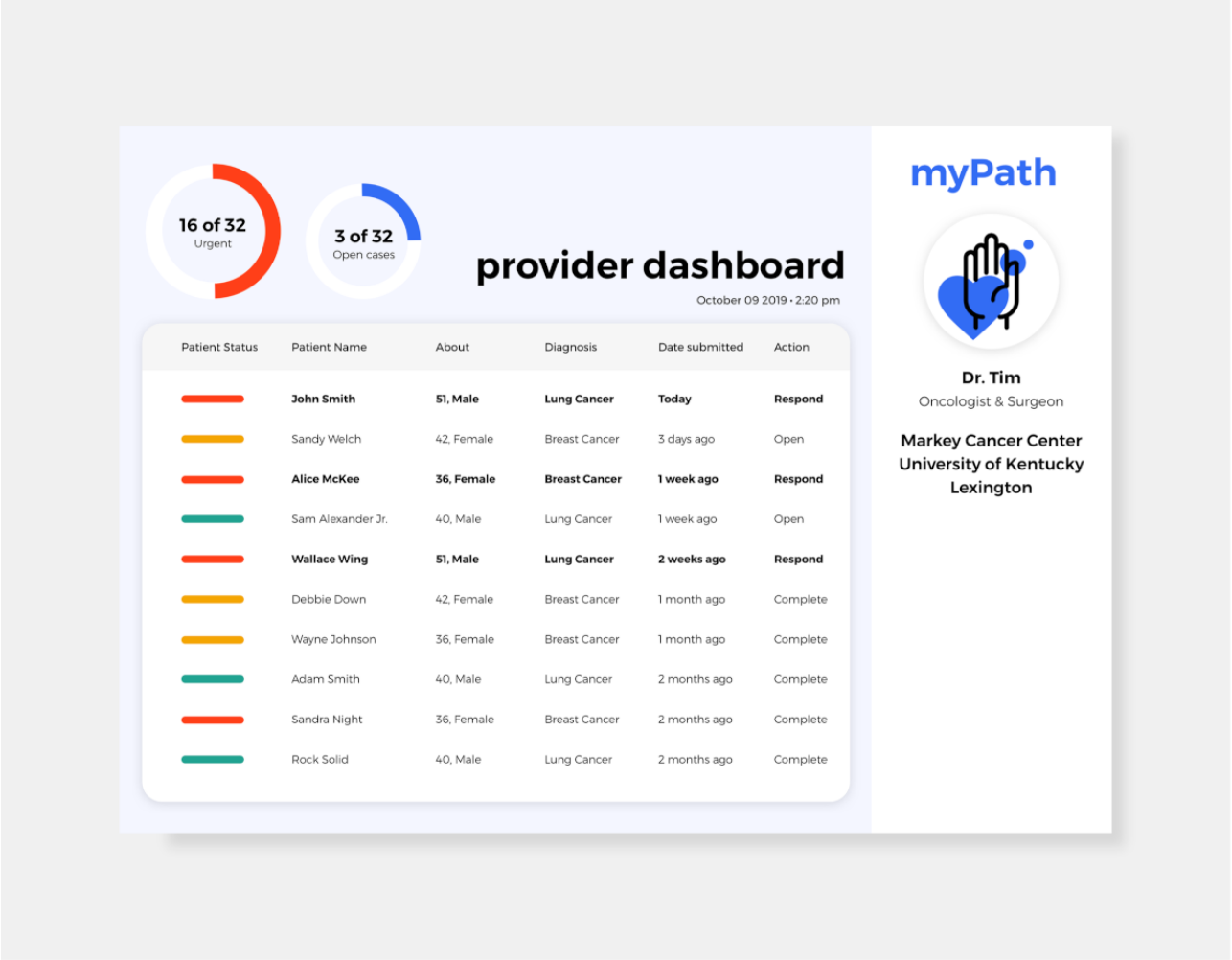


Figure 5. Clinical dashboard prototype.



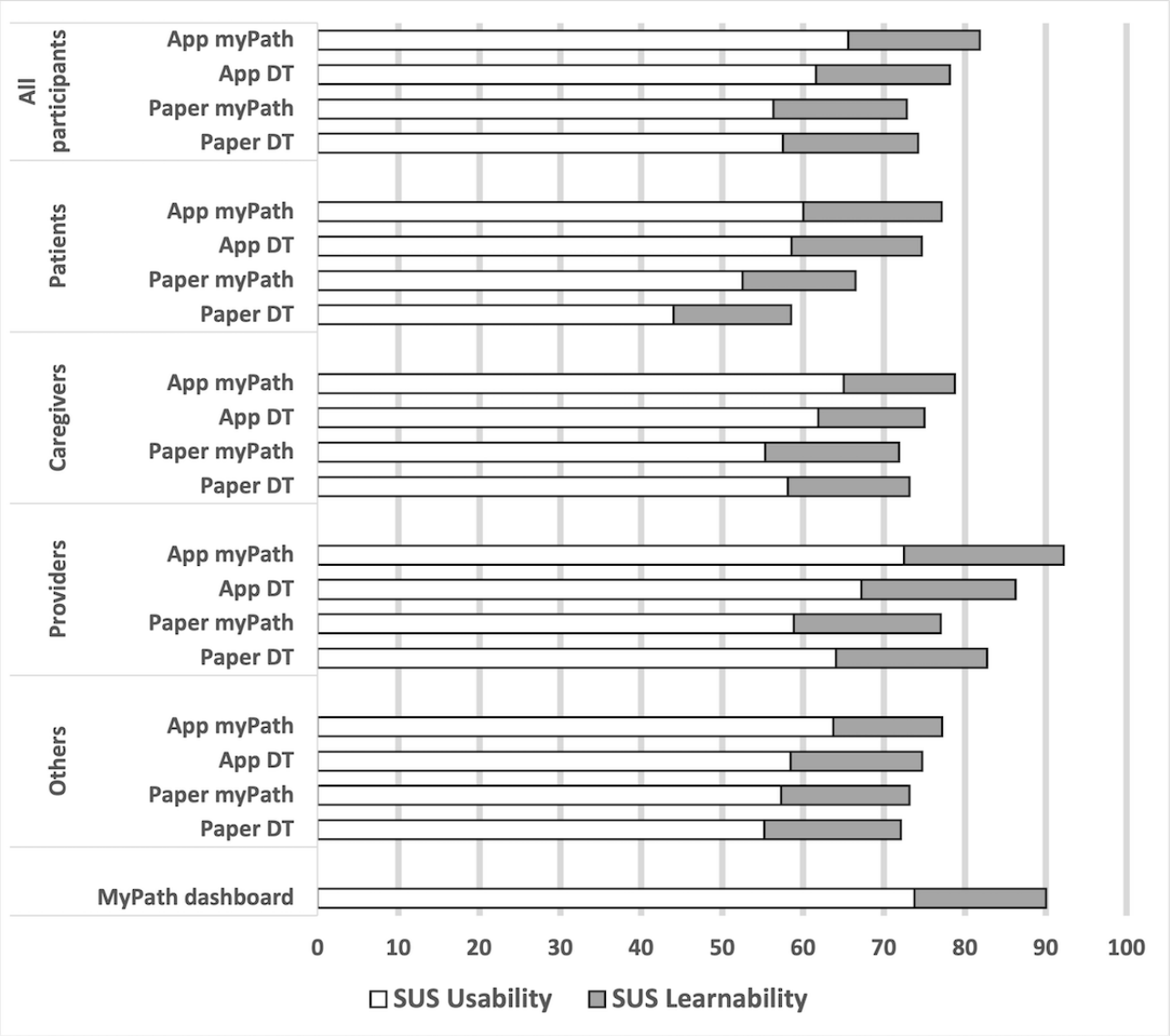
Prototype Testing

Overview

On both days (October 9 and 10, 2019) and at both locations (MCC clinics and MCCAN Annual Meeting), 86 participants completed a usability survey. Of the 86 participants, 46 (53%) evaluated the paper versions of both instruments of the NCCN DT and myPath, 34 (40%) reviewed the app versions, and 6 (7%) assessed the myPath dashboard. After removing surveys with missing data in the SUS questions, of the 86 participants, 38 (44%) evaluated the paper versions. Of these 38 participants, 5 (13%) were patients with cancer, 8 (21%) were caregivers, 11 (29%) were providers, there were 12 (32%) others (eg, technologists, payers, and service providers), and there were 2 (5%) with missing data on this question. Of the 27 participants who evaluated the app versions, 7 (26%) were patients with cancer or cancer survivors, 4 (15%) were caregivers, 8 (30%) were providers, and there were 8 (30%) others. Of the 6 participants who evaluated the dashboard, 4 (67%) were providers and 2 (33%) were researchers. Overall, more than half of the participants were aged >50 years, and 96% (83/86) used a smartphone. Among the 83 smartphone users, 64 (77%) had an iPhone.

Figure 6 shows the results of usability surveys with participants, plotted by participant and prototype assessed. We show the sum and 2 SUS subscales for the digital myPath (App myPath), the NCCN DDT (App DT), the paper prototype of myPath (Paper myPath), NCCN DT (Paper DT), and the provider dashboard. The error bars represent the SDs of the subscales. On average, the overall SUS scores of both instruments for both the paper and app versions as well as the dashboard are higher than a cutoff value at 68-70, showing that both instruments, in either format, have above-average usability [41,43]. Specific comparisons, including statistical testing results, are presented in the next sections. However, because this study was not designed or powered to detect the differences among instruments, the comparative results reported here should be considered preliminary evidence.

Figure 6. Stacked bar chart of System Usability Scale (SUS) subscales. DT: Distress Thermometer.



myPath App Versus Paper DT

Across participants, the digital myPath app showed the highest perceived combined usability (mean 81.9, SD 15.2) compared with the current gold standard of distress management for patients with cancer, the Paper NCCC DT (mean 74.2, SD 15.8). Testing of the SUS subscales showed that the myPath app had significantly better usability than the Paper DT (t_{63}=2.611; *P*=.01; Cohen *d*=0.657, 95% CI 0.148-1.161), whereas the learnability did not differ between the instruments (t_{63}=–0.311; *P*=.76; Cohen *d*=–0.078, 95% CI –0.571 to 0.416).

App Versus Paper Version of the Same Instrument

To discern differences that could be attributed to content versus digitization, we compared paper and digital approaches (Figure 6). Here, the digital version of myPath showed a significantly higher overall SUS score than the paper version (t_{63}=2.345; *P*=.02; Cohen *d*=0.59, 95% CI 0.084-1.092), and this difference was significant for the usability scale (t_{63}=2.991; *P*=.004; Cohen *d*=0.753, 95% CI 0.24-1.261), although not for the learnability scale (t_{63}=–0.157; *P*=.88; Cohen *d*=–0.04, 95% CI –0.533 to 0.454). However, the difference between the DDT and Paper DT was not statistically significant (*P*=.34 and *P*=.24, respectively). Across participant roles, providers (mean 92.19, SD 7.61) reported significantly higher SUS scores (t_{17}=3.164; *P*=.006; Cohen *d*=1.47, 95% CI 0.419-2.488) for the myPath app than for the paper myPath (mean 77.05, SD 11.82). No significant difference was found when comparing providers' ratings of DDT versus Paper DT (t_{17}=0.567; *P*=.58; Cohen *d*=0.263, 95% CI –0.655 to 1.174).

myPath Versus NCCN DT in the Same Format

The paper myPath tool had a slightly lower overall SUS score (mean 72.8, SD 15.3) than the paper DT (mean 74.2, SD 15.8). In contrast, the myPath app (mean 81.9, SD 15.2) had a slightly higher SUS score than the DDT (mean 78.1, SD 17.1). However, these differences were not statistically significant (*P*=.60 and *P*=.20, respectively). These comparisons were also not statistically different in the 2 SUS subscales.

Patients' Versus Providers' Ratings of the Same Instrument

Compared with patients, providers reported higher SUS scores for all instruments (Figure 6). In the 1-way analysis of variance, the overall SUS score for the Paper DT differed significantly between participant roles ($F_{3,32}$=3.28; P=.03). Specifically, patients reported a significantly lower overall SUS score than providers (Cohen d=–24.23, SE=7.85; P=.03; 95% CI –46.24 to –2.21). This difference was mainly seen in the usability subscale ($F_{3,32}$=3.21; P=.04) and not the learnability scale ($F_{3,32}$=1.902; P=.15). The same test did not show significantly different SUS scores between participant roles for the paper myPath, DT, or myPath app.

Discussion

Summary of Results

In this study, we report the results of an ongoing effort to improve the resilience of an oncology system, and a health care system in general, that is struggling to provide equitable access to care for patients both within and outside of the clinical encounter. Specifically, we identified the problem of remote distress monitoring for patients with cancer as an emblematic first step by which a public–private partnership could demonstrate the value of connected cancer care through innovative design methods that democratize the development of solutions. In our efforts, we began with the recognition that successful implementation of connected health solutions in real-world settings requires more than providing broadband access to patients and their care teams. Such an approach requires a careful restructuring of the local workflows and communication channels needed to give patients and their providers the confidence to engage in proactive care irrespective of physical distance or scheduled appointments. Moreover, re-engineering workflows is not something that can be done successfully from the top down; adaptation must be driven by the very individuals who understand the local context. Moreover, this process must have testable, objective outcomes to drive iteration to a clear win state that recognizes both institutional and individual needs.

To facilitate this local adaptation, we created a design and implementation framework (Table 2) informed by our collective experience in human factors, cognitive engineering, anthropology, public health, health communication, design, epidemiology, and clinical medicine. The framework began with a problem-identification stage, which framed geographic areas for contextual inquiry through the combined triangulation of epidemiologic data from registries of cancer burden with industry-level databases on broadband access. The results of these quantitative studies were then used to guide further discovery, problem refinement, and ideation through ethnographic observations, interviews, and semistructured group meetings in those areas of Appalachian Kentucky struggling with the double burden of poor cancer outcomes and lack of access to broadband. These locally co-designed solutions were then prototyped, refined, and tested as part of an implementation solution in Appalachian Kentucky. Clinical trials are currently underway to test the efficacy of these solutions.

Table 2. LAUNCH (Linking and Amplifying User-Centered Networks through Connected Health) framework: lessons learned.

Step	Innovation	Lessons learned
Identify	GIS[a] analysis: data sharing across organizations and analytic teams helped to identify double-burden regions	Calibration of parameters across data sets took time; once calibrated, the resulting geographic maps served to focus community efforts
Discover	Cognitive ethnography: used a blended methodology combining cognitively based protocol analysis (focusing on how individuals perceive, process, and use cancer information) with ethnographic techniques from medical anthropology designed to elucidate culture, norms, roles, and values	Cognitive inquiry proved to be useful for improving wording, formatting, and sequencing at the individual user level, whereas the ethnographic work helped address systemic implementation at the community level
Define	Tactical refinement: a high-level, cross-sector *think tank* was convened to review data and establish priorities in follow-up to information gathered during the discovery phase	Strategic discussions were most productive when they transcended misaligned incentives to identify mutually agreeable objectives across the full ecosystem of care within the targeted communities
Ideate and refine	Co-design: multi-stakeholder teams of clinical and community representatives worked in collaboration with technology developers to co-design an end-to-end system for monitoring and reducing patients' distress	Co-design worked best when it followed a rapid sprint model for prototyping and then testing the key components needed to support productive interactions among patients, caregivers, and clinical teams
Implement	Clinical adaptation of reusable components: the LAUNCH development process yields a reusable library of technologies and protocols, which can then be adapted and evaluated locally within functional systems	Effective clinical care within communities requires adaptation to customize telemedical components using the resources, infrastructure, and people available within local ecosystems of care linked through accountable data structures
Test	LAUNCH-PAD[b]: a platform that allows for pragmatic assembly and testing of crucial components safely within clinical settings	The implementation science needed to customize service structures in a timely and responsive fashion should adhere to pragmatic trial evaluation approaches

[a]GIS: geographic information system.

[b]PAD: Platform for Agile Development.

Our goal in publishing these results at this time is to give an indication of how use of the framework could help inform design decisions from the ground up as communities adapt to the complex interplay of remote telemedical options and in-person consultation in a period of intense change in health care. In our case, the ethnographic work we conducted in collaboration with MCC and MCCAN highlighted a need to adapt the traditional verbiage and approach of the NCCN-mandated stress measures to improve buy-in (adoption) and increase comprehension, consider a wellness frame (as opposed to a sick frame), and increase patient and provider engagement. This is just one aspect of system redesign that may often be overlooked without careful examination of patients' (people's) values and predilections in the geographic regions in which they are served [16]. Once these were identified, we engaged in co-design efforts with these groups to improve how the instrument could address community prioritized needs in a manner that is locally acceptable, effective, and sustainable.

The results from our iterative evaluation process showed incremental improvements in usability and learnability among patients as we progressed through our iterative development path from the paper-based NCCN DT to a reworded version of the paper-based instrument to a digital adaptation of the NCCN DT and finally to an electronic adaptation of the myPath app. Providers, on the other hand, seemed to be less comfortable with the earlier iteration of the paper-based myPath tool than patients. This seemed to be because clinical staff were more familiar with the traditional language in the NCCN DT and thus felt more comfortable with the existing verbiage. Nevertheless, providers responded with even higher usability and learnability ratings than patients when exposed to the final electronic version of the myPath app. Providers also offered extremely high ratings for the myPath dashboard. From our experience, it seems that the clinical staff grew more appreciative of the form and intent of the project the closer it progressed to its final operable format. The oncology teams were especially taken with the enhanced capacity the dashboard offered in management of caseloads across patients.

Implications

When we began our efforts, we relied on an evolving evidence base in cancer care suggesting that strategic deployment of remote, *point-of-need* technologies can offer instrumental gains for improving medical outcomes, protecting quality of life from the burden and cost of travel to appointments, and for offering intervention opportunities well before emergency services may be needed. In other words, we were following a typical evolutionary path for implementation in cancer care; one that—just like telemedicine more generally—would take years to complete.

Since then, we have watched as the necessities for physical distancing during the COVID-19 pandemic have pushed policy makers and health care administrators to move more rapidly on making connected care an integral part of 21st century medicine. Changes have included provisions for reimbursement by the Centers for Medicare & Medicaid Services, a relaxation of jurisdictional barriers across state lines, a softening of privacy restrictions under the Health Insurance Portability and Accountability Act, and the provision of financial incentives for building out broadband support for medicine within underresourced communities. It is unclear at this time how much of this policy change will continue after the pandemic subsides. What is clear is that a focus on HCD will be crucial as we iterate forward toward new service models in a rapidly changing and continuously challenging time. What is also clear is that these changes in the medical landscape will go well beyond the benefits they may convey to cancer care; they will be applicable to all other facets of care as we move to create an antifragile system for patients and the professionals who care for them.

Strengths, Limitations, and Future Direction

This paper presents a framework for guiding HCD activities at the community level. It also offers insight into the process of applying the framework to the specific objective of improving distress monitoring processes for patients with cancer living in rural areas surrounding MCC in Lexington. As noted earlier, these efforts are part of the LAUNCH initiative [11]; therefore, it is only one part of a larger, unfolding story. Clinical pilot studies are currently underway to gauge the overall impact that our HCD efforts are expected to have on system and patient metrics. The true test of a community-based HCD approach will not be completely evident until much farther down the line.

Another limitation of this study is that it offers only one example of how a community-driven design approach could be applied to meet the needs of an academic medical center serving a largely rural catchment area. The way to scale the approach, we believe, is to offer pathways by which other communities around the country could apply best practices in HCD to their own local jurisdictions following a true platform-based model of deployment [44]. Future efforts are planned to create a *Platform for Agile Development* (ie, *LAUNCH-PAD*) to facilitate local resource matching, knowledge transfer, and a guided application of HCD resources and data within the local context.

Conclusions

In this work, we described the implementation and testing of a co-design framework developed to address a global need for rapid health system innovation that generates effective and locally sustainable solutions that can scale laterally to promote resilience in the system. We hypothesized that this antifragile approach, bridging lead-user innovation with equity, diversity, and inclusion through co-design, could promote resilience in rapidly changing and increasingly uncertain times. Through this process we showed that codeveloped solutions addressing community-defined problems could be produced quickly, with broad stakeholder input, balancing goal-oriented, time-limited development with open and inclusive dialog. In conclusion, we encourage bringing together people with lived expertise and diverse and even dissimilar views because this creates the circumstances and creative outputs required for resilient, scalable, and locally acceptable solutions.

Acknowledgments

The authors would like to acknowledge their public health, clinical, and community partners across the state of Kentucky, patients and their families in Kentucky who are affected by cancer, and the health care professionals and informal caregivers who care for those with cancer, and especially the People's Rural Telephone Cooperative for helping the authors to understand how to provide a path forward on connecting patients in rural locations to their care teams. Special thanks to Meghan Johnson and Anna McCowan from the University of Kentucky for invaluable support and to Dr Kevin Patrick for his early vision and contributions to this work. This work was supported by the Behavioral and Community - Based Research Shared Resource Facility at the University of Kentucky Markey Cancer Center (P30CA177558). The University of Kentucky received funding from ICF Macro, Inc, to conduct LAUNCH (Linking and Amplifying User-Centered Networks through Connected Health)–related activities (2018-2019). Amgen provided funding for the University of California San Diego contributions, whereas the Federal Communications Commission provided funding for the Connect2HealthFCC Task Force activities. Further material can be viewed in the Multimedia Appendix 1.

Disclaimer

The statements, analyses, and conclusions set forth in this paper are those of the named authors and do not necessarily reflect the views of the authors' respective organizational affiliations or other organizations that are part of the LAUNCH (Linking and Amplifying User-Centered Networks through Connected Health) Collaborative, including the Federal Communications Commission and the National Cancer Institute.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Think tank overview.
[PNG File , 904 KB - jmir_v24i4e29492_app1.png]

Multimedia Appendix 2
Deep Dive 1: empower people.
[PNG File , 869 KB - jmir_v24i4e29492_app2.png]

Multimedia Appendix 3
Deep Dive 2: activate communities.
[PNG File , 844 KB - jmir_v24i4e29492_app3.png]

Multimedia Appendix 4
Deep Dive 3: pay for value.
[PNG File , 843 KB - jmir_v24i4e29492_app4.png]

Multimedia Appendix 5
Kentucky Design Sprint in pictures.
[PNG File , 1262 KB - jmir_v24i4e29492_app5.png]

References

1. Bashshur RL, Doarn CR, Frenk JM, Kvedar JC, Shannon GW, Woolliscroft JO. Beyond the COVID pandemic, telemedicine, and health care. Telemed J E Health 2020 Nov;26(11):1310-1313. [doi: 10.1089/tmj.2020.0328] [Medline: 32809913]
2. Butler SM. After COVID-19: thinking differently about running the health care system. JAMA 2020 Jun 23;323(24):2450-2451. [doi: 10.1001/jama.2020.8484] [Medline: 32573658]
3. Keesara S, Jonas A, Schulman K. Covid-19 and health care's digital revolution. N Engl J Med 2020 Jun 04;382(23):e82. [doi: 10.1056/nejmp2005835]
4. Slomski A. Telehealth success spurs a call for greater post-COVID-19 license portability. JAMA 2020 Sep 15;324(11):1021-1022. [doi: 10.1001/jama.2020.9142] [Medline: 32930757]
5. Hirko K, Kerver JM, Ford S, Szafranski C, Beckett J, Kitchen C, et al. Telehealth in response to the COVID-19 pandemic: implications for rural health disparities. J Am Med Inform Assoc 2020 Nov 01;27(11):1816-1818 [FREE Full text] [doi: 10.1093/jamia/ocaa156] [Medline: 32589735]
6. Sharpless NE. COVID-19 and cancer. Science 2020 Jun 19;368(6497):1290. [doi: 10.1126/science.abd3377] [Medline: 32554570]

7. Bendell A. Building Anti-Fragile Organisations: Risk, Opportunity and Governance in a Turbulent World. Farnham, United Kingdom: Ashgate Publishing Limited; 2014.
8. Wheelock A, Bechtel C, Leff B. Human-centered design and trust in medicine. JAMA 2020 Dec 15;324(23):2369-2370. [doi: 10.1001/jama.2020.21080] [Medline: 33320236]
9. von Hippel E. Democratizing innovation: the evolving phenomenon of user innovation. Int J Innov Sci 2009 Mar;1(1):29-40. [doi: 10.1260/175722209787951224]
10. Improving cancer-related outcomes with connected health. President's Cancer Panel. 2016. URL: https://prescancerpanel.cancer.gov/report/connectedhealth/ [accessed 2022-01-20]
11. Hesse BW, Ahern D, Ellison M, Aronoff-Spencer E, Vanderpool RC, Onyeije K, et al. Barn-raising on the digital frontier: the L.A.U.N.C.H. collaborative. J Appalachian Health 2020;2(1):6-20 [FREE Full text]
12. McComsey M, Ahern DK, Vanderpool RC, Mullett TW, Chih MY, Johnson M, et al. Experiencing cancer in Appalachian Kentucky. J Appalachian Health 2020;2(3):74-116 [FREE Full text] [doi: 10.13023/jah.0203.09]
13. Wilson RJ, Ryerson AB, Singh SD, King JB. Cancer incidence in Appalachia, 2004–2011. Cancer Epidemiol Biomarkers Prev 2016 Jan 27;25(2):250-258. [doi: 10.1158/1055-9965.epi-15-0946]
14. Rodriguez S, Vanderford NL, Huang B, Vanderpool RC. A social-ecological review of cancer disparities in Kentucky. South Med J 2018 Apr;111(4):213-219 [FREE Full text] [doi: 10.14423/SMJ.0000000000000794] [Medline: 29719033]
15. Laeeq Khan M, Welser HT, Cisneros C, Manatong G, Idris IK. Digital inequality in the Appalachian Ohio: understanding how demographics, internet access, and skills can shape vital information use (VIU). Telematics Informatics 2020 Jul;50:101380. [doi: 10.1016/j.tele.2020.101380]
16. Valdez RS, Gibbons MC, Siegel ER, Kukafka R, Brennan PF. Designing consumer health IT to enhance usability among different racial and ethnic groups within the United States. Health Technol 2012 Jul 13;2(4):225-233. [doi: 10.1007/s12553-012-0031-6]
17. Eller R. Uneven Ground: Appalachia Since 1945. Lexington, Kentucky: University Press of Kentucky; 2008.
18. Samaras GM, Horst RL. A systems engineering perspective on the human-centered design of health information systems. J Biomed Inform 2005 Feb;38(1):61-74 [FREE Full text] [doi: 10.1016/j.jbi.2004.11.013] [Medline: 15694886]
19. Norman DA, Stappers PJ. DesignX: complex sociotechnical systems. She Ji J Design Econ Innov 2015;1(2):83-106. [doi: 10.1016/j.sheji.2016.01.002]
20. Steen M, Manschot M, De Koning N. Benefits of co-design in service design projects. Int J Design 2011;5(2):53-60.
21. Participatory Design Principles and Practices. Boca Raton, Florida: CRC Press; 2017.
22. Carroll J. Making Use: Scenario-Based Design of Human-Computer Interactions. Cambridge: MIT Press; 2000.
23. Holtzblatt K, Wendell JB, Wood S. Rapid Contextual Design: A How-to Guide to Key Techniques for User-Centered Design. Burlington, Massachusetts, United States: Morgan Kaufmann; 2004.
24. Signorelli GR, Lehocki F, Mora Fernández M, O'Neill G, O'Connor D, Brennan L, et al. A research roadmap: connected health as an enabler of cancer patient support. J Med Internet Res 2019 Oct 29;21(10):e14360 [FREE Full text] [doi: 10.2196/14360] [Medline: 31663861]
25. Alfano CM, Mayer DK, Beckjord E, Ahern DK, Galioto M, Sheldon LK, et al. Mending disconnects in cancer care: setting an agenda for research, practice, and policy. JCO Clin Cancer Inform 2020 Jun;4:539-546 [FREE Full text] [doi: 10.1200/CCI.20.00046] [Medline: 32543897]
26. de Moor JS, Mariotto AB, Parry C, Alfano CM, Padgett L, Kent EE, et al. Cancer survivors in the United States: prevalence across the survivorship trajectory and implications for care. Cancer Epidemiol Biomarkers Prev 2013 Apr;22(4):561-570 [FREE Full text] [doi: 10.1158/1055-9965.EPI-12-1356] [Medline: 23535024]
27. Alfano CM, Smith T, de Moor JS, Glasgow RE, Khoury MJ, Hawkins NA, et al. An action plan for translating cancer survivorship research into care. J Natl Cancer Inst 2014 Sep 22;106(11):dju287 [FREE Full text] [doi: 10.1093/jnci/dju287] [Medline: 25249551]
28. Howell D, Mayer DK, Fielding R, Eicher M, Verdonck-de Leeuw IM, Johansen C, Global Partners for Self-Management in Cancer. Management of cancer and health after the clinic visit: a call to action for self-management in cancer care. J Natl Cancer Inst 2021 May 04;113(5):523-531 [FREE Full text] [doi: 10.1093/jnci/djaa083] [Medline: 32525530]
29. Hung P, Deng S, Zahnd WE, Adams SA, Olatosi B, Crouch EL, et al. Geographic disparities in residential proximity to colorectal and cervical cancer care providers. Cancer 2020 Mar 01;126(5):1068-1076 [FREE Full text] [doi: 10.1002/cncr.32594] [Medline: 31702829]
30. Ambroggi M, Biasini C, Del Giovane C, Fornari F, Cavanna L. Distance as a barrier to cancer diagnosis and treatment: review of the literature. Oncologist 2015 Dec;20(12):1378-1385 [FREE Full text] [doi: 10.1634/theoncologist.2015-0110] [Medline: 26512045]
31. Max-Onakpoya E, Madamori O, Grant F, Vanderpool R, Chih MY, Ahern DK, et al. Augmenting cloud connectivity with opportunistic networks for rural remote patient monitoring. In: Proceedings of the International Conference on Computing, Networking and Communications (ICNC). 2020 Presented at: International Conference on Computing, Networking and Communications (ICNC); Feb 17-20, 2020; Big Island, HI, USA. [doi: 10.1109/icnc47757.2020.9049733]

32. Roberts ET, Mehrotra A. Assessment of disparities in digital access among medicare beneficiaries and implications for telemedicine. JAMA Intern Med 2020 Oct 01;180(10):1386-1389 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2666] [Medline: 32744601]
33. Riba MB, Donovan KA, Andersen B, Braun I, Breitbart WS, Brewer BW, et al. Distress management, version 3.2019, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw 2019 Oct 01;17(10):1229-1249 [FREE Full text] [doi: 10.6004/jnccn.2019.0048] [Medline: 31590149]
34. Pirl WF, Jacobsen PB, Deshields TL. Opportunities for improving psychosocial care for cancer survivors. J Clin Oncol 2013 Jun 01;31(16):1920-1921. [doi: 10.1200/JCO.2012.48.0566] [Medline: 23610106]
35. Cutillo A, O'Hea E, Person S, Lessard D, Harralson T, Boudreaux E. The Distress Thermometer: cutoff points and clinical use. Oncol Nurs Forum 2017 May 01;44(3):329-336 [FREE Full text] [doi: 10.1188/17.ONF.329-336] [Medline: 29493167]
36. Optimal resources for cancer care (2020 Standards). Center to Advance Palliative Care. URL: https://www.capc.org/documents/474/ [accessed 2021-12-26]
37. Jacobsen P, Norton WE. The role of implementation science in improving distress assessment and management in oncology: a commentary on "Screening for psychosocial distress among patients with cancer: implications for clinical practice, healthcare policy, and dissemination to enhance cancer survivorship". Transl Behav Med 2019 Mar 01;9(2):292-295 [FREE Full text] [doi: 10.1093/tbm/ibz022] [Medline: 30870569]
38. Design kit: the human-centered design toolkit. IDEO. URL: https://www.ideo.com/post/design-kit [accessed 2021-12-26]
39. Canfora G, Cimitile A, Garcia F, Piattini M, Visaggio CA. Evaluating performances of pair designing in industry. J Syst Softw 2007 Aug;80(8):1317-1327. [doi: 10.1016/j.jss.2006.11.004]
40. Brooke J. SUS: a 'quick and dirty' usability scale. In: Usability Evaluation In Industry. Boca Raton, Florida: CRC Press; 1996.
41. Sauro J, Lewis JR. Quantifying the User Experience Practical Statistics for User Research. Amsterdam, Netherlands: Elsevier Science; 2016.
42. Šidák Z. Rectangular confidence regions for the means of multivariate normal distributions. J Am Stat Assoc 1967 Jun;62(318):626-633. [doi: 10.1080/01621459.1967.10482935]
43. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Human Comput Interac 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
44. Parker GG, Alstyne MW, Choudhary SP. Platform Revolution – How Networked Markets Are Transforming the Economy–and How to Make Them Work for You: How Networked Markets are Transforming the Economy - and How to Make Them Work For You. New York: W. W. Norton & Company; 2016.

Abbreviations

DDT: Digital Distress Thermometer
DT: Distress Thermometer
HCD: human-centered design
LAUNCH: Linking and Amplifying User-Centered Networks through Connected Health
MCC: Markey Cancer Center
MCCAN: Markey Cancer Center Affiliate Network
NCCN: National Comprehensive Cancer Network
SUS: System Usability Scale

Edited by R Kukafka; submitted 08.04.21; peer-reviewed by A Bredart, M Rodriguez; comments to author 18.05.21; revised version received 21.06.21; accepted 21.10.21; published 12.04.22.

Please cite as:
Aronoff-Spencer E, McComsey M, Chih MY, Hubenko A, Baker C, Kim J, Ahern DK, Gibbons MC, Cafazzo JA, Nyakairu P, Vanderpool RC, Mullett TW, Hesse BW
Designing a Framework for Remote Cancer Care Through Community Co-design: Participatory Development Study
J Med Internet Res 2022;24(4):e29492
URL: https://www.jmir.org/2022/4/e29492
doi:10.2196/29492
PMID:35412457

Original Paper

Agreement Between Self-reports and Photos to Assess e-Cigarette Device and Liquid Characteristics in Wave 1 of the Vaping and Patterns of e-Cigarette Use Research Study: Web-Based Longitudinal Cohort Study

Elizabeth Crespi[1], MPH; Jeffrey J Hardesty[1], MPH; Qinghua Nian[1], PhD; Joshua Sinamo[1], BA; Kevin Welding[1], PhD; Ryan David Kennedy[1], PhD; Joanna E Cohen[1], PhD

Institute for Global Tobacco Control, Department of Health, Behavior & Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

Corresponding Author:
Elizabeth Crespi, MPH
Institute for Global Tobacco Control
Department of Health, Behavior & Society
Johns Hopkins Bloomberg School of Public Health
2213 McElderry Street
Baltimore, MD, 21205
United States
Phone: 1 410 614 5378
Fax: 1 410 614 1003
Email: ecrespi2@jhu.edu

Abstract

Background: e-Cigarette device and liquid characteristics are highly customizable; these characteristics impact nicotine delivery and exposure to toxic constituents. It is critical to understand optimal methods for measuring these characteristics to accurately assess their impacts on user behavior and health.

Objective: To inform future survey development, we assessed the agreement between responses from survey participants (self-reports) and photos uploaded by participants and the quantity of usable data derived from each approach.

Methods: Adult regular e-cigarette users (≥5 days per week) aged ≥21 years (N=1209) were asked questions about and submitted photos of their most used e-cigarette device (1209/1209, 100%) and liquid (1132/1209, 93.63%). Device variables assessed included brand, model, reusability, refillability, display, and adjustable power. Liquid variables included brand, flavor, nicotine concentration, nicotine formulation, and bottle size. For each variable, percentage agreement was calculated where self-report and photo data were available. Krippendorff α and intraclass correlation coefficient (ICC) were calculated for categorical and continuous variables, respectively. Results were stratified by device (disposable, reusable with disposable pods or cartridges, and reusable with refillable pods, cartridges, or tanks) and liquid (customized and noncustomized) type. The sample size for each calculation ranged from 3.89% (47/1209; model of disposable devices) to 95.12% (1150/1209; device reusability).

Results: Percentage agreement between photos and self-reports was substantial to very high across device and liquid types for all variables except nicotine concentration. These results are consistent with Krippendorff α calculations, except where prevalence bias was suspected. ICC results for nicotine concentration and bottle size were lower than percentage agreement, likely because ICC accounts for the level of disagreement between values. Agreement varied by device and liquid type. For example, percentage agreement for device brand was higher among users of reusable devices (94%) than among users of disposable devices (75%). Low percentage agreement may result from poor participant knowledge of characteristics, user modifications of devices inconsistent with manufacturer-intended use, inaccurate or incomplete information on websites, or photo submissions that are not a participant's most used device or liquid. The number of excluded values (eg, self-report was "don't know" or no photo submitted) differed between self-reports and photos; for questions asked to participants, self-reports had more usable data than photos for all variables except device model and nicotine formulation.

Conclusions: Photos and self-reports yield data of similar accuracy for most variables assessed in this study: device brand, device model, reusability, adjustable power, display, refillability, liquid brand, flavor, and bottle size. Self-reports provided more

data for all variables except device model and nicotine formulation. Using these approaches simultaneously may optimize data quantity and quality. Future research should examine how to assess nicotine concentration and variables not included in this study (eg, wattage and resistance) and the resource requirements of these approaches.

(J Med Internet Res 2022;24(4):e33656) doi:10.2196/33656

KEYWORDS

tobacco; e-cigarette; methodology; internet; photo; survey; self-report

Introduction

e-Cigarettes are devices that heat a liquid to produce an aerosol for inhalation. The device itself typically consists of a battery, an electrical heater (eg, an atomizer or coil), a container (eg, cartridge, pod, or tank) to hold liquid, and a mouthpiece for inhalation [1,2]. e-Cigarette liquids often comprise nicotine, propylene glycol, vegetable glycerin, and flavorants, but can also contain tetrahydrocannabinol, cannabidiol, and vitamins [1,3-5]. e-Cigarette devices and liquids are highly customizable; for example, adjustable power and modifiable coils allow users to alter their device and its settings, and users can mix flavors in various combinations and proportions [5,6]. In 2014, users were able to select from >450 e-cigarette device brands available for sale on the internet [7]. More recent research suggests that a similar number of device brands existed in 2017. Although 178 of the device brands for sale in 2014 had ceased operation as of 2017, many were replaced by newly emerged brands [8]. In the same time frame, the number of unique flavors increased from 7700 [7] to >15,500 [8]. In addition, the number of websites selling refillable modifiable devices increased from 117 to 190 between 2014 and 2017, whereas the number of websites selling disposable cig-a-like devices (ie, devices that look similar to cigarettes) decreased by approximately 50 in the same period [8]. This demonstrates the popularity of customizable devices in the market and the quick evolution of the e-cigarette device and liquid market.

Various constituents that may be associated with negative health outcomes have been found in e-cigarette liquids and aerosols, including aldehydes, carbonyls, and heavy metals among others [9]. e-Cigarette users may have vastly different experiences and exposures to toxic constituents depending on the duration of vaping sessions [10], e-cigarette liquid contents [10,11], device types [12,13], and device settings [6,14]. Evidence also indicates that these characteristics may be important determinants of nicotine delivery and influence the risk and severity of e-cigarette dependence [5,6]. It is important to monitor the use of these device and liquid characteristics to better understand e-cigarette use behaviors and patterns and their impact on health to inform policy decisions about these products. However, the wide range of products available, their high level of customizability, and the rapidly evolving marketplace presents challenges for measurement.

Self-report data in health research can be subject to several biases (eg, social desirability, recall period, sampling approach, or selective recall) that affect the validity and reliability of the data [15]. Although self-reported measures have been shown to be reliable for assessing cigarette use [16,17], their reliability for assessing e-cigarette use varies depending on the particular measure used (eg, number of days in the past 30 days and sessions per day) [18-20] and their reliability for certain e-cigarette device characteristics, such as voltage and resistance, is insufficient [21]. Thus, there is a need to examine novel methods for assessing e-cigarette device and liquid characteristics, such as submission and coding of photos of e-cigarette devices and liquids, to understand the potential advantages and disadvantages of these data collection methods and whether various e-cigarette device or liquid characteristics warrant different approaches to measurement. This study assessed the agreement between self-report and photo-coded data for certain e-cigarette device and liquid characteristics to better understand the potential challenges and advantages of each approach.

Methods

Study Sample and Protocols

Data were from wave 1 (May 2020 to October 2020) of the Vaping and Patterns of E-cigarette Use Research study, which is a US-based longitudinal cohort study following regular vapers (≥5 days per week) aged ≥21 years through a web-based survey about e-cigarette use patterns and behaviors. Participants of wave 1 were recruited via web-based advertisements on Facebook, Instagram, and Craigslist and flyers and business cards distributed by vape shops. Advertisements were posted in 125 US cities selected for their potential to yield respondents who use e-cigarettes (ie, relatively high population).

After clicking an advertisement, participants were directed to a web-based survey hosted by Virginia Commonwealth University's REDCap (Research Electronic Data Capture; Vanderbilt University), a Health Insurance Portability and Accountability Act–compliant secure web application for building and managing web-based surveys and databases. Before answering any questions, participants reviewed a consent form and certificate of confidentiality and were asked if they would like to continue with the survey. Then, they provided their contact information (ie, full name, phone number, email address, mailing address, and date of birth).

In addition to answering survey questions about their use patterns and behaviors, participants were asked to submit photos of their most used e-cigarette device and their most used liquid for that device. Upon completion, participants received US $10 Amazon gift codes via mail. To exclude potential bot activity, survey responses were reviewed for non-English or Spanish alphabet (0/2813, 0%), data suggesting inattention or very low knowledge of e-cigarettes or liquids (eg, indicating that they are aged 25 years and began vaping at the age of 50 years or self-reporting JUUL as a disposable device; 38/2813, 1.35%),

invalid mailing addresses (13/2813, 0.46%), answering the minimum number of questions (this suggests participant may have previous knowledge of survey skip logic; 0/2813, 0%), completing the survey in <5 minutes (24/2813, 0.85%), missing attention check questions (64/2813, 2.28%), straight lining (0/2813, 0%), failed or incomplete identity authentication (eg, did not provide additional information such as utility bill to confirm identity; 765/2813, 27.2%), or multiple survey attempts (162/2813, 5.76%). Participants (464/2813, 16.5%) were also excluded if their device photos were from the internet (as determined by a Google search describing the image or a Google Reverse Image Search), submitted multiple times, not of an e-cigarette device, or taken in a store. JUUL, Vaporesso, and Voopoo were included as examples in the question prompt for device brand, and records suspected to be bots or professional survey takers frequently reported these 3 brands. Therefore, we decided to consider participants who self-reported one of these 3 brands and submitted a photo of a device that did not match the reported brand to be invalid. In all, 0.11% (3/2813) of the records were not reviewed for various reasons (eg, completed survey after the survey closed) and were excluded from the final data set.

Upon completion of the data collection wave, participants were also excluded if they were found to be highly suspicious in a review of suspicious records (71/2813, 2.52%). Records were comprehensively reviewed if they had incentives returned to the sender or contained data issues such as reporting devices intended for cannabis or oils or beginning to use e-cigarettes before they were commercially available in the United States. Records were considered highly suspicious and were excluded if they contained 1 significant data issue (eg, self-reported liquid contains cannabis derivatives) or a combination of multiple data issues (eg, mismatched self-report and photo device or liquid, patterned responses, or self-reported device or liquid characteristics that do not match self-reported device or liquid). A total of 7875 participants completed the screener, of which 4289 (54.46%) participants were eligible. Of the 4289 participants, 2813 (65.59%) participants completed the survey. Of the 2813 participants, 1604 (57.02%) were excluded after implementing the strategies listed above to avoid bots and professional survey takers, resulting in a sample size of 1209 (42.98%). Researchers (EC and JH) used the Google search engine to search text and markings in submitted photos of devices and liquids to identify the device brand and model and liquid brand and flavor by visually matching the submitted photo with Google search results. Then, the device brand and model and liquid brand and flavor were searched on Google to identify, in the following order of priority, manufacturer, academic (ie, journal articles), retail, and review sites for the given device and liquid and record information about key characteristics of the devices and liquids. One site was sufficient for confirming the information for a characteristic; however, up to 3 sites were searched for each characteristic before categorizing the information as missing. If discrepancies were found between sites for a characteristic, information from the site with highest priority was used. If the device brand and model or liquid brand and flavor could not be identified or the information for a particular characteristic could not be found in the photo or on the web, this information was considered missing. An initial round of reliability testing was conducted for this process to ensure high reliability (≥90%) between researchers.

Ethics Approval

All participants provided informed consent. The institutional review board at the Virginia Commonwealth University (no HM20015004) approved the study protocol. The Johns Hopkins Bloomberg School of Public Health Institutional Review Board (no 9277) approved reliance on the Virginia Commonwealth University Institutional Review Board.

Measurements

Device Brand and Model

Self-reported device brand and model were assessed using the following question: “What is the brand AND model of the device (eg, JUUL, Vaporesso Luxe, Voopoo Drag 2, etc.)?” As participants who self-reported JUUL, Vaporesso, or Voopoo were excluded if their self-reported and photo brand mismatched, these 3 self-reported brands were excluded from all device brand calculations (270/1209, 22.33%). Device brand and model from photos were determined by searching the identifying text or markings in the submitted photo on Google and finding a visual match on the web. For all calculations, device brand and model were assessed individually rather than as a combined variable.

Device Reusability

Self-reported reusability of the device was assessed using the following question: “Is the device (1) reusable (ie, you recharge the device when the battery life is low or at 0%) or (2) disposable (ie, you discard entire device when the battery life is low or at 0%)?” For photos, websites obtained from the Google search conducted for each device were reviewed for mentions of the device being disposable or reusable. Then, the records were coded based on the information on the website.

Device Refillability

Self-reported refillability of the device was assessed using the following question: “When the device runs out of e-liquid, do you TYPICALLY (1) discard the empty cartridge or pod and replace with a new and unused cartridge or pod prefilled with e-liquid or (2) refill the empty tank/cartridge/pod with e-liquid from a larger container(s) of e-liquid?” Given that disposable devices cannot be refillable, this question was asked only to participants who indicated that their device was reusable. The information on whether the device was refillable or nonrefillable was obtained from photos by searching the device brand and model on Google and extracting details from websites.

Visual Representation of Adjustable Settings (Device Display)

Self-reported presence of a visual representation of adjustable settings on the device was assessed using the following question: “Does the device have a VISUAL DISPLAY that allows you to see the wattage or other vape settings?” Response options included “yes” and “no.” This question was asked only to participants who indicated that their device was reusable. For photos, a visual representation of adjustable settings was defined as a display (eg, a screen, small light, or dial) that shows

information about settings that can be adjusted by the user; this information was confirmed from websites for the given device.

Adjustable Power

Self-reported ability to adjust device power was assessed using the following question: "Does the device have SETTINGS that allow you to modify power or vapor volume?" Response options included "yes" and "no." This question was asked only to participants who indicated that their device was reusable. For photos, a device was considered to have adjustable power if websites for the given device indicated that the user can customize the wattage or voltage using a dial or button (not by changing the coil or other internal parts).

Liquid Type

Liquid type was assessed using participants' responses to the following question: "Is your most used e-liquid a (1) customized flavor blend-mixed yourself, (2) customized flavor blend-mixed for you by someone else (3) non-customized flavor?" This question was asked only to users of reusable devices with refillable pods, cartridges, or tanks; users of disposable devices or reusable devices with disposable pods or cartridges were presumed to have noncustomized liquids. Agreement between photos and self-reports on liquid type was not assessed; rather, liquid type was used to stratify results by those using customized and noncustomized liquids.

Liquid Brand

Self-reported liquid brand was assessed using the following question: "Do you know the brand of the [cartridge or pod (eg, JUUL, blu, VUSE, etc.]/e-liquid container (eg, Naked 100, Beard Vape, Milkman, etc.)]? [If yes] please specify the brand." As disposable devices do not have separate liquids, this question was not asked to users of disposable devices. Participants who indicated that they use a customized flavor blend were also excluded from this variable as they were not asked about the brand of their liquid. Photos of the liquids were assessed for brand by searching any identifying text or markings in the photo on Google and reviewing websites for a visual match.

Liquid Flavor

Self-reported liquid flavor was assessed using the following question: "What is the flavor of the [device's (for disposable devices)] e-liquid?" Response options included (1) tobacco; (2) tobacco menthol or menthol; (3) mint; (4) a flavor such as fruit, candy, alcohol, coffee, vanilla, or other food/drink; and (5) no flavor. Liquid flavors from photos were identified using flavor descriptions on the website for the given liquid, and then categorized using the e-cigarette liquid flavor wheel developed by Krusemann et al [22]. If a website was unavailable but the flavor was clearly listed on the container, the photo was used to code the flavor. Then, the flavors were grouped to match the survey question on flavors (ie, tobacco, menthol or tobacco menthol, mint, other, or unflavored).

Nicotine Concentration (mg/mL)

Self-reported nicotine concentration of the liquid was assessed using the following question: "Do you know how much nicotine is in the [device's] e-liquid/flavor blend? [If yes] please specify." Respondents indicated whether they reported the concentration as mg/mL or a percentage. When participants reported nicotine concentrations as a percentage, the reported value was multiplied by 10 to obtain the mg/mL; exceptions include JUUL, which is reported to be 59 mg/mL for their 5% pods and 35 mg/mL for their 3% pods [23]; NJOY Ace, which is reported to be 58 mg/mL for their 5% pods and 28 mg/mL for their 2.4% pods [24]; and NJOY Daily, which is reported to be 69 mg/mL (rich tobacco flavor) or 68 mg/mL (menthol flavor) for their 6% pods and 51 mg/mL for their 4.5% pods [25]. For photos, nicotine concentration (in mg/mL) was extracted from photos of the liquid bottles or from the manufacturer, academic, retail, or review sites. If the nicotine concentration was not mentioned on the container and multiple concentrations were available on the web for the given liquid, concentration was considered missing for that liquid.

Nicotine Formulation

Self-reported nicotine formulation of the liquid was assessed using the following question: "Does the [device's] e-liquid/flavor blend contain nicotine salts?" Response options included "yes" and "no." For photos, nicotine formulation was coded as either nicotine salt or free-base nicotine. As websites frequently report when a liquid is a salt but fail to report when a liquid is free-base and 27 mg/mL was the highest confirmed free-base nicotine liquid concentration in our sample, liquids with nicotine concentration ≤27 mg/mL, for which formulation could not be found on any website, were considered to be free-base. Liquids (2/1209, 0.17%) with nicotine concentrations >27 mg/mL, with formulation not found on any website (Vuse Solo Chai and Glas pods) were investigated further and considered to be salts because other liquids by the same brand are exclusively salts, YouTube reviewers indicated the liquid is likely a salt, or other similar devices frequently use salts.

Liquid Bottle Size (mL)

Self-reported liquid bottle size was assessed using the following question: "Do you know the bottle size (in milliliters) of your most used e-liquid/flavor blend that you last purchased? [If yes] please specify the bottle size (in mL)." This question was asked only to participants who indicated their device was reusable and refillable. For photos, bottle size (in mL) was extracted from photos of the liquid bottles or from the manufacturer, academic, retail, or review sites. If the bottle size was not mentioned on the bottle and multiple sizes were available on the web for the given liquid, bottle size was considered missing for this liquid.

Statistical Analysis

Percentage agreement, calculated as the number of records for which self-report and photo data were concordant divided by the total number of records with available self-report and photo data, was calculated for the following variables: device brand, device model, device reusability, device refillability, device display, adjustable power, liquid brand, liquid flavor, nicotine concentration, nicotine formulation, and liquid bottle size. As percentage agreement does not account for agreement expected by chance, Krippendorff α was also calculated for nominal categorical variables. This method was chosen for its versatility in the number of raters and types of data [26,27]. For continuous

variables, intraclass correlation coefficient (ICC) estimates and 95% CIs were also calculated based on a mean rating (k=2), absolute-agreement, 2-way mixed effects model [28].

Percentage agreement, Krippendorff α, and ICC calculations were also stratified by self-reported liquid type (ie, noncustomized or customized liquid) and self-reported device type (ie, disposable device, reusable device with disposable pods or cartridges, or reusable device with refillable pods, cartridges, or tanks). Calculations were conducted using Microsoft Excel and Stata (version 16.1; StataCorp).

Results

Sample Characteristics

Among our sample, 34.33% (415/1209) were aged 21-29 years, 36.15% (437/1209) were aged 30-39 years, 19.35% (234/1209) were aged 40-49 years, and the remaining 10.17% (123/1209) were aged ≥50 years (Table 1). In addition, 53.02% (641/1209) of the participants were female, 44.99% (544/1209) were male, 0.99% (12/1209) of participants selected the "Other" option for gender, and 0.99% (12/1209) of participants selected "prefer not to answer." Most participants were White (919/1209, 76.01%), followed by multiracial (133/1209, 11%), Black (48/1209, 3.97%), "Other" (36/1209, 2.98%), Asian (24/1209, 1.99%), and American Indian/Alaska Native (9/1209, 0.74%) and 1.99% (24/1209) of the participants selected "prefer not to answer." Most participants (1088/1209, 89.99%) used e-cigarettes 7 days per week. Of the 1209 participants, 713 (58.97%) did not smoke in the last 30 days; the remaining 496 (41.03%) smoked cigarettes ≥1 time in the last 30 days.

Table 1. Sociodemographic characteristics of survey participants (N=1209).

Sociodemographic characteristic	Participants, n (%)
Geographic location	
Midwest	216 (17.87)
Northeast	130 (10.75)
South	491 (40.61)
West	372 (30.77)
Age (years)	
21-24	182 (15.05)
25-29	233 (19.27)
30-34	248 (20.51)
35-39	189 (15.63)
40-44	146 (12.08)
45-49	88 (7.28)
≥50	123 (10.17)
Gender	
Male	545 (45.08)
Female	646 (53.43)
Other[a]	10 (0.83)
Prefer not to answer	8 (0.66)
Race	
American Indian or Alaska Native only	9 (0.74)
Asian or Asian American only	26 (2.15)
Black or African American only	45 (3.72)
Native Hawaiian or Pacific Islander only	4 (0.33)
White only	920 (76.09)
Other	42 (3.47)
Multiracial	134 (11.08)
Prefer not to answer	29 (2.39)
Ethnicity	
Hispanic, Latino, or of Spanish origin	129 (10.67)
Non-Hispanic, Latino, or of Spanish origin	1061 (87.76)
Prefer not to answer	19 (1.57)
Annual household income (US $)	
0-39,999	583 (48.22)
40,000-59,999	287 (23.74)
60,000-99,999	209 (17.29)
≥100,000	101 (8.35)
Prefer not to answer	29 (2.39)
e-Cigarette use (days per week)	
5	74 (6.12)
6	26 (2.15)
7	1109 (91.73)
Smoking status	

Sociodemographic characteristic	Participants, n (%)
Nonsmoker[b]	711 (58.81)
Smoker[b]	498 (41.19)

[a]Other gender includes transgender individuals, nonbinary individuals, and so on.

[b]Participants were considered non-smokers if they had not smoked in the past 30 days. Participants were considered smokers if they smoked cigarettes ≥1 time in the past 30 days.

Availability of Data for Both Photos and Self-reports

The number of values excluded ranged from 4.88% (59/1209; device reusability) to 56.49% (683/1209; bottle size; Table 2). Photo values were excluded if device brand or model or liquid brand or flavor could not be identified from the photo (ie, no photo submitted, photo of the liquid was from the internet, poor quality of the photo, multiple devices or liquids in the photo, photo was of refillable or third-party pod or cartridge, or no match was found in Google searches) or the information could not be found on the internet after identifying the device brand or model or liquid brand or flavor. Self-reported values were excluded if the participant self-reported "I don't know," skip logic prevented participant from being asked the given question (eg, participants who indicated that their liquid was customized were not asked the brand of their liquid), or the self-reported response was not able to be cleaned owing to lack of clarity (eg, liquid brand reported as "local vape shop"). Of the 1209 participants, the resulting sample sizes for calculations ranged from 47 (3.89% for device model of disposable devices) to 1150 (95.12% for device reusability; Table 3).

Table 2. Excluded values for photos and self-report data by variable (N=1209).

Variables	Photo only, n (%)	Self-report only, n (%)		Both, n (%)		Total, n (%)
		Excluded for other reasons[a]	Excluded because question was not asked to participant	Excluded for other reasons[a]	Excluded because question was not asked to participant	
Device brand	42 (3.47)	289[b] (23.90)	0 (0)	0 (0)	0 (0)	331 (27.38)
Device model	29 (2.40)	363 (30.02)	0 (0)	211 (17.45)	0 (0)	603 (49.88)
Device reusability	59 (4.88)	0 (0)	0 (0)	0 (0)	0 (0)	59 (4.88)
Device refillability	51 (4.22)	0 (0)	111 (9.18)	0 (0)	8 (0.66)	170 (14.06)
Device display	51 (4.22)	0 (0)	111 (9.18)	0 (0)	8 (0.66)	170 (14.06)
Adjustable power	51 (4.22)	0 (0)	111 (9.18)	0 (0)	8 (0.66)	170 (14.06)
Liquid brand	70 (5.79)	52 (4.3)	307 (25.39)	26 (2.15)	106 (8.77)	561 (46.40)
Liquid flavor	317 (26.22)	37 (3.06)	0 (0)	21 (1.74)	0 (0)	375 (31.02)
Nicotine concentration	321 (26.55)	50 (4.14)	0 (0)	63 (5.21)	0 (0)	434 (35.90)
Nicotine formulation	224 (18.53)	240 (19.85)	0 (0)	80 (6.62)	0 (0)	544 (45.00)
Liquid bottle size	167 (13.81)	72 (5.96)	352 (29.11)	37 (3.06)	55 (4.55)	683 (56.49)

[a]Other reasons include if the participant self-reported "I don't know" or the self-reported response was not able to be cleaned owing to lack of clarity (eg, liquid brand reported as "local vape shop").

[b]Includes records that were excluded because the brand was JUUL, Vaporesso, or Voopoo (n=270).

Table 3. Sample size for calculations by liquid and device type.

Variables	Device types, n (%)				Liquid types, n (%)	
	Overall (n=1209)	DD[a] (n=119)	RDD[b] (n=288)	RDR[c] (n=802)	Noncustomized (n=915)	Customized (n=294)
Device brand	878 (72.62)	108 (90.8)	141 (48.9)	629 (78.4)	645 (70.5)	233 (79.3)
Device model	606 (50.12)	47 (39.5)	57 (19.8)	502 (62.6)	447 (48.9)	159 (54.1)
Device reusability	1150 (95.12)	111 (93.3)	280 (97.2)	759 (94.6)	878 (95.9)	272 (92.5)
Device refillability	1039 (85.94)	N/A[d]	280 (97.2)	759 (94.6)	767 (83.8)	272 (92.5)
Device display	1039 (85.94)	N/A	280 (97.2)	759 (94.6)	767 (83.8)	272 (92.5)
Adjustable power	1039 (85.94)	N/A	280 (97.2)	759 (94.6)	767 (83.8)	272 (92.5)
Liquid brand	648 (53.59)	N/A	251 (87.2)	397 (49.5)	648 (70.8)	N/A
Liquid flavor	834 (68.98)	92 (77.3)	166 (57.6)	576 (71.8)	687 (75.1)	147 (50)
Nicotine concentration	775 (64.10)	85 (71.4)	137 (47.6)	553 (68.9)	602 (65.8)	173 (58.8)
Nicotine formulation	665 (55)	61 (51.3)	106 (36.8)	498 (62.1)	519 (56.7)	146 (49.7)
Liquid bottle size	526 (43.51)	N/A	N/A	526 (65.6)	382 (41.7)	145 (49.3)

[a]DD: disposable devices.

[b]RDD: reusable devices with disposable pods or cartridges.

[c]RDR: reusable devices with refillable pods, cartridges, or tanks.

[d]N/A: not applicable.

Agreement for Device and Liquid Characteristics Between Photos and Self-reports

Percentage agreement was high (≥80%) between photos and self-reports for device reusability, adjustable power, device display, device refillability, and liquid brand (Table 4). Very high agreement (≥91%) was also observed for device brand for all device and liquid types except disposable devices (75%). Substantial agreement (61%-80%) was found for device model for disposable devices and refillable devices; however, agreement was very high (91.2%) for reusable devices with disposable pods or cartridges, though the sample size for this calculation was limited (57/1209, 4.71%). Percentage agreement was high for liquid flavor, though reusable devices with disposable pods or cartridges had lower agreement (79.5%) than other device types (≥91.3%). Moderate to substantial agreement was found for nicotine concentration across device and liquid types (56.2% for customized liquids to 69% for refillable devices); however, this agreement was lower than that for other variables. Percentage agreement varied widely for nicotine formulation (58.5% for reusable devices with disposable pods or cartridges to 93.6% for refillable devices), though it was generally high. Substantial agreement was also found for bottle size overall (74.3%), though agreement was low for customized liquids (64.6%).

These results were largely supported by Krippendorff α calculations; however, agreement based on Krippendorff α was lower than the percentage agreement for several variables (Table 5). Results from the ICC calculations for nicotine concentration and bottle size (Table 6) show lower agreement for these variables than the results of the percentage agreement calculations.

Table 4. Results of percentage agreement calculations by liquid and device type.

Variables	Device types (%)				Liquid types (%)	
	Overall	DD[a]	RDD[b]	RDR[c]	Noncustomized	Customized
Device brand	92	75	97.9	93.6	91.3	94
Device model	72.6	74.5	91.2	70.3	72.5	73
Device reusability	98.8	91	99.3	99.7	98.4	100
Device refillability	96.7	N/A[d]	96.8	96.7	96.3	97.8
Device display	92.9	N/A	92.9	92.9	92.8	93
Adjustable power	93.9	N/A	97.1	92.8	95	90.8
Liquid brand	86.7	N/A	96	80.9	86.7	N/A
Liquid flavor	89.9	91.3	79.5	92.7	89.4	92.6
Nicotine concentration	66.3	65.9	56.2	69	68.1	60.1
Nicotine formulation	86.2	73.8	58.5	93.6	85	90.4
Liquid bottle size	74.3	N/A	N/A	74.3	78	64.6

[a]DD: disposable devices.

[b]RDD: reusable devices with disposable pods or cartridges.

[c]RDR: reusable devices with refillable pods, cartridges, or tanks.

[d]N/A: not applicable.

Table 5. Results of Krippendorff α calculations by liquid and device type.

Variables	Device types (Krippendorff α)				Liquid types (Krippendorff α)	
	Overall	DD[a]	RDD[b]	RDR[c]	Noncustomized	Customized
Device reusability	.93	−0.04	0	0	.93	1
Device refillability	.92	N/A[d]	−0.01	−0.02	.92	−0.01
Device display	.86	N/A	.13	.84	.85	.83
Adjustable power	.88	N/A	.32	.82	.90	.74
Liquid flavor	.78	.61	.66	.69	.79	.45
Nicotine formulation	.72	−0.14	−0.18	.86	.70	.76

[a]DD: disposable devices.

[b]RDD: reusable devices with disposable pods or cartridges.

[c]RDR: reusable devices with refillable pods, cartridges, or tanks.

[d]N/A: not applicable.

Table 6. Results of intraclass correlation coefficient calculations by liquid and device type.

Variables	Device types, estimate (95% CI)				Liquid types, estimate (95% CI)	
	Overall	DD[a]	RDD[b]	RDR[c]	Noncustomized	Customized
Nicotine concentration	0.21 (0.14 to 0.27)	0.01 (−0.14 to 0.18)	0.04 (−0.12 to 0.20)	0.18 (0.10 to 0.26)	0.20 (0.13 to 0.28)	0.16 (0.02 to 0.30)
Liquid container size	0.38 (0.30 to 0.45)	N/A[d]	N/A	0.38 (0.30 to 0.45)	0.35 (0.26 to 0.44)	0.47 (0.33 to 0.59)

[a]DD: disposable devices.

[b]RDD: reusable devices with disposable pods or cartridges.

[c]RDR: reusable devices with refillable pods, cartridges, or tanks.

[d]N/A: not applicable.

Discussion

Principal Findings

Although we found substantial to almost perfect agreement between photos and self-report for all variables measured in this study, agreement and ICC for nicotine concentration was substantially lower than those for other variables assessed. As previous research has identified [29], this may be a result of participants' poor understanding of nicotine concentration labeling and particularly the differences in the units of reported concentrations (mg/mL vs percentage); some users reported JUUL pods at a concentration of 5 mg/mL, though JUUL sells pods at only 35 mg/mL (3%) or 59 mg/mL (5%) [23].

In addition, we found the lowest agreement for device brand and model for participants who were using disposable devices and the highest agreement for those using reusable devices with disposable pods or cartridges. Despite finding the highest agreement for device brand and model among users of reusable devices with disposable pods or cartridges, agreement for liquid flavor was lowest among these users. These differences may be owing to low knowledge of these characteristics among users of certain device types or inaccurate or incomplete information about these characteristics on websites for certain device types. The wide variation in agreement for nicotine formulation across device types suggests a need for future research into how to evaluate the nicotine formulation of a liquid and how to ensure adequate labeling of nicotine formulation, so that participants' self-reports can be more accurate. Given that the lowest agreement was observed in users of disposable devices and reusable devices with disposable pods or cartridges, it may be that users of nonrefillable devices tend to be less informed about the nicotine formulation of their liquid, possibly owing to poor labeling or lack of concern about nicotine formulation. In addition, these may be a result of challenges in finding accurate information about nicotine formulation for disposable devices and reusable devices with disposable pods or cartridges on websites. Customized liquids presented a unique challenge in assessing bottle size, with lower agreement than noncustomized liquids. This is likely because of challenges in coding the photos of customized liquids, as users sometimes refill bottles obtained from previous purchases of brand-name liquids but may be reporting the quantity of refill rather than the actual size of the bottle. It is also possible that users of customized liquids have submitted photos of a noncustomized liquid owing to concerns that we will be unable to use data from an unlabeled customized liquid bottle.

Krippendorff α varied from the percentage agreement for certain variables, which is likely owing to differences in the prevalence of certain characteristics [26,27]. For example, the prevalence of reusable devices with disposable pods or cartridges in our sample with device displays (24/1209, 1.99% for self-report; 2/1209, 0.17% for photo) was low; Krippendorff α can be affected by extreme values of prevalence for a given measure [26,27]. In addition, results of the ICC calculations were lower than percentage agreement for nicotine concentration and bottle size. As ICC also accounts for the magnitude of the differences between values that disagree, this suggests that, when the self-report and photo values for nicotine concentration and bottle size disagree, the magnitude of the difference between the values is relatively large.

Disagreement between self-reports and photos may also be caused by inaccuracies in website data used to code photos (information about device and liquid characteristics are sometimes inconsistent across websites), user modifications of devices in manners inconsistent with manufacturer intended use of the product (eg, some consumers refill JUUL pods, which are intended to be disposable [30], with other liquids) [31], or submission of photos that are not a participant's most used device or liquid. These issues present unique challenges in understanding e-cigarette device and liquid characteristics and warrant future research into understanding the prevalence of these issues and opportunities for potential solutions.

It is also important to note that the number of excluded values differ between self-reports and photos. This is an important consideration in deciding on an approach as missing values can reduce the sample size or lead to a requirement for more resources for recruitment to obtain a sufficient sample size. When discounting self-reported values that were excluded owing to the skip logic of the survey, self-reports have more usable data for all variables except device model and nicotine formulation. Values were excluded based on the skip logic in our survey only to avoid asking unnecessary questions or questions to which users are unlikely to provide reliable responses. Users of disposable devices (119/1209, 9.84%) were not asked about their device refillability, adjustable power, presence of a device display, liquid brand, or liquid bottle size; users of reusable devices with disposable pods or cartridges (288/1209, 23.82%) were not asked about their liquid container size; and users of customized liquids (294/1209, 24.32%) were not asked the brand of their liquid. Although both self-reports and photos may produce fairly accurate results for device model and nicotine formulation, as indicated by the generally high agreement for these variables, photos may ultimately be superior to self-reports for device model and nicotine formulation because they capture more usable data.

The results outlined in this study provide valuable information about assessing several characteristics of e-cigarette devices and liquids. The agreement between photos and self-reports was substantial to very high for all variables included in this study except nicotine concentration. In addition, photos may provide more usable data for device model and nicotine formulation, whereas self-reports may provide more usable data for the remaining variables: device brand, device reusability, adjustable power, device display, device refillability, liquid brand, liquid flavor, nicotine concentration, and liquid bottle size (for users of refillable devices). However, using both of these approaches in tandem may allow for higher data quantity and quality, as values missing from one approach can be supplemented using the other (ie, values missing from photo data may be filled in using data from self-reports and vice versa) and data can be cross-checked between the 2 approaches.

Strengths and Limitations

The strengths of this study include the large sample size, which allows for analyses of subgroups of participants such as users

of various device types. In addition, participants were not aware that they would be required to submit photos for the survey until after the self-report questions had been answered. Therefore, it is unlikely that the agreement data were affected as a result of participants being more accurate in their self-report owing to knowledge of the photo-uploading part of the study. However, the reverse may not be true; it is possible that some participants were more likely to upload photos of the device or liquid they self-reported because they had already reported information about the device or liquid.

Although the questions included in the survey were selected from or based on previously validated questions (eg, PhenX Toolkit) or established surveys (eg, Population Assessment of Tobacco and Health and International Tobacco Control survey) where possible, novel questions were created if a given characteristic had not been previously assessed or validated (eg, device reusability). In addition, the rigorous data review procedures in this study (eg, eliminating participants with very low knowledge of e-cigarette devices or liquids; photos from the internet; or photos with a brand that did not match the self-reported brand for JUUL, Voopoo, and Vaporesso and requesting a utility bill for participants who failed an initial identity authentication) may have resulted in a sample with a larger proportion of highly conscientious participants, and thus, higher agreement between self-reports and photos than we may otherwise have seen. JUUL, Voopoo, and Vaporesso were excluded from device brand analyses. As these were provided in the question prompt for brand as examples, bots and other invalid submissions frequently listed these as brand and submitted photos of other devices. These submissions were excluded from our analyses and, therefore, any valid self-reported JUUL, Voopoo, and Vaporesso submissions would be a match for photo brand. Therefore, the results for brand calculations cannot be extended to include these 3 brands. Results should be interpreted carefully with the understanding that "I don't know" or missing responses were excluded from the percentage agreement, Krippendorff α, and ICC calculations; therefore, the level of agreement found in this study applies only to complete responses. Self-reported bottle size was assessed only among users of refillable devices owing to concerns that participants may not reliably report this information; therefore, the results for this variable may not extend to users of disposable devices or reusable devices with disposable pods or cartridges. Owing to time and resource constraints, our survey only included questions about and provided a photo submission option for participants' most used device and liquid; it is possible that agreement between self-report and photo data may vary when asking about users' alternative devices and liquids. In addition, only 11 characteristics were included in the analyses; results cannot be applied to other variables such as device wattage, voltage, or resistance. The distribution of our sample with respect to gender and race was similar to that of daily e-cigarette users in the nationally representative Tobacco Use Supplement to the Current Population Survey; however, participants in our study were relatively young and had low income. It is possible that participants from different backgrounds may report more or less consistently between self-reports and photos. Totally, 89.99% (1088/1209) of our sample used e-cigarettes on 7 days per week; this may have selected for more knowledgeable e-cigarette users. Percentage agreement may be lower in a population that uses e-cigarettes less frequently as compared with this study. As we decided not to establish a gold standard between self-report and photo data, we cannot establish which method yielded more accurate results. The time requirements for collecting and processing data for each method also was not considered in this study.

Conclusions

e-Cigarette device and liquid data from both self-reports and user-submitted photos can present challenges. The high agreement between self-reports and photos suggests that these 2 methods yield data of similar accuracy for several variables: device brand, device model, device reusability, presence of adjustable wattage, presence of a display, device refillability, liquid brand, liquid flavor, and liquid bottle size. Self-reports provided a higher quantity of data than photos for all these variables except device model and nicotine formulation, for which photos provided a higher quantity of data. Although self-reports may be sufficient in certain studies and for specific variables, using these 2 approaches in tandem presents an opportunity to optimize the quality and quantity of data as it allows data to be cross-checked between 2 sources and provides an additional source when data are missing from one source. Further research is needed to understand how to assess nicotine concentration and other variables not included in these analyses (eg, wattage and resistance), how consistently and accurately the users of disposable devices and reusable devices with disposable pods or cartridges report on their liquid bottle size, the time and resource requirements for successful implementation of each of these 2 approaches (ie, photos and self-reports), and potential new innovative techniques for assessing e-cigarette device and liquid characteristics (eg, video recordings and daily e-cigarette use diaries).

Acknowledgments

The authors thank Katie Li and Vikas Vattipally for their contributions to participant recruitment, photo coding, and data cleaning for this study. They would also like to thank Virginia Commonwealth University's Survey and Evaluation Research Laboratory for their work in data collection. Research reported in this publication was supported by National Institute on Drug Abuse and Food and Drug Administration Center for Tobacco Products under award number U54DA036105. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the Food and Drug Administration.

Authors' Contributions

EC conceived the research question. EC and JJH coded the data. EC, JJH, QN, JS, and KW cleaned the data. EC and QN conducted the data analysis. EC wrote the first draft of the manuscript. All authors contributed to the design of the study, critically reviewed the drafts of the manuscript, and read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Breland A, Soule E, Lopez A, Ramôa C, El-Hellani A, Eissenberg T. Electronic cigarettes: what are they and what do they do? Ann N Y Acad Sci 2017 Apr;1394(1):5-30 [FREE Full text] [doi: 10.1111/nyas.12977] [Medline: 26774031]
2. Vaping devices (Electronic Cigarettes) DrugFacts. National Institutes of Health. URL: https://www.drugabuse.gov/publications/drugfacts/vaping-devices-electronic-cigarettes [accessed 2021-08-13]
3. The facts on e-cigarette use among youth and young adults. Know the Risks: E-cigarettes & Young People. URL: https://e-cigarettes.surgeongeneral.gov/ [accessed 2021-04-01]
4. Outbreak of lung injury associated with the use of e-cigarette, or vaping, products. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/tobacco/basic_information/e-cigarettes/severe-lung-disease.html [accessed 2021-04-01]
5. National Academies of Sciences, Engineering, and Medicine. Public Health Consequences of E-Cigarettes. Washington, DC: The National Academies Press; 2018.
6. DeVito EE, Krishnan-Sarin S. E-cigarettes: impact of e-liquid components and device characteristics on nicotine exposure. Curr Neuropharmacol 2018 May 01;16(4):438-459. [doi: 10.2174/1570159x15666171016164430]
7. Zhu S, Sun JY, Bonnevie E, Cummins SE, Gamst A, Yin L, et al. Four hundred and sixty brands of e-cigarettes and counting: implications for product regulation. Tob Control 2014 Jul 16;23 Suppl 3(suppl 3):iii3-iii9 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051670] [Medline: 24935895]
8. Hsu G, Sun JY, Zhu S. Evolution of electronic cigarette brands from 2013-2014 to 2016-2017: analysis of brand websites. J Med Internet Res 2018 Mar 12;20(3):e80 [FREE Full text] [doi: 10.2196/jmir.8550] [Medline: 29530840]
9. Eshraghian E, Al-Delaimy W. A review of constituents identified in e-cigarette liquids and aerosols. Tob Prev Cessat 2021 Feb 11;7(February):10-15 [FREE Full text] [doi: 10.18332/tpc/131111] [Medline: 33585727]
10. Fuoco F, Buonanno G, Stabile L, Vigo P. Influential parameters on particle concentration and size distribution in the mainstream of e-cigarettes. Environ Pollut 2014 Jan;184:523-529. [doi: 10.1016/j.envpol.2013.10.010] [Medline: 24172659]
11. Yan XS, D'Ruiz C. Effects of using electronic cigarettes on nicotine delivery and cardiovascular function in comparison with regular cigarettes. Regul Toxicol Pharmacol 2015 Feb;71(1):24-34. [doi: 10.1016/j.yrtph.2014.11.004] [Medline: 25460033]
12. Melstrom P, Koszowski B, Thanner M, Hoh E, King B, Bunnell R, et al. Measuring PM2.5, ultrafine particles, nicotine air and wipe samples following the use of electronic cigarettes. Nicotine Tob Res 2017 Sep 01;19(9):1055-1061. [doi: 10.1093/ntr/ntx058] [Medline: 28340080]
13. Farsalinos KE, Yannovits N, Sarri T, Voudris V, Poulas K. Protocol proposal for, and evaluation of, consistency in nicotine delivery from the liquid to the aerosol of electronic cigarettes atomizers: regulatory implications. Addiction 2016 Jun 21;111(6):1069-1076. [doi: 10.1111/add.13299] [Medline: 26756124]
14. Havel CM, Benowitz NL, Jacob P, St Helen G. An electronic cigarette vaping machine for the characterization of aerosol delivery and composition. Nicotine Tob Res 2017 Oct 01;19(10):1224-1231 [FREE Full text] [doi: 10.1093/ntr/ntw147] [Medline: 27281605]
15. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. J Multidiscip Healthc 2016 May;9:211-217 [FREE Full text] [doi: 10.2147/JMDH.S104807] [Medline: 27217764]
16. Brigham J, Lessov-Schlaggar C, Javitz H, Krasnow RE, Tildesley E, Andrews J, et al. Validity of recall of tobacco use in two prospective cohorts. Am J Epidemiol 2010 Oct 01;172(7):828-835 [FREE Full text] [doi: 10.1093/aje/kwq179] [Medline: 20720099]
17. Soulakova J, Hartman A, Liu B, Willis G, Augustine S. Reliability of adult self-reported smoking history: data from the tobacco use supplement to the current population survey 2002-2003 cohort. Nicotine Tob Res 2012 Aug;14(8):952-960 [FREE Full text] [doi: 10.1093/ntr/ntr313] [Medline: 22318688]
18. Doran N, Correa JB, Myers MG, Tully L. Associations between self-reported and biological measures of nicotine consumption among young adult nondaily cigarette smokers. Am J Addict 2020 Nov 02;29(6):471-475. [doi: 10.1111/ajad.13052] [Medline: 32358904]
19. Yamaguchi N, Kechter A, Schiff S, Braymiller J, Ceasar RC, Simpson KA, et al. Critical challenges and creative solutions for quantifying nicotine vaping: qualitative reports from young adults. Nicotine Tob Res 2022 Feb 14;24(3):416-420. [doi: 10.1093/ntr/ntab074] [Medline: 33846743]

20. Li Z, Benowitz-Fredericks C, Ling P, Cohen J, Thrul J. Assessing young adults' ENDS use via ecological momentary assessment and a smart bluetooth enabled ENDS device. Nicotine Tob Res 2021 May 04;23(5):842-848 [FREE Full text] [doi: 10.1093/ntr/ntaa205] [Medline: 33031497]
21. Rudy AK, Leventhal AM, Goldenson NI, Eissenberg T. Assessing electronic cigarette effects and regulatory impact: challenges with user self-reported device power. Drug Alcohol Depend 2017 Oct 01;179:337-340 [FREE Full text] [doi: 10.1016/j.drugalcdep.2017.07.031] [Medline: 28843084]
22. Krüsemann EJ, Boesveldt S, de Graaf K, Talhout R. An e-liquid flavor wheel: a shared vocabulary based on systematically reviewing e-liquid flavor classifications in literature. Nicotine Tob Res 2019 Sep 19;21(10):1310-1319 [FREE Full text] [doi: 10.1093/ntr/nty101] [Medline: 29788484]
23. JUUL homepage. JUUL. URL: https://www.juul.com/resources/What-is-JUUL-Vape-Liquid-All-JUUL-Pod-Flavors [accessed 2021-03-12]
24. NJOY ace pods. NJOY. URL: https://faq.njoy.com/njoy-ace-pods [accessed 2021-04-12]
25. What is the nicotine content of the Njoy daily? NJOY. URL: https://faq.njoy.com/njoy-daily/what-is-the-nicotine-content-of-the-njoy-daily [accessed 2021-04-13]
26. Krippendorff K. Estimating the reliability, systematic error and random error of interval data. Educ Psychol Measure 2016 Jul 02;30(1):61-70. [doi: 10.1177/001316447003000105]
27. 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]
28. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropractic Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
29. Morean M, Wackowski O, Eissenberg T, Delnevo C, Krishnan-Sarin S. Adolescents and young adults have difficulty understanding nicotine concentration labels on vaping products presented as mg/ml and percent nicotine. Nicotine Tob Res 2021 Aug 04;23(8):1389-1397 [FREE Full text] [doi: 10.1093/ntr/ntab007] [Medline: 33433626]
30. What are JUULpods? Ingredients and flavors. JUUL. URL: https://www.juul.com/learn/pods [accessed 2021-03-25]
31. Ramamurthi D, Chau C, Jackler RK. JUUL and other stealth vaporisers: hiding the habit from parents and teachers. Tob Control 2018 Sep 15;28(6):610-616. [doi: 10.1136/tobaccocontrol-2018-054455] [Medline: 30219794]

Abbreviations

ICC: intraclass correlation coefficient
REDCap: Research Electronic Data Capture

Edited by T Leung; submitted 21.09.21; peer-reviewed by J Pepper, N Felicione; comments to author 29.10.21; revised version received 08.12.21; accepted 11.02.22; published 27.04.22.

Please cite as:
Crespi E, Hardesty JJ, Nian Q, Sinamo J, Welding K, Kennedy RD, Cohen JE
Agreement Between Self-reports and Photos to Assess e-Cigarette Device and Liquid Characteristics in Wave 1 of the Vaping and Patterns of e-Cigarette Use Research Study: Web-Based Longitudinal Cohort Study
J Med Internet Res 2022;24(4):e33656
URL: https://www.jmir.org/2022/4/e33656
doi:10.2196/33656
PMID:35475727

Original Paper

Using Smartphone Sensor Paradata and Personalized Machine Learning Models to Infer Participants' Well-being: Ecological Momentary Assessment

Alexander Hart[1], MSc; Dorota Reis[1], PhD; Elisabeth Prestele[2], PhD; Nicholas C Jacobson[3], PhD

[1]Research Group Applied Statistical Modeling, Department of Psychology, Saarland University, Saarbrücken, Germany

[2]Research Group Diagnostics, Differential and Personality Psychology, Methods and Evaluation, Department of Psychology, University of Koblenz-Landau, Landau, Germany

[3]Center for Technology and Behavioral Health, Departments of Biomedical Data Science and Psychiatry, Geisel School of Medicine, Dartmouth College, Lebanon, NH, United States

Corresponding Author:
Alexander Hart, MSc
Research Group Applied Statistical Modeling
Department of Psychology
Saarland University
Campus A2 4
Saarbrücken, 66123
Germany
Phone: 49 6813023130
Email: hart.research@pm.me

Abstract

Background: Sensors embedded in smartphones allow for the passive momentary quantification of people's states in the context of their daily lives in real time. Such data could be useful for alleviating the burden of ecological momentary assessments and increasing utility in clinical assessments. Despite existing research on using passive sensor data to assess participants' moment-to-moment states and activity levels, only limited research has investigated temporally linking sensor assessment and self-reported assessment to further integrate the 2 methodologies.

Objective: We investigated whether sparse movement-related sensor data can be used to train machine learning models that are able to infer states of individuals' work-related rumination, fatigue, mood, arousal, life engagement, and sleep quality. Sensor data were only collected while the participants filled out the questionnaires on their smartphones.

Methods: We trained personalized machine learning models on data from employees (N=158) who participated in a 3-week ecological momentary assessment study.

Results: The results suggested that passive smartphone sensor data paired with personalized machine learning models can be used to infer individuals' self-reported states at later measurement occasions. The mean R^2 was approximately 0.31 (SD 0.29), and more than half of the participants (119/158, 75.3%) had an R^2 of ≥0.18. Accuracy was only slightly attenuated compared with earlier studies and ranged from 38.41% to 51.38%.

Conclusions: Personalized machine learning models and temporally linked passive sensing data have the capability to infer a sizable proportion of variance in individuals' daily self-reported states. Further research is needed to investigate factors that affect the accuracy and reliability of the inference.

(J Med Internet Res 2022;24(4):e34015) doi:10.2196/34015

KEYWORDS

digital biomarkers; machine learning; ecological momentary assessment; smartphone sensors; internal states; paradata; accelerometer; gyroscope; mood; mobile phone

Introduction

Background

In mental health care, learning about the trajectories of a patient's psychological strain often requires repeated verbal interactions. Interviewing the patient in person is not always preferable because of the economic burden posed on clinicians and patients alike. A substitute for these interviews are questionnaires that patients answer in pen and paper diaries or on their mobile devices. Although these methods allow for a relatively cheap and fine-grained examination of the participants' condition, responding multiple times to the same questionnaire can still be tiring and burdensome for patients, leading to increased noncompliance. In the last decade, biomarkers from mobile sensors have emerged as a promising alternative to infer patients' psychological conditions (eg, depression severity). However, thus far, many of the proposed methods incidentally collect plenty of unrelated private data, require specialized apps on the patients' devices, or depend on the environment (eg, cell phone signal strength). Inferring a patient's condition from movement sensors commonly available in smartphones and recorded only while participants fill out a questionnaire could be an alternative that has not been explored so far. Therefore, in this study, we investigated the predictive capabilities of such sparse data collected during an unrelated web-based study. Although this method initially relies on questioning the patient via identical web-based questionnaires, participant burden could be attenuated in the long run by rotating questionnaire topics or omitting parts of the questionnaire.

Theoretical Background

In the last 40 years, our understanding of fluctuations and trajectories of psychological characteristics has benefited greatly from the adoption of ecological momentary assessment (EMA). EMA studies typically focus on assessing a person's state at a set of moments throughout the day. This schedule is then repeated over a number of days of interest (see Shiffman et al [1] for a thorough introduction to the methodology). EMA moves the assessment closer to the real-life occurrence of relevant phenomena. For example, it allows for the recording of mood states before medication is misused [2]. This fine-grained assessment allows investigations to be conducted on the interplay between variables in everyday life while diminishing memory biases [3]. Using EMA across a considerable length of time consequently allows an investigator to inspect and monitor the course of an individual's responses without requiring direct interaction between the participant and investigator. Consequently, EMA could be a fruitful tool in researchers' and practitioners' toolsets as it can be used to discover more about the moment-to-moment changes that are taking place in a person's condition. However, answering a sizable number of questions several times a day over many days can also be disruptive and time-consuming for EMA participants. Therefore, concerns have been raised about whether the burden it imposes on the individual may even undermine the effort of collecting ecologically valid data [1,4].

Reducing burden in general is of fundamental concern for research ethics [5] and, thus, in the long run, it is of concern for patients and practitioners using frequently repeated interviews. Participant attrition, which can be an outcome of burden, has been highlighted as a major threat to validity in eHealth research [6,7]. In the case of EMA methodology, increased burden has also been linked to reduced data quantity and quality [8,9]. Reasons for the additional burden may vary from study to study, but sampling frequency and number of items per prompt have repeatedly been identified as relevant factors [9-13]. Consequently, a common strategy for reducing burden in EMA studies is to shorten the questionnaire [9,14]. At the same time, omitting items from a questionnaire can negatively affect the reliability or validity of the measurement. This poses a dilemma for using EMA to its full potential in research and clinical application.

A way to overcome this limitation is to passively collect behavioral data. With behavioral data, one could attempt to infer participants' states without requiring them to explicitly report their symptoms. Smart devices, which can monitor their users' behavior in multiple ways [15], enable researchers to gather, explore, and leverage such data. Today, many field studies using EMA or comparable methodologies are already using smartphones for assessment [13,16]. There have also been several successful attempts to identify behavioral information that indicates psychological characteristics, often termed digital biomarkers or behavioral markers, which comprise the field of personal sensing [17]. Scholarly articles on the exploration of passively collected data range from inferring participants' traits from their smartphone use [18-20] to inferring momentary expressions [21]. As the field is quite young, it still comprises a variety of techniques, operationalizations, and outcomes and requires more evidence regarding the effectiveness of approaches and measurement validity (see Mohr et al [17] for an overview pertaining to mental health). Focusing on the areas of mental health and well-being, Yim et al [22] recently reviewed studies that explicitly conducted personal sensing alongside smartphone-based EMA or substituted passive sensing for EMA in the context of major depression. Most studies successfully identified participants with depressive symptoms and inferred their stress levels or their levels of (negative) emotion. Consequently, passive sensing appears to be a fruitful option to overcome burden by passively collecting information from participant biomarkers. At the same time, the studies reviewed by Yim et al [22] showed considerable heterogeneity in terms of the sensor data, how the data were collected, and which methods were used to analyze the data.

Open Challenges for Alleviating the EMA Burden With Personal Sensing

Many passive sensing approaches seem to work equally well for inferring states related to mental health. However, the heterogeneity in these approaches leaves unclear which methodological choices should inform a reliable and applicable approach that could alleviate burden in EMA surveys. For example, Mohr et al [17] identified open challenges pertaining to study quality, reproducibility, variability, uncertainty, and privacy. This investigation addressed 3 of these major challenges.

First, passive sensing studies have shown a distinct lack of agreement on the implemented validation strategies. At the same time, choosing an inappropriate validation strategy for the intended application threatens the validity of the inference. In this vein, Saeb et al [23] investigated the cross-validation scheme of 64 passive sensing studies that aimed to infer clinical outcomes. They found that 45% of the studies used a cross-validation that overestimated the capabilities of the models. Hence, choosing a correct training and cross-validation method is of special concern when data from passive sensing will be used to infer unobserved characteristics of the participants. One of the most popular methods in the studies reviewed by Yim et al [22] and Saeb et al [23] is training and cross-validating the models on a random subset of the entire sample. However, Saeb et al [23] noted that personalized models, in which an individual's past relationships between indicators and states are used to derive models that can infer future states based on future indicator data, might be more appropriate. The viability of such personalized models has only been explored in a few studies related to EMA [24-26]. For example, Asselbergs et al [24] collected the self-reported mood of 27 participants at 5 time points each day over the course of 6 weeks. In addition, the authors unobtrusively logged information about phone calls, SMS text messages, screen activation, app use, and camera use. Asselbergs et al [24] used these data to compute personalized models to infer each participant's mood using forward stepwise regression. Within an error margin of 0.5 around the observed scores, the models inferred between 55% and 76% of the responses on average. However, the authors were not able to replicate the rate of 93% that had been presented in an earlier study [27]. As previous research has not explored which validation strategy can be deemed most appropriate for the use case of substituting EMA responses with collected sensor data, our study compared the most common approaches of training and cross-validating models across the entire sample to computing personalized models for each individual. Furthermore, we contrasted the performance of 2 popular machine learning algorithms that are common in passive sensing studies.

A second challenge in determining an appropriate method to alleviate EMA burden is the variety of available sensors and other use data. This confronts researchers with countless degrees of freedom in how features are computed and modeled [17]. Mixing sensors with distinct characteristics (eg, device orientation and geolocation) might underestimate the importance of one feature in favor of the other and, thus, lead to a biased evaluation of the predictive capabilities of a sensor in a certain setting. This mixing and matching of sensors may further lead to an unwanted interdependence of sensor readings and person and environmental characteristics [17]. Cell phone or Wi-Fi signal strength might not be the same indicator between a city and a rural area, and battery capacity as an indicator might be confounded with participants' choice of smartphone. Furthermore, Bähr et al [28] raised concerns about several quality issues related to geolocation data. Unfortunately, geolocation data have been a favorite for inferring depressive symptoms. Underestimating the external factors influencing these data may lead to failed replication when transitioning from small-scale validation studies to studies with larger and more diverse samples. For example, Saeb et al [29] were able to train classifiers to infer the participants' (N=208) semantic location (eg, at home, at friends' place, or at a restaurant) from sensor data but then could not find a substantial connection to self-reported depressive symptoms. Besides the possibility that there may be no such connection, the authors partially attributed this discrepancy from previous literature to sampling from the broader American population instead of relying on samples that had been restricted to a single location (eg, available students). Such interdependencies between sample and sensor readings must be considered when choosing sensors that may be capable of being substituted for or used to complement self-reported data in a wide variety of studies. To avoid such issues, one could also restrict the investigation to a set of sensors that are for the most part independent from environmental interference. Therefore, in this investigation, we chose data from sensors that detect device movement, which are common in smart devices.

A third challenge in substituting EMA responses with sensor data is that the chosen method must not coincidentally inflict burden in any other way. Most studies investigating mobile sensing have collected and combined considerable quantities of data from various sensors throughout the day. This leads to significant requirements regarding storage and processing capabilities on the side of the investigators and might inflict considerable battery drain and unresponsiveness on the participant's device [30]. Furthermore, collecting large quantities of sensor data may encompass information that is inherently personally identifiable (eg, names of other connected devices) or that is able to reveal sensitive details about a person (eg, geolocation) [30]. This level of detail is not necessarily needed to infer a mental state. For example, a feature Saeb et al [31] used to infer depression was the distance covered by participants, not the participants' actual geolocations. However, once data with such detail and semantic information have been collected for a person, the data can be used to reconstruct a large portion of the person's daily habits without the need for elaborate analysis methods [32]. At the same time, practitioners and participants might overestimate the informational security of software [33]. Participants might further dismiss privacy concerns about activity data as part of a boundary management strategy [34,35]. If the requirements for a certain health condition or the study incentives outweigh their concerns, participants might agree to data collection that they would not have consented to under other circumstances. Therefore, modern methods should strive to maximize participant privacy at a technical level. Ideally, data collection methods will ensure this data sparsity without sacrificing predictive performance or potential treatment efficacy [36,37]. Therefore, in our investigation, we chose to restrict the data collection to the times when participants were responding to their EMA questionnaires.

This Study: Concurrent Sensing

Our approach for tackling the aforementioned challenges involved restricting sensor dependencies and narrowing measurement time to a required minimum. Even with such sparse data, personalized models could be fit to a participant's peculiarities within a relatively short amount of time and accurately infer future self-reports only from future sensor data.

We chose 2 sensors that are widespread in smartphones: the acceleration and orientation sensors [38]. These 2 sensors track how participants move and rotate their devices. As sensor readings are device-centric, they are mostly independent of environmental factors such as signal strength. Although accelerometers and gyroscopes are part of the sensor ensemble of many studies, it is rather uncommon for them to be used as the sole source for studies pertaining to EMA. Nevertheless, previous research has demonstrated that they can be suitable for detecting participants' conditions. For example, wireless accelerometers have been used in medical applications to monitor Parkinson disease [39]. Data from acceleration and orientation sensors have also allowed researchers to infer participants' emotions [40,41]. Furthermore, Kern et al [42] showed that the accelerometer data recorded alongside the survey provided information about survey completion conditions (eg, whether a participant moved while taking the survey). Well-being has also historically been linked to posture [43]. In this regard, Kuhlmann et al [44] explored the possibility of inferring momentary subjective well-being from smartphone tilt in a recent preprint and were at least partially successful. Many more studies have shown that movement-related data can be used to infer information about the device holder, from location to personality traits, thus underscoring the predictive capabilities of such data. Summarizing several of these studies, Kröger et al [45] recently even voiced privacy concerns about the accelerometer. Consequently, also for these 2 sensors, there is a need for solutions that allow sensor data to be collected more sparsely while retaining meaningful predictive capabilities.

A solution to this problem is to collect data from sensors only while participants fill out the EMA questionnaire. Data collected in this way are usually referred to as the paradata of the survey. According to Kreuter [46], paradata are behavioral by-products of computer-assisted data collection that are often neglected in studies. In the past, they have primarily been used to reduce survey error within the total survey error framework [47]; for example, by uncovering participants' insecurity while responding to particular survey items [48]. In the same manner that paradata can be collected from web surveys, it is possible to record sensor readings directly from participants' web browsers without relying on additional applications. Such data are naturally restricted to the moments when participants interact with the survey. In this way, readings between time points always correspond to a similar task and are therefore somewhat standardized. This implicit standardization could be helpful for reducing heterogeneity in the data. While completing the survey, participants inevitably move their phones in a particular way. Thus, it is likely that sensor data collected from moment to moment also hold a relevant amount of information that is usually covered up by continuously recorded data. Fluctuations between time points may be more informative than features extracted from sensor readings throughout the day. For example, the change in mean acceleration along a spatial axis while completing the questionnaire might be indicative of increased arousal during that particular measurement period. This subtle information might be discarded in comparison with stronger signals such as acceleration from walking during the entire day.

In our study, we explored such sensor paradata as a solution for alleviating burden in EMA methodology. We collected sensor readings from the accelerometer and gyroscope alongside an unrelated EMA study. Participants in this study were 158 employees who were recruited to report on factors pertaining to their work-related stress and after-work detachment 6 days a week for 3 weeks. We then used those data to train machine learning models on the first 13 mornings of the data collection period to infer all of a participant's self-reported states on the last 5 mornings based on the sensor data from these days. Furthermore, this study contrasted personalized models with between-participant models for the purpose of inferring states. Personalized (or idiographically weighted) models have already been demonstrated to be feasible solutions [23-26]. However, there is no evidence about whether between-participant machine learning models might not also be a feasible or better solution when using movement sensor paradata. Finally, we inspected the results of the algorithms regarding commonalities between personalized models that may lead to a comprehensible interpretation of how characteristics of movement are related to successful inferences. Taken together, this study aimed to answer the following open questions pertaining to substituting privacy-friendly personal sensing for EMA responses: (1) Can future self-reported states be accurately inferred from sparse accelerometer and gyroscope paradata by models trained on past data? (2) Do between-person and personalized models perform differently when inferring such self-reports? (3) Are there sensor features that are particularly suitable for inferring states?

Methods

Recruitment

The data for this investigation were collected as part of a research project by Reis and Prestele [49]. In this project, 158 employees from different professions voluntarily participated in an experience sampling assessment for 3 weeks. The mean age of the sample was 41.6 (SD 10.9) years, and 67.1% (106/158) of the participants were women. Most of the participants (125/158, 79.1%) worked >36 hours per week. The participants were further incentivized to complete ≥50% of the measurement points with a compensation of €30 (US $34.13) or the opportunity to access 3 weeks of web-based mindfulness stress reduction training subsequent to the end of the study.

Data Collection and Procedure

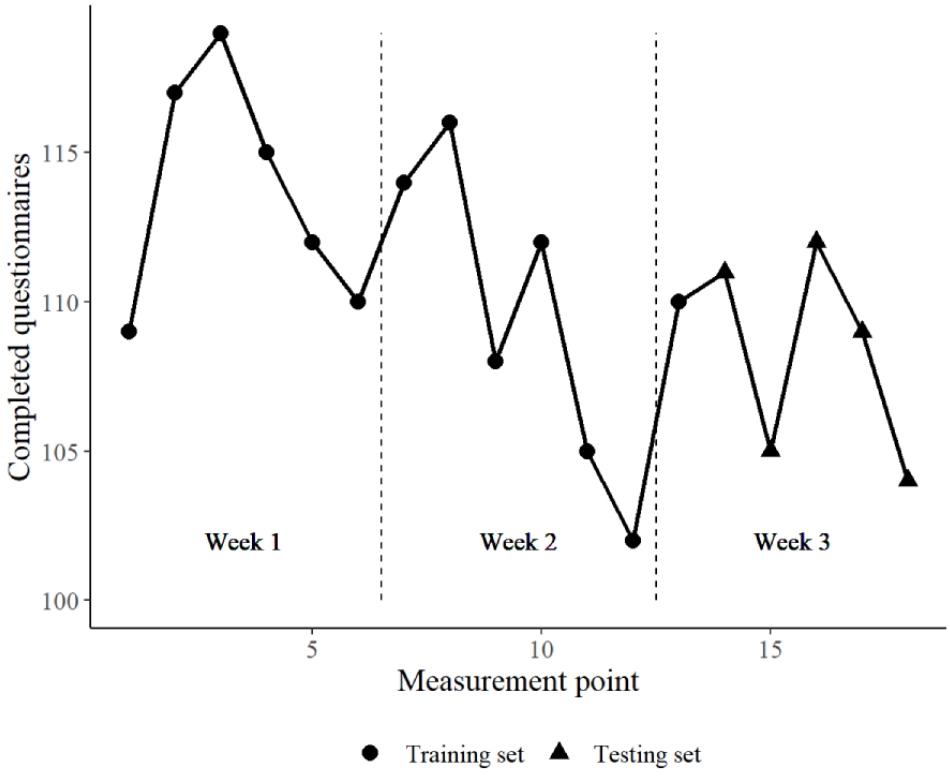
After providing informed consent and completing an intake survey, the participants began the experience sampling procedure. For each workday during the 3 weeks, the participants were prompted in the morning, directly after work, and in the evening to fill out a short questionnaire. On Saturdays, they were asked to fill out only the morning questionnaire. The content of the questionnaire varied throughout the day. Further information about the research project by Reis and Prestele [49] is provided in the publication and in the corresponding repository.

Owing to the time-varying nature of the study's measurement points, we had to choose a subset of constructs that we would try to infer from the sensor data. We decided to omit constructs

that were exclusively related to the participants' judgments of their workplace experiences. The remaining constructs were sleep quality, life engagement, work-related rumination, fatigue, and mood (with 2 subscales). Fatigue and mood were assessed at each measurement occasion, whereas the other 3 constructs were only assessed in the morning. Using all measurement occasions for those 2 predictors would have made them incomparable with the other 3 constructs because of a much larger data corpus (48 instead of 18 measurements) and would have required the inference of mixed trajectories (from morning to evening and from day to day instead of from day to day only). Consequently, we decided to also restrict our mood and fatigue subsample to the morning measurement occasions. This resulted in 18 measurements for mood, fatigue, and sleep quality and 15 measurements for the remaining outcomes as those were not asked about on Mondays. Although all constructs were intended to be predictors, mediators, or outcomes for the model in the initial study, this paper refers to all of them as predicted outcomes of the sensor-related models in this study.

Sensor data were acquired by assessing the JavaScript device orientation application programming interface of the web browser at 2 Hz while the participants filled out the questionnaire, which provided data for the acceleration of the device on 3 axes and the tilting of the device at 3 angles. Although the questionnaires could be completed on a website on any computer, the participants were encouraged to use smartphones, and the questionnaires were optimized for presentation on smaller screens. Most participants complied with this recommendation, resulting in 1995 processable sensor streams out of 2204 measurements, already excluding missing or irregular data (eg, streams with an SD of 0, indicating measurement failure). Overall, adherence in this subsample was comparable between measurement occasions (Figure 1), with a visible decline over the course of the assessment. Nevertheless, the participants completed 15 out of 18 surveys on average (mean 15.11, SD 4.06). On all measurement occasions, between 65% and 75% of the questionnaires were completed, which is on par with previous EMA studies [13]. Over the course of the assessment, there was a noticeable decline in compliance common to EMA studies [9]. The start of the study and Saturdays appeared to be the least favored days by the participants. Consequently, we concluded that compliance in our sample was comparable with that reported in earlier EMA studies.

Figure 1. Number of questionnaires completed during the 3 weeks of assessment. Adherence varied throughout the weeks, with a declining trend toward the end of the assessment.



Measures

Overview

As the sensor paradata for this study were collected alongside an EMA study, we were restricted to scales that the authors had chosen for investigating their hypotheses. All the self-reported measures have been validated and are commonly used in occupational health psychology. For further information regarding the underlying research project, please refer to the article by Reis and Prestele [49].

Fatigue

Momentary fatigue was measured on a 5-point rating scale ranging from 1 (not at all) to 5 (extremely) using a subset of 4 items from the Profiles of Mood State [50].

Mood and Arousal

Mood and arousal were assessed on 6 bipolar adjective pairs of the Multidimensional Mood State Questionnaire (Mehrdimensionaler Befindlichkeitsfragebogen) [51]. The participants expressed their momentary mood on a 6-point scale regarding good mood (feeling well, good, satisfied, or happy) and tense arousal (feeling tense or restless).

Life Engagement

Life engagement was measured using the Utrecht General Engagement Scale–3 [52], a shortened version of the Utrecht General Engagement Scale [53], which is a generalized version of the Utrecht Work Engagement Scale [54,55], which inquires about vigor, dedication, and absorption with a single item per dimension. The participants were asked to report the levels of life engagement they had experienced the previous evening on a 5-point rating scale ranging from 1 (not at all) to 5 (extremely).

Rumination

Work-related rumination was assessed with a selection from the items by Flaxman et al [56] (*I worried about things I need to do at work, I worried about how I would deal with a work task or issue, and my thoughts kept returning to a stressful situation at work*) that were adapted for state use from the perseverative cognition scale [57]. The participants reported the levels of rumination they had experienced the previous evening on a 5-point rating scale ranging from 1 (not at all) to 5 (extremely).

Sleep Quality

The sleep quality of the previous night was assessed using the sleep quality subscale of the Standardized Sleep Inventory [58]. The participants rated three adjectives (good, undisturbed, and ample) on a 5-point agreement scale ranging from 1 (not at all) to 5 (very much).

Extracted Features

To extract features based on the sensor data [17], we followed the recommendations of Hoogendoorn and Funk [59], who proposed that information should be aggregated within participants by computing descriptives such as central tendency, range, and variability. These recommendations were targeted toward working with self-tracking data from wearable activity trackers, thus resembling our approach of recording part of the participants' movements while they filled out the questionnaires. For each measurement occasion, we computed the mean, the maximum of the absolute readings, the SD, the root mean square of successive differences, and the SD of successive differences across all measurements. In addition, we computed autocorrelations and partial autocorrelations up to a lag of 15. In our case, autocorrelation features imply that the device was moved or rotated similarly in a periodic manner (eg, a lag-6 autocorrelation on the z-axis would imply that the smartphone was similarly accelerated along the z-axis every 3 seconds). We also extracted the 2 highest-power frequencies from a Fourier analysis as an additional feature of periodicity in the data.

Statistical Analysis

To investigate relationships between sensor data and outcomes, the sensor streams were arranged to follow the order of the scales during the morning assessment, thereby constructing a continuous sensor stream across all pages of the questionnaire. We then extracted features from each stream for every measurement occasion, and these features served as predictor variables for the later models. The median length of a sensor recording period across participants and measurement occasions was 68.00 seconds (5th percentile=35.00 seconds; 95th percentile=171.00 seconds) if the measurement occasion included life engagement and rumination and 45.00 seconds (5th percentile=21.20 seconds; 95th percentile=107.00 seconds) on days where only fatigue, mood, and arousal were assessed. Next, we split the data into a training set and a testing set following a 70%/30% testing scheme along the timeline of the assessment, resulting in 1455 (mood, fatigue, and sleep) or 1119 (rumination and life engagement) cases in the training set and 543 cases in the testing set. Consequently, we used our models to infer the last 5 mornings of week 3 from learning the participants' idiosyncrasies on all previous mornings of the assessment. The final data set consisted of 230 variables (ie, 6 outcomes and 224 predictors), which we analyzed with 2 different machine learning algorithms using the *caret* package in R (R Foundation for Statistical Computing) with a separate model for each outcome. After analyzing the complete data set, individual models were computed for each participant. We then looked into relationships between self-report characteristics and feature performance that could have determined the predictive capabilities of the models.

We evaluated model performance by inspecting the R^2 values of the resulting models that inferred the participants' states using values in the testing set. In this way, we could assess how well the inferred data points recreated the observed data. For comparability with earlier studies, we further computed the accuracy criterion by LiKamWa et al [27] and Asselbergs et al [24] for the personalized models, allowing for an error margin of 0.5 around the observed value to classify the inferred value as either correct or incorrect. Finally, we explored the relationship between the sensor features and outcomes by clustering feature importance over all outcomes by means of a latent profile analysis (LPA).

Machine Learning Algorithms

To analyze the sensor features, we chose 2 algorithms suitable for supervised learning that have been found to be reliable for different problems: random forest [60] and a penalized general linear model (GLM) [61]. Random forests make use of decision trees, whereas the penalized GLM uses regularization to fit a linear model to the data. The implementations that we chose in R were *ranger* (random forests) and *glmnet* (least absolute shrinkage and selection operator and elastic net regularized GLM). All algorithms were trained with their default values in the *caret* package and a 10-fold cross-validation of the training set.

Ethics Approval

The study received ethics approval from the Ethics Committee of the Department of Psychology of the University of Koblenz-Landau (145_2018). Informed consent was obtained from all individual participants.

Results

Model Performance

Initially, we explored the models' capabilities to infer outcomes in the complete training sample from the set of 224 predictors. The models included the features as well as the unique participant identifier (ie, a character string corresponding to the respective case) and the measurement occasion identifier (ie, an integer corresponding to the place of the measurement within the study) to provide the algorithms with information about the relationship of the repeated measurements. These results resemble a common approach in personal sensing in which a fraction of the entire data set is used to train the models. The results for all models and outcomes are shown in Table 1.

Table 1. R^2 and the root mean square error (RMSE) for all models and outcomes using the full feature set.

Outcome	Random forest		Penalized GLM[a]	
	R^2	RMSE	R^2	RMSE
Training sample				
Sleep quality	0.16	0.87	*0.19* [b]	0.85
Fatigue	0.21	0.85	*0.28*	0.80
Good mood	0.25	0.88	*0.33*	0.82
Tense arousal	0.24	1.13	*0.37*	1.01
Life engagement	0.14	1.03	*0.22*	0.99
Rumination	0.24	1.05	*0.39*	0.93
Testing sample				
Sleep quality	0.12	0.92	*0.16*	0.90
Fatigue	0.21	0.94	*0.29*	0.89
Good mood	0.28	0.94	*0.35*	0.89
Tense arousal	0.23	1.14	*0.32*	1.07
Life engagement	0.21	1.21	*0.25*	1.17
Rumination	0.31	1.06	*0.40*	0.98

[a]GLM: general linear model.

[b]Italics indicate the highest R^2 values for each outcome.

In the training sample, the penalized GLM performed best on all outcome measures. In the testing sample, the penalized GLM again performed the best followed by the random forest model.

Inspecting the most impactful predictors revealed that each model chose one or more unique participant identifiers as their primary source of information. Given that the rating scales restricted the participants' answers to ranging between 1 and 5, we concluded that the models might have identified prototypical participant trajectories that were able to represent a larger proportion of the sample. Considering these results, it was unclear whether the sensor data would provide any meaningful information about the participants' states on their own when the models were trained on between-participant data. Therefore, we continued the analysis by splitting the predictor set into the sensor data and the participant identifier.

Omitting the participant identifier resulted in a reduction in the predictive performance of both models. The R^2 values for the random forests ranged from 0.01 (rumination and life engagement) to 0.05 (good mood) in the training sample and was approximately 0.01 in the testing sample for all outcomes. The R^2 values for the penalized GLM ranged from 0.01 (sleep quality and rumination) to 0.03 (good mood and tense arousal) in the training sample and ranged from <0.01 (life engagement and rumination) to 0.01 (all other outcomes) in the testing sample.

Inspecting the importance of the predictors for the models using only the identifiers showed that the random forests and the penalized GLM chose the same participant identifiers to infer the outcomes except for tense arousal, which both models still inferred equally well. Such similarities were not found in the models without the person identifier.

On the basis of these results, it appeared that the sensor data obtained from the questionnaire were not suitable for training models that could replace self-reports.

Personalized Models

After analyzing the data for the complete sample, we continued to train the models on the individual trajectory of each

participant. Model performance was then evaluated by inferring the same participant's responses on the last 5 mornings. We trained a total of 948 possible models for each algorithm. With the missing data now weighted against each individual model, 302 models could not be computed using random forests (sleep quality: 46/302, 15.2%; fatigue: 41/302, 13.6%; good mood: 48/302, 15.9%; tense arousal: 57/302, 18.9%; life engagement: 52/302, 17.2%; rumination: 58/302, 19.2%). Another 283 models could not be computed for the penalized GLM (sleep quality: 43/283, 15.2%; fatigue: 40/283, 14.1%; good mood: 44/283, 15.5%; tense arousal: 54/283, 19.1%; life engagement: 48/283, 17%; rumination: 54/283, 19.1%). Most (208/302, 75.5%) of the missing models were related to the same 38 participants with very low compliance in the morning questionnaire.

For all outcomes, the R^2 values archived in the testing sample ranged from 0 to 1 for the random forests and penalized GLMs. For 75.3% (119/158) of the participants, the models showed an R^2 value of ≥0.18. The descriptive statistics for the R^2 distributions are presented in Table 2. Furthermore, we computed the accuracy criterion used in earlier studies [24,27]. The mean accuracy of the random forest models ranged from 41.37% (SD 30.86%) for life engagement to 51.38% (SD 32.05%) for good mood (sleep quality: mean 46.4%, SD 31.5%; fatigue: mean 42.4%, SD 29.29%; tense arousal: mean 41.88%, SD 29.62%; rumination: mean 45.06%, SD 33.78%). The mean accuracy for the penalized GLM ranged from 38.41% (SD 29.18%) for rumination to 48.19% (SD 31.48%) for good mood (sleep quality: mean 42.03%, SD 31.46%; fatigue: mean 43.92%, SD 29.87%; tense arousal: mean 41.19%, SD 28.58%; life engagement: mean 40.08%, SD 27.42%). On average, between 38.41% and 51.38% of the participants' answers on the last 5 measurement occasions could be inferred from models trained on the individual sensor paradata of the first 13 measurements. Relating the accuracy to the mean compliance of participants during the last 5 mornings (mean 4.44, SD 0.82), this means that the models correctly inferred between 1.71 and 2.28 answers on average.

Table 2. Descriptive statistics for the distributions of R^2 values achieved on the testing set. Except for the tense arousal outcome, the penalized general linear model (GLM) performed better at inferring the last 5 mornings.

Outcome	Values, mean (SD)	Percentile 0	25th percentile	50th percentile	75th percentile	100th percentile
Random forest						
Sleep quality	0.27 (0.28)	0	0.03	0.18	0.44	1
Fatigue	0.32 (0.32)	0	0.05	0.21	0.54	1
Good mood	0.28 (0.28)	0	0.05	0.2	0.43	1
Tense arousal	0.34 (0.28)	0	0.1	0.3	0.52	1
Life engagement	0.3 (0.3)	0	0.04	0.2	0.49	0.99
Rumination	0.27 (0.29)	0	0.03	0.16	0.41	1
Penalized GLM						
Sleep quality	0.35 (0.33)	0	0.05	0.26	0.63	1
Fatigue	0.28 (0.27)	0	0.05	0.21	0.47	1
Good mood	0.3 (0.28)	0	0.07	0.21	0.51	1
Tense arousal	0.34 (0.3)	0	0.07	0.26	0.59	1
Life engagement	0.33 (0.29)	0	0.07	0.26	0.5	1
Rumination	0.31 (0.31)	0	0.07	0.18	0.48	1

Further Analyses Regarding Features and Performance

To analyze the importance of the features, we summed the feature importance of all the individual models for each outcome. The algorithms predominantly used autocorrelations and partial autocorrelations from both sensors. Both algorithms tended to harness the higher-order autocorrelations, representing a correlation of values between 2 and 5 seconds rather than the lag-1 or lag-2 autocorrelations. The same held true for the partial autocorrelations. For the penalized GLM, the set of important features was complemented by the mean acceleration along the y-axis (vertical movement) for good mood, tense arousal, and rumination. In addition, the mean acceleration along the x-axis (horizontal movement) appeared to be important for inferring tense arousal. The random forest models used the measurement occasion to infer fatigue, tense arousal, and rumination. The algorithm also used the maximum acceleration along the y-axis, the maximum rotation around α (turning the phone to landscape mode), and the root mean square of successive differences of the rotation around α to infer good mood. It also used the SD of the acceleration along the z-axis (moving the phone back and forth) to infer life engagement. None of the algorithms showed a pattern that could be interpreted point-blank in terms of posture or exerted force.

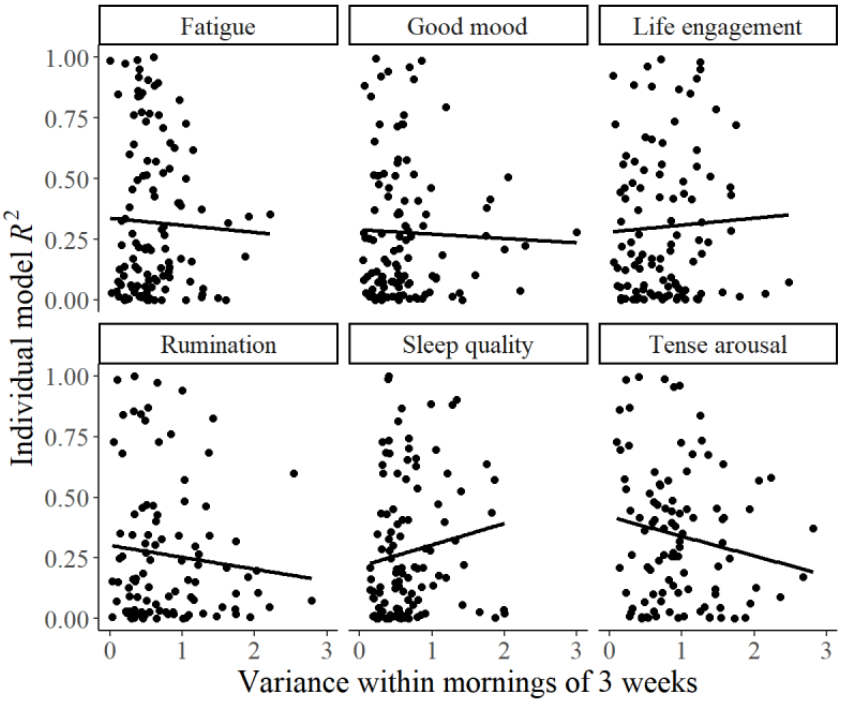
We attempted to identify further similarities between the models by clustering the predictors based on their importance across all outcomes. To determine an empirical number of clusters, we conducted an LPA using the *mclust* package in R [62]. Inspecting the progression of the Bayesian information criterion

(BIC) and the integrated completed likelihood for the random forests, both clustering algorithms favored a 1-cluster solution (BIC_{EEI}=−3828.916; integrated completed $likelihood_{EEI}$=−3828.916) regardless of the cluster parameterization [62]. We explored this further by testing the preferred solution of equally sized and distributed clusters (EII parameterization) with a bootstrap likelihood ratio test, testing successive cluster solutions for an improvement in model fit. This analysis also pointed toward a single-cluster solution (likelihood ratio test $statistic_{1\ vs\ 2}$=7.98; *P*=.54).

Contrary to the LPA for the random forests, the results for the penalized GLMs indicated a solution with multiple clusters. The BIC peaked between a solution with 5 and 7 clusters, with 5 clusters that were variable in size and rotation (VVE parameterized) as the favored outcome. However, the cluster solutions found for the penalized GLM aggregated the features into clusters of similar performance across all models instead of clusters targeted at individual outcomes. The cluster containing the features with the highest importance contained the most features (n=86) and, therefore, provided no advantage over the unclustered importance sums.

Furthermore, we explored the relationships between the variability in the outcomes and model performance. All of the following correlations are Holm-corrected [63] for multiple comparisons. Correlation analyses (2-sided) between the square root–transformed individual models' R^2 results and the log-transformed variability of the self-report measures revealed no meaningful results for either algorithm. This was true for all outcomes as well as for the correlation between performance and variability within 3 weeks or only within the last week. The relationships between the (untransformed) model performance and the 3-week variance are depicted in Figure 2. Next, we included only results with $R^2 \geq 0.01$ in the analysis to explore whether more or less variability in the outcomes was correlated with models that explained ≥1% of the variance. This analysis also revealed no meaningful relationships.

Figure 2. Scatterplots exploring the relationships between the variability of the outcomes across 3 weeks and the performance of the individual random forest models. Although some trends were depicted in the plot, we found no substantial correlations between variability and model performance.



We similarly analyzed the relationships between the questionnaires completed in the first 2 weeks and model performance. As we were not able to correct the skewed distribution of the compliance variable (most participants were very compliant), we calculated a 2-sided percentage bend correlation [64] as a nonparametric alternative to the Pearson correlation coefficient. This resulted in a negative relationship between compliance and model performance for the good mood outcome ($r_{percentage\ bend}(99)$=−0.32; *P*=.02).

Finally, we visually inspected the outcome trajectories as well as the distributions of the sensor feature values between the 25% best-performing models and the 25% worst-performing models. Both visual analyses revealed no evident differences between the best- and worst-performing models.

Discussion

Principal Findings

This study examined the utility of sensor paradata that were passively collected while participants filled out EMA questionnaires to infer participants' future self-reported mental states. Our results suggested that the sparse data collected only from movement-related sensors allowed us to infer participants' self-reports on several outcomes related to mental well-being to some degree. For half of the participants, our models already performed well on all 6 outcomes. Using the accuracy criterion by Asselbergs et al [24], the mood-related personalized models in this study were able to replicate the lower boundary of previous work. Given that our models relied only on sparse data

from the 2 movement-related sensors, this result is quite remarkable. This was even more remarkable given that the sample in our study was 5 times the size and consisted of a rather heterogeneous group of employees. This gives us confidence that further development in this area will refine the assessment to a point where passive sensing can actually replace self-reports.

Another result of our exploration was that our approach did not appear to be suitable for predictions trained on data from all participants, but it was suitable for modeling the states of the participants idiographically. The models resulting from training on the complete data set relied on the peculiarities of interindividual responses and started treating individual participant identifiers as the major source of information. Given the restricted ranges of Likert-type rating scales, deriving prototypes is a perfectly reasonable approach for capturing a good portion of the common variance between people, which could heuristically explain the results of the algorithms. Consequently, the models based on participant identifiers also worked, but the information from passive data had little to no impact on the final models. Another explanation could be that random variation between participants (eg, sensor readings that differed by device [44,65]) might have concealed relevant variance. This supports the point of Saeb et al [23] that models focusing on inferring future data from the same source should use personalized models. Consequently, we also advise future studies to clearly determine and disclose whether their goal is to classify participants or to infer participants' future states with the trained models when choosing on which part of the data models should be trained and tested. As predicting the commonalities of a sample might most likely not be the goal of all studies using passive data and EMA, we suggest that the default of training between-participant models should be treated with caution when inferring participant mental states from sensor data instead of classifying passive data patterns. However, as demonstrated by Jacobson and Chung [25], such models can most likely be used to inform personalized models.

A final question we investigated was whether the working models could be traced back to a pattern of successful predictors that would be indicative of a certain outcome. In the related analyses, we found that, despite finding that the algorithms performed similarly, they treated the data differently. This was not too surprising as random forests rely on subsets based on chance and pick up on nonlinear relationships, whereas penalized GLMs rely on the complete data. For the random forest models, the LPA suggested a 1-cluster solution regarding the importance of features, but the LPA of the penalized GLMs showed no such classification tendencies. Instead, the analysis clustered indicators according to their impact with respect to all outcomes. Although both outcomes are semantically appealing, we would recommend that any interpretation be adapted with caution in future research on the relationships between sensor readings and psychometric outcomes.

We also did not find any intuitive relationships between model performance and the variability of the outcome data, participant compliance, the trajectory of the outcomes, or the distribution of the feature values. Although a singular negative relationship between the collected data and the model performance appeared to be significant, we assume this to be purely by chance despite the applied correction. As we asked about both mood subscales on the same questionnaire page, discovering an effect for only 1 subscale points to the difference in failed models between the 2 outcomes rather than a true correlation with compliance. Consequently, we would not expect to find this association between compliance and model performance in future studies.

Potential Applications

We presented evidence that sparse movement sensor data collected in a privacy-friendly manner contain valuable information about the participant's psychological state. However, so far it has not become explicit how these data might be used to significantly alleviate participant burden in EMA studies. At the moment, we envision 2 approaches that may foster participants' compliance.

First, researchers might be able to omit arbitrary questionnaires after the training period. Reducing the length of the assessment comes with direct benefits for experienced burden and participant compliance [9]. Researchers would start with all relevant questionnaires in the survey and attempt to train personalized models on the relationship between sensor paradata features and scale values. Once those models are successfully established for the required number of participants and constructs, researchers could then omit arbitrary questionnaires from the survey and infer these missing responses by means of the trained models and the sensor paradata from the remaining questionnaires.

Second, researchers might be able to change the content of the survey during the assessment without compromising the completeness of the data set. Researchers would start with a subset of the relevant questionnaires. After training the personalized models on these data, one or more questionnaires could be replaced with other questionnaires. The trained models could then be used to infer the now unobserved constructs. Furthermore, if personalized models are trained on the added constructs, the resulting models could be used to infer the previously unobserved constructs. In this way, longer assessments could be shortened significantly to reduce the initial burden. Participants may also find the assessment less burdensome because of the regularly refreshed novelty.

The latter technique shares similarities with planned missingness designs [66] where certain items of a scale are randomized between participants. However, in contrast to planned missingness, the missing data are not inferred based on the entire sample's answering behavior but through the biomarker provided by the individual participant. Given that further research can improve the accuracy of our method to the rates presented by LiKamWa et al [27], rotating questionnaires flanked by biomarkers might provide data with higher intrapersonal validity than planned missingness designs.

Both methods will still require an initial period to gather enough data to train the personalized models. In addition, we also assume that participants will be required to fill out questionnaires of a similar format after the training period, thus adhering to a semistandardized procedure. Other activities that might be enjoyable for the participants (eg, playing games) but

require more physical activity might considerably affect the validity of the computed sensor features. However, these boundaries of applicability are subject to further systematic research as well as questions about the quantity of viable omissions and substitutions, the length of the required training period, and whether nonrectangular questionnaire formats (eg, slider bars and swipe choice) are suitable for training the personalized models.

Limitations and Future Directions

One of the strongest points of this study is that we were able to demonstrate that models trained on paradata that can be collected while participants fill out a survey can be used to infer self-reported states over time. As demonstrated, this can be done completely unobtrusively alongside an already existing research project without a dedicated study setting focused on the exploration of passive data or special laboratory hardware. However, as this study is still exploratory, we want to emphasize some limitations that come with the degrees of freedom that the researchers will use when conducting research on passive data. As outlined in our theoretical introduction, changing the parameters may lead to different results in future examinations on the topic.

First, choosing the same resolution for the sensors (2 Hz) might be crucial for replicating our results. Smartphone sensors of movement values are usually able to report up to approximately 100 readings per second (100 Hz). Choosing a frequency that is appropriate for the task is an important part of analyzing sensor data [65]. In our study, these rather conservative values were chosen based on three deliberations targeting the applicability of the methods presented in this study:

- Sensor output should remain manageable by statistical software that is typically used in the social sciences. Sensor readings become big data very quickly, where traditional data analytic methods might still suffice, but data have to be handled via external databases. Data preparation then requires experience with software and programming that is not part of many medical or social science curricula. For comparison, the low-resolution sensor readings in this study already added up to approximately 7.5 million data points.
- Sensor data collection should be applicable to many different survey frameworks without the need for proprietary software. This perspective does not comprise the technical implementation regarding the application programming interface, which will be the same for all frameworks; rather, it is concerned with the space required to save the sensor streams. For example, in this study, we used the method of temporarily collecting the data per survey page in a long string of text in the web browser's memory and then saved it to the survey's database. Some survey frameworks might be limited even further when storing large chunks of text to a study variable.
- Sensor data collection should not interfere with the main questionnaire or device use. Collecting large chunks of data in the web browser's memory may slow down the experience of using the questionnaire when older smartphones are used. As the training of the sensor models still relies on the self-reported outcomes, unnecessary dropout owing to inconvenience when completing the questionnaire should be avoided. In settings where dedicated apps are used, the sampling frequency could easily be increased.

Regarding the overall variation in the performance of the individual models, retrospectively, it would have been preferable to collect data on the specific devices that might have indicated some issues related to a family of devices or an operating system. Therefore, we recommend that future studies should at least record basic parameters such as screen resolution and the web browser that was used to rule out some device-specific issues.

Sensor data are also subject to the same *black box* problem that is immanent in any study in which experimenters and participants rarely meet each other. We do not know exactly what the participants were doing or experiencing while filling out the questionnaire. Consequently, a very poor-performing model might simply originate from a participant's irregular behavior when filling out the questionnaire. However, contrary to the same issues with self-report data, combining sensors with pretrained classification algorithms might be fruitful for determining participants' activity levels for each measurement and informing the models beyond the aggregated features [25,42].

Finally, although we were able to validate our method for a variety of psychological constructs, it is unclear whether this transitions to all possible time frames and topics. As demonstrated in our analyses, models will not learn when there are not sufficient data. This might limit the implementation of this personalized approach to studies that comprise contexts with at least 10 to 20 measurement points. Our study was also unable to answer questions about the underlying processes that determine a good model or features responsible for being a good marker regarding the outcomes we investigated. Consequently, the process between features and mental state itself remains largely unexplored. As we demonstrated in this study that data can be acquired almost effortlessly, we encourage future studies to examine and define the connections between the mind and the behavior we can record concurrently with the self-report.

Conclusions

In our study, we demonstrated that a few unobtrusively collected movement sensor data are a capable foundation to train models that are able to infer a range of psychological constructs with sizable effects [67]. This study contributes to previous research by inspecting differences between models trained on between-person variance and personalized models. Our study showed that validating such models by inferring participants' states in an independent data set can prove to be a fruitful approach. This applies even when relying on a privacy-friendly small amount of data that were only collected during measurement occasions of the self-report data collection. Finally, we expanded on previous work on combining personalized models with sensor data in the EMA context using a large and diverse sample. Despite basing our models only on movement-related sensors, we were able to replicate a comparable degree of accuracy. This worked best for outcomes that have already been demonstrated to work well in the

literature (ie, mood) but also showed its potential regarding other outcomes (eg, life engagement).

Although sensor readings of physiological properties have a strong root in the history of medicine and psychology, the exploration of the relationships between everyday fluctuations in psychometric values and smartphone sensors is still in its infancy. Much more research on this topic will be required before it can become a valuable tool in mental health. We contributed to this goal by demonstrating that the accessible method of obtaining sparse movement sensor data that we outlined can be temporally interlinked with EMA studies without sacrificing accuracy compared with studies with specialized hardware, software, or a much larger data set. In the future, such methods might be able to provide valuable insights into the mental well-being of the participants while reducing burden in research and clinical application.

Acknowledgments

The authors acknowledge the support from the Deutsche Forschungsgemeinschaft (German Research Foundation) and Saarland University within the Open Access Publishing funding program. They would like to thank MSc Matti Meyer for his assistance in survey preparation and collection of the ecological momentary assessment data.

Conflicts of Interest

None declared.

References

1. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
2. Papp LM, Kouros CD, Witt HK, Curtin JJ, Blumenstock SM, Barringer A. Real-time momentary mood as a predictor of college students' prescription drug misuse in daily life: direct links and the moderating role of background mental health. Exp Clin Psychopharmacol (forthcoming) 2021 Jun 10. [doi: 10.1037/pha0000500] [Medline: 34110882]
3. Schwarz N. Retrospective and concurrent self-reports: the rationale for real-time data capture. In: Stone A, Shiffman S, Atienza A, Nebeling L, editors. The science of real-time data capture: self-reports in health research. Oxford, UK: Oxford University Press; May 3, 2007:11-26.
4. Hufford MR. Special methodological challenges and opportunities in ecological momentary assessment. In: Stone A, Shiffman S, Atienza A, Nebeling L, editors. The science of real-time data capture: self-reports in health research. Oxford, UK: Oxford University Press; May 3, 2007:54-75.
5. Lingler JH, Schmidt KL, Gentry AL, Hu L, Terhorst LA. A new measure of research participant burden: brief report. J Empir Res Hum Res Ethics 2014 Oct;9(4):46-49 [FREE Full text] [doi: 10.1177/1556264614545037] [Medline: 26125079]
6. 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]
7. 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]
8. Fuller-Tyszkiewicz M, Skouteris H, Richardson B, Blore J, Holmes M, Mills J. Does the burden of the experience sampling method undermine data quality in state body image research? Body Image 2013 Sep;10(4):607-613. [doi: 10.1016/j.bodyim.2013.06.003] [Medline: 23856302]
9. Eisele G, Vachon H, Lafit G, Kuppens P, Houben M, Myin-Germeys I, et al. The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. Assessment 2022 Mar;29(2):136-151. [doi: 10.1177/1073191120957102] [Medline: 32909448]
10. Vachon H, Viechtbauer W, Rintala A, Myin-Germeys I. Compliance and retention with the experience sampling method over the continuum of severe mental disorders: meta-analysis and recommendations. J Med Internet Res 2019 Dec 06;21(12):e14475 [FREE Full text] [doi: 10.2196/14475] [Medline: 31808748]
11. Ono M, Schneider S, Junghaenel DU, Stone AA. What affects the completion of ecological momentary assessments in chronic pain research? An individual patient data meta-analysis. J Med Internet Res 2019 Feb 05;21(2):e11398 [FREE Full text] [doi: 10.2196/11398] [Medline: 30720437]
12. Jones A, Remmerswaal D, Verveer I, Robinson E, Franken IH, Wen CK, et al. Compliance with ecological momentary assessment protocols in substance users: a meta-analysis. Addiction 2019 Apr;114(4):609-619 [FREE Full text] [doi: 10.1111/add.14503] [Medline: 30461120]
13. Williams MT, Lewthwaite H, Fraysse F, Gajewska A, Ignatavicius J, Ferrar K. Compliance with mobile ecological momentary assessment of self-reported health-related behaviors and psychological constructs in adults: systematic review and meta-analysis. J Med Internet Res 2021 Mar 03;23(3):e17023 [FREE Full text] [doi: 10.2196/17023] [Medline: 33656451]
14. Shrout PE, Lane SP. Psychometrics. In: Mehl MR, Conner TS, editors. Handbook of research methods for studying daily life. New York, NY, USA: The Guilford Press; 2012:302-320.

15. Miller G. The smartphone psychology manifesto. Perspect Psychol Sci 2012 May;7(3):221-237. [doi: 10.1177/1745691612441215] [Medline: 26168460]
16. Trull TJ, Ebner-Priemer U. The role of ambulatory assessment in psychological science. Curr Dir Psychol Sci 2014 Dec;23(6):466-470 [FREE Full text] [doi: 10.1177/0963721414550706] [Medline: 25530686]
17. Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 2017 May 08;13:23-47 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
18. Wang H, Marsella S. Assessing personality through objective behavioral sensing. In: Proceedings of the 7th International Conference on Affective Computing and Intelligent Interaction. 2017 Presented at: ACII '17; October 23-26, 2017; San Antonio, TX, USA p. 131-137. [doi: 10.1109/acii.2017.8273590]
19. Wang W, Harari GM, Wang R, Müller SR, Mirjafari S, Masaba K, et al. Sensing behavioral change over time: using within-person variability features from mobile sensing to predict personality traits. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Sep 18;2(3):1-21. [doi: 10.1145/3264951]
20. Stachl C, Au Q, Schoedel R, Gosling SD, Harari GM, Buschek D, et al. Predicting personality from patterns of behavior collected with smartphones. Proc Natl Acad Sci U S A 2020 Jul 28;117(30):17680-17687 [FREE Full text] [doi: 10.1073/pnas.1920484117] [Medline: 32665436]
21. Rüegger D, Stieger M, Nißen M, Allemand M, Fleisch E, Kowatsch T. How are personality states associated with smartphone data? Eur J Pers 2020 Sep 01;34(5):687-713. [doi: 10.1002/per.2309]
22. Yim SJ, Lui LM, Lee Y, Rosenblat JD, Ragguett RM, Park C, et al. The utility of smartphone-based, ecological momentary assessment for depressive symptoms. J Affect Disord 2020 Sep 01;274:602-609. [doi: 10.1016/j.jad.2020.05.116] [Medline: 32663993]
23. Saeb S, Lonini L, Jayaraman A, Mohr DC, Kording KP. The need to approximate the use-case in clinical machine learning. Gigascience 2017 May 01;6(5):1-9 [FREE Full text] [doi: 10.1093/gigascience/gix019] [Medline: 28327985]
24. Asselbergs J, Ruwaard J, Ejdys M, Schrader N, Sijbrandij M, Riper H. Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: an explorative study. J Med Internet Res 2016 Mar 29;18(3):e72 [FREE Full text] [doi: 10.2196/jmir.5505] [Medline: 27025287]
25. Jacobson NC, Chung YJ. Passive sensing of prediction of moment-to-moment depressed mood among undergraduates with clinical levels of depression sample using smartphones. Sensors (Basel) 2020 Jun 24;20(12):3572 [FREE Full text] [doi: 10.3390/s20123572] [Medline: 32599801]
26. Sultana M, Al-Jefri M, Lee J. Using machine learning and smartphone and smartwatch data to detect emotional states and transitions: exploratory study. JMIR Mhealth Uhealth 2020 Sep 29;8(9):e17818 [FREE Full text] [doi: 10.2196/17818] [Medline: 32990638]
27. LiKamWa R, Liu Y, Lane ND, Zhong L. MoodScope: building a mood sensor from smartphone usage patterns. In: Proceeding of the 11th annual international conference on Mobile systems, applications, and services. 2013 Jun Presented at: MobiSys '13; June 25-28, 2013; Taipei, Taiwan p. 389-402. [doi: 10.1145/2462456.2464449]
28. Bähr S, Haas GC, Keusch F, Kreuter F, Trappmann M. Missing data and other measurement quality issues in mobile geolocation sensor data. Soc Sci Comput Rev 2020 Aug 06;40(1):212-235. [doi: 10.1177/0894439320944118]
29. Saeb S, Lattie EG, Kording KP, Mohr DC. Mobile phone detection of semantic location and its relationship to depression and anxiety. JMIR Mhealth Uhealth 2017 Aug 10;5(8):e112 [FREE Full text] [doi: 10.2196/mhealth.7297] [Medline: 28798010]
30. Harari GM, Lane ND, Wang R, Crosier BS, Campbell AT, Gosling SD. Using smartphones to collect behavioral data in psychological science: opportunities, practical considerations, and challenges. Perspect Psychol Sci 2016 Nov;11(6):838-854 [FREE Full text] [doi: 10.1177/1745691616650285] [Medline: 27899727]
31. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. J Med Internet Res 2015 Jul 15;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
32. de Montjoye YA, Hidalgo CA, Verleysen M, Blondel VD. Unique in the Crowd: the privacy bounds of human mobility. Sci Rep 2013;3:1376 [FREE Full text] [doi: 10.1038/srep01376] [Medline: 23524645]
33. Williams PA. When trust defies common security sense. Health Informatics J 2008 Sep;14(3):211-221 [FREE Full text] [doi: 10.1177/1081180X08092831] [Medline: 18775827]
34. Zimmer M, Kumar P, Vitak J, Liao Y, Chamberlain Kritikos K. 'There's nothing really they can do with this information': unpacking how users manage privacy boundaries for personal fitness information. Inf Commun Soc 2020;23(7):1020-1037. [doi: 10.1080/1369118x.2018.1543442]
35. Petronio S. Boundaries of privacy: dialectics of disclosure. Albany, NY, USA: SUNY Press; 2002.
36. Jacobson NC, Bentley KH, Walton A, Wang SB, Fortgang RG, Millner AJ, et al. Ethical dilemmas posed by mobile health and machine learning in psychiatry research. Bull World Health Organ 2020 Apr 01;98(4):270-276 [FREE Full text] [doi: 10.2471/BLT.19.237107] [Medline: 32284651]
37. Carter A, Liddle J, Hall W, Chenery H. Mobile phones in research and treatment: ethical guidelines and future directions. JMIR Mhealth Uhealth 2015 Oct 16;3(4):e95 [FREE Full text] [doi: 10.2196/mhealth.4538] [Medline: 26474545]

38. Lane ND, Miluzzo E, Lu H, Peebles D, Choudhury T, Campbell AT. A survey of mobile phone sensing. IEEE Commun Mag 2010 Sep;48(9):140-150. [doi: 10.1109/mcom.2010.5560598]
39. LeMoyne R, Mastroianni T, Grundfest W. Wireless accelerometer configuration for monitoring Parkinson's disease hand tremor. Adv Parkinsons Dis 2013 May;2(2):62-67. [doi: 10.4236/apd.2013.22012]
40. Olsen AF, Torresen J. Smartphone accelerometer data used for detecting human emotions. In: Proceedings of the 3rd International Conference on Systems and Informatics. 2016 Presented at: ICSAI '16; November 19-21, 2016; Shanghai, China p. 410-415. [doi: 10.1109/icsai.2016.7810990]
41. Piskioulis O, Tzafilkou K, Economides A. Emotion detection through smartphone's accelerometer and gyroscope sensors. In: Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization. 2021 Jun Presented at: UMAP '21; June 21-25, 2021; Utrecht, Netherlands p. 130-137. [doi: 10.1145/3450613.3456822]
42. Kern C, Höhne JK, Schlosser S, Revilla M. Completion conditions and response behavior in smartphone surveys: a prediction approach using acceleration data. Soc Sci Comput Rev 2020 Dec 04;39(6):1253-1271. [doi: 10.1177/0894439320971233]
43. Riskind J. They stoop to conquer: guiding and self-regulatory functions of physical posture after success and failure. J Pers Soc Psychol 1984;47(3):479-493. [doi: 10.1037/0022-3514.47.3.479]
44. Kuhlmann T, Garaizar P, Reips UD. Smartphone sensor accuracy varies from device to device in mobile research: the case of spatial orientation. Behav Res Methods 2021 Feb;53(1):22-33 [FREE Full text] [doi: 10.3758/s13428-020-01404-5] [Medline: 32472500]
45. Kröger JL, Raschke P, Bhuiyan TR. Privacy implications of accelerometer data: a review of possible inferences. In: Proceedings of the 3rd International Conference on Cryptography, Security and Privacy. 2019 Jan Presented at: ICCSP '19; January 19-21, 2019; Kuala Lumpur, Malaysia p. 81-87.
46. Kreuter F. Improving surveys with paradata: Introduction. In: Kreuter F, editor. Improving surveys with Paradata: analytic uses of process information. Hoboken, NJ, USA: John Wiley & Sons; 2013:1-9.
47. Groves RM, Lyberg L. Total survey error: past, present, and future. Public Opin Q 2011 Feb 09;74(5):849-879. [doi: 10.1093/poq/nfq065]
48. Lenzner T, Kaczmirek L, Lenzner A. Cognitive burden of survey questions and response times: a psycholinguistic experiment. Appl Cognit Psychol 2009 Aug 20;24(7):1003-1020. [doi: 10.1002/acp.1602]
49. Reis D, Prestele E. The role of trait and state perfectionism in psychological detachment from daily job demands. Stress Health 2020 Apr;36(2):228-245. [doi: 10.1002/smi.2901] [Medline: 31612593]
50. McNair DM, Lorr M, Droppleman LF. Revised manual for the Profile of Mood States. San Diego, CA, USA: Educational and Industrial Testing Services; 1992.
51. von Steyer R, Schwenkmezger P, Notz P, Eid M. Der mehrdimensionale Befindlichkeitsfra-gebogen (MDBF). Göttingen, Germany: Hogrefe Publishing Group; 1997.
52. Schaufeli WB, Shimazu A, Hakanen J, Salanova M, De Witte H. An ultra-short measure for work engagement. Eur J Psychol Assess 2019 Jul;35(4):577-591. [doi: 10.1027/1015-5759/a000430]
53. Schaufeli WB. General engagement: conceptualization and measurement with the Utrecht General Engagement Scale (UGES). J Well-being Assess 2017 Aug 17;1(1-3):9-24. [doi: 10.1007/s41543-017-0001-x]
54. Schaufeli W, Bakker A. Utrecht Work Engagement Scale: preliminary manual. Occupational Health Psychology Unit, Utrecht University. 2004 Dec. URL: https://www.wilmarschaufeli.nl/publications/Schaufeli/Test%20Manuals/Test_manual_UWES_English.pdf [accessed 2018-05-01]
55. Schaufeli WB, Bakker A, Salanova M. The measurement of work engagement with a short questionnaire: a cross-national study. Educ Psychol Meas 2006 Aug 1;66(4):701-716. [doi: 10.1177/0013164405282471]
56. Flaxman PE, Stride CB, Söderberg M, Lloyd J, Guenole N, Bond FW. Relationships between two dimensions of employee perfectionism, postwork cognitive processing, and work day functioning. Eur J Work Organ Psychol 2018;27(1):56-69. [doi: 10.1080/1359432x.2017.1391792]
57. Flaxman PE, Ménard J, Bond FW, Kinman G. Academics' experiences of a respite from work: effects of self-critical perfectionism and perseverative cognition on postrespite well-being. J Appl Psychol 2012 Jul;97(4):854-865. [doi: 10.1037/a0028055] [Medline: 22545621]
58. Görtelmeyer R. Schlaffragebogen A und B - Revidierte Fassung (SF-A/R und SF-B/R). Göttingen, Germany: Hogrefe Publishing Group; 2011.
59. Hoogendoorn M, Funk B. Machine learning for the quantified self. Cham, Switzerland: Springer International Publishing; 2018.
60. Breiman LR. Random forests. Mach Learn 2001;45(1):5-32. [doi: 10.1023/A:1010933404324]
61. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33(1):1-22 [FREE Full text] [Medline: 20808728]
62. Scrucca L, Fop M, Murphy TB, Raftery AE. mclust 5: clustering, classification and density estimation using gaussian finite mixture models. R J 2016 Aug;8(1):289-317 [FREE Full text] [Medline: 27818791]
63. Holm S. A simple sequentially rejective multiple test procedure. Scand J Statist 1979;6(2):6. [doi: 10.2307/4615733]
64. Wilcox RR. The percentage bend correlation coefficient. Psychometrika 1994 Dec;59(4):601-616. [doi: 10.1007/bf02294395]

65. Stisen A, Blunck H, Bhattacharya S, Prentow TS, Kjærgaard MB, Dey A, et al. Smart devices are different: assessing and MitigatingMobile sensing heterogeneities for activity recognition. In: Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems. 2015 Nov Presented at: SenSys '15; December 26-28, 2021; Bali Island, Indonesia p. 127-140. [doi: 10.1145/2809695.2809718]
66. Graham JW, Taylor BJ, Olchowski AE, Cumsille PE. Planned missing data designs in psychological research. Psychol Methods 2006 Dec;11(4):323-343. [doi: 10.1037/1082-989X.11.4.323] [Medline: 17154750]
67. Funder DC, Ozer DJ. Evaluating effect size in psychological research: sense and nonsense. Adv Methods Pract Psychol Sci 2019 May 08;2(2):156-168. [doi: 10.1177/2515245919847202]

Abbreviations

BIC: Bayesian information criterion
EMA: ecological momentary assessment
GLM: general linear model
LPA: latent profile analysis

Edited by R Kukafka; submitted 04.10.21; peer-reviewed by E Chan, G Coombs; comments to author 06.01.22; revised version received 02.03.22; accepted 13.03.22; published 28.04.22.

Please cite as:
Hart A, Reis D, Prestele E, Jacobson NC
Using Smartphone Sensor Paradata and Personalized Machine Learning Models to Infer Participants' Well-being: Ecological Momentary Assessment
J Med Internet Res 2022;24(4):e34015
URL: https://www.jmir.org/2022/4/e34015
doi:10.2196/34015
PMID:35482397

Original Paper

The Moderating Effects of Disability on Mobile Internet Use Among Older Adults: Population-Based Cross-sectional Study

Eunjin Yang[1], PhD; Kyung Hee Lee[1], PhD

Mo-Im Kim Nursing Research Institute, Yonsei University College of Nursing, Seoul, Republic of Korea

Corresponding Author:
Kyung Hee Lee, PhD
Mo-Im Kim Nursing Research Institute
Yonsei University College of Nursing
50-1 Yonsei-ro
Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 2 2228 3321
Email: kyungheelee@yuhs.ac

Abstract

Background: The preferred devices to access the internet are changing from personal computers to mobile devices, and the number of older adults with or without disabilities is rapidly increasing in an aging society. However, little is known about the moderating effects of disability on mobile internet use among older adults.

Objective: This study aimed to examine the levels of mobile internet use and factors associated with this use among older adults according to their disabilities. In addition, moderating effects of disability on mobile internet use were investigated.

Methods: This study consisted of a secondary data analysis using the 2020 Digital Divide Survey conducted in South Korea. The single inclusion criterion was participants being aged 55 years or older; accordingly, 2243 people without disabilities and 1386 people with disabilities were included in the study. Multiple regression analyses considering complex sample designs were conducted to identify mobile internet use factors and to test the moderating effects of disability on mobile internet use.

Results: Older adults with disabilities used mobile internet less than older adults without disabilities. However, disability status had moderating effects on the relationships between mobile internet use and (1) operational skills regarding mobile devices (B=0.31, *P*=.004), (2) internet use skills (B=1.46, *P*<.001), (3) motivation to use digital devices (B=0.46, *P*=.01), and (4) attitude toward new technology (B=0.50, *P*=.002). The results revealed that these positive relationships were stronger among older adults with disabilities than among adults without disabilities.

Conclusions: Although older adults and people with disabilities are considered vulnerable populations regarding technology adoption, disability creates a stronger association between several determinants and actual mobile internet use. Therefore, policy makers and practitioners should pay attention to older adults with disabilities to deliver appropriate information-literacy education. Older adults with disabilities could be the primary beneficiaries of mobile services and new technology.

(J Med Internet Res 2022;24(4):e37127) doi:10.2196/37127

KEYWORDS

older adults; people with disabilities; digital divide; mobile phone use

Introduction

During the COVID-19 pandemic, most offline activities and services were transferred online according to the regulations and guidelines for infectious disease response [1,2]. In particular, South Korea implemented principles and policies based on information and communications technology (ICT) to respond to infectious diseases [3]; because of this, digital transformation is accelerating in the country. Use of ICT is becoming more crucial in this highly connected information society. In addition to these situations, the digital divide—defined as the gap between ICT users and nonusers in terms of access to or use of ICT based on their economic, regional, physical, and social factors—is becoming serious [4]. Discourse around the digital divide in society has changed the issue from material access to that of effective use of technology [5,6]. South Korea has already achieved a highly connected network society through

high internet penetration; therefore, digital divide issues, such as skills and benefits from information use, are emerging beyond material access issues [7].

Among various populations, older adults are less likely to adopt, diffuse, or use ICT [8]. Therefore, with the prevalence of non–face-to-face activities and social services, older adults who do not use the internet are more likely to be disadvantaged [9]. Older adults who are nonusers of the internet face double the burden, that of digital and social exclusion. They can become easily isolated, feel lonely, and face many difficulties in using the digital technologies that pervade daily life compared to internet users [10]. In addition, older adults generally have poor health status and are prone to disabilities, such as brain lesions, visual impairment, and hearing impairment. These disabilities may impact their technology adoption and mobile internet usage [11].

In terms of populations with disabilities, there are conflicting opinions regarding technology use. Previous studies reported that people who have disabilities experience the digital divide more frequently [12]. On the other hand, there exists a differing view that people with disabilities want to use online information and digital technology to make their voices heard and participate in the online community [13]. Once people with disabilities can connect to the internet and mobile accessibility issues are resolved, the online world can potentially become a physical barrier–free environment for them [14]. Therefore, digital technology, including mobile internet use, is considered both a challenge and an opportunity for people with disabilities [15].

Mobile internet use among older adults and people with disabilities has risen over time, and these trends increased dramatically during the COVID-19 pandemic [16,17]. In addition, the main devices used to access and use the internet are changing from personal computers to smart mobile devices. Approximately 91.5% of the total population aged 3 years and above use mobile internet, and 93.1% of those aged 6 years and above possess smartphones in South Korea [18]. This means that everyone who owns a smart mobile device can use ICT services, regardless of the time and place. Mobile internet use significantly affects people's chances to obtain information and build social capital. Therefore, there is a gap in this regard between mobile internet users and nonusers [19].

Previous studies revealed many factors associated with mobile internet use in various dimensions, including social, cultural, personal, material, and motivational aspects [20]. According to van Dijk's digital divide theory, motivation, materials, and skills are continuously and recursively associated with digital information usage and participation in society [6]. Moreover, skills, attitudes, motivation, and internet efficacy are closely linked to the use of internet information [20]. Many studies have investigated the relationship between these associated factors and internet use among people with disabilities and older adults [16,20]. Some studies have reported that people with disabilities are less likely to use the internet than those without disabilities [21,22], and they have compared the digital divide among people with disabilities and those without [15,21]. One study reported a moderating effect of disability on smartphone online activities; however, it examined all age groups, and the results showed that the moderating effects of disability existed in the relationships between smartphone online activities and the attitude and social support factors [16]. However, older adults may show different patterns of mobile internet usage compared to other age groups, and the relationship between mobile internet use and its associated factors could differ according to their health and disability status [23]. To fill in this research gap, research regarding older adults' mobile internet use according to their disabilities is needed. Even if older adults were less likely to use the internet than younger adults, many of their health care needs, such as reliable health information, could be addressed by information disseminated by the internet. Therefore, before preparing strategies to facilitate older adults' mobile use, it may be helpful to understand the relationships between several factors associated with technology and the adults' actual mobile internet use. Accordingly, this study examined the levels of mobile internet use and its related factors among older adults based on their disabilities. In addition, it investigated the moderating effects of disability on mobile internet use.

Methods

Data and Study Participants

We used data from the 2020 Digital Divide Survey, which was conducted in South Korea from September to December 2020 through face-to-face interviews using structured questions. This survey was managed by the Ministry of Science and ICT and the National Information Society Agency; it has been implemented annually since 2002. The survey examined the effectiveness of policy to bridge the digital divide between the general population and vulnerable populations, such as people with disabilities, older adults, and North Korean defectors, among others. These data were collected using a complex survey design. Older adults were recruited through stratified proportional sampling; people with disabilities were recruited through proportional sampling by group characteristics (eg, age, gender, type of disability, and geographic area). In this study, people with disabilities were defined as those who were registered as such according to the national registration process after medical diagnosis. In terms of disability types, physical, brain lesion, visual, hearing, and speech disabilities were included in this survey among the 15 types of disabilities defined in the Enforcement Decree of the Act on Welfare of Persons with Disabilities of South Korea [24]. The participants' only inclusion criterion in this study was being aged 55 years or above; therefore, of the 2200 people with disabilities, 871 (39.59%) who were aged below 55 years were excluded, leaving 1329 people. Although the survey for older adults did not specifically target people with disabilities, a question assessing the presence of disability was included in the questionnaire. However, details regarding the types of disability were not assessed. Therefore, among 2300 older adults without disabilities, 57 people who answered as having a registered disability, based on legal standards, were reclassified as people with disabilities. In the end, 2243 people without disabilities and 1386 people with disabilities were included in this study.

Variables

Dependent Variable

The level of mobile internet use was calculated based on three points:

1. Whether people used the internet recently in the past month or not and days of use (1 item).
2. Whether people used a variety of internet services (13 items, including general search, email, online content services, social networks and cloud services, and information searches regarding daily life).
3. Whether people used advanced internet services (12 items, including information generation and sharing, online networking, and participation in social activities and the economy).

Items for points 2 and 3 were measured on a 4-point scale. The level of mobile internet use was calculated by considering item weights (ie, 0.4, 0.4, and 0.2, respectively, for each of the three points). The score was calculated based on the guidance provided by the National Information Society Agency, and the total score ranged from 0 to 100.

Independent Variables

Independent variables included operational skills regarding mobile devices, internet use skills, motivation to use digital devices, and attitude toward new technology. Operational skills regarding mobile devices were defined as skills for operating mobile device hardware and software [25,26]; these were investigated using seven items on a 4-point scale, ranging from 1 (strongly disagree) to 4 (strongly agree). The seven items were as follows: (1) configuring mobile devices; (2) accessing Wi-Fi networks; (3) transferring files from mobile devices to personal computers; (4) transferring files or photos from the participants' mobile devices to those of other people; (5) installing, deleting, and updating mobile apps; (6) dealing with malware in mobile devices; and (7) creating documents on mobile devices. The possible score ranged from 7 to 28, and Cronbach α was .94 in this study.

Skills for internet use, also called strategic internet skills, are comprised of the capacity to access and manage information using digital devices to reach particular goals [25,26]. They were measured using four items on a 4-point scale, ranging from 1 (strongly disagree) to 4 (strongly agree). The items were as follows:

1. I can connect and communicate with others through the internet and cooperate with them for problem solving and work.
2. I can exchange opinions about political and social issues and problems through the internet and participate in various activities, such as discussion, donation, and solving of public problems.
3. I can protect myself and others from risk factors associated with internet use, such as leaking of private information.
4. I can understand others' opinions, accept different views, and use the internet responsibly, while not accessing illegal media or infringing on other people's rights.

The total score ranged from 4 to 16, and the Cronbach α was .88 in this study.

The scale for motivation to use digital devices included five items on a 4-point scale, ranging from 1 (strongly disagree) to 4 (strongly agree). Items were related to participants' (1) eagerness to obtain information using digital devices, (2) wish to become acquainted with many persons through digital devices, (3) wish to be entertained through digital devices, (4) need for self-development, and (5) need to express one's opinion through digital devices. The total score ranged from 5 to 20, and the Cronbach α was .88 in this study.

The scale for attitude toward new technology comprised six items on a 4-point scale, ranging from 1 (strongly disagree) to 4 (strongly agree). The items were as follows:

1. I tend to adapt well to new technology and products.
2. I am confident in using new technology and products by myself.
3. When I use new technology and products, I do well as compared with others.
4. I think digital technology is essential to continue economic activities.
5. I try to learn new technology actively.
6. I think I am a lifetime learner and enjoy the classes needed to learn new technology.

The total score ranged from 6 to 24, and the Cronbach α was .88 in this study.

Covariates

The demographic characteristics of age (continuous), gender (female vs male), education (below vs above high school), living arrangements (living alone vs with others), household income (below vs above ₩4,000,000/month; a currency exchange rate of ₩1=US $0.00082 is applicable), and living areas (urban vs rural) were selected as covariates [12,20].

Statistical Analysis

Data analysis was carried out using SPSS Statistics for Windows (version 26.0; IBM Corp). Data were analyzed using complex sampling analysis considering sample weight. Comparison of demographic characteristics and variables between older adults with disabilities and those without disabilities was conducted using the *t* test (2-tailed) and Rao - Scott chi - square tests. A multiple regression analysis was applied to examine the association between independent variables and mobile internet use. To explore the moderating effect of disability, first, the covariates, independent variables (ie, operational skills, internet use skills, motivation, and attitude), and moderator (ie, disability) were entered into the model as a block. Then, four interaction terms (ie, operational skills × disability, internet use skills × disability, motivation × disability, and attitude × disability) were entered into each model; therefore, four separate models were created. Statistical significance was set at the .05 level in all analyses. The slope of the moderation effect of disability was presented using the EasyFlow Statistics macro in Excel [27].

Ethical Considerations

The Korean Statistics Act guaranteed information protection for all participants, and participants were notified about this before the survey. This study was a secondary analysis study using publicly available data; therefore, the Institutional Review Board (IRB) approved this study as an exempt study (IRB No. 4-2021-1743). All study processes were performed according to the relevant guidelines and regulations.

Results

Characteristics and Variables of Groups With Disabilities and Those Without Disabilities

Participants' general characteristics and independent variables based on disability status are summarized in Table 1. In terms of disability types among the group of persons with disabilities, over half of them (61.85%) were physical disabilities. Most demographic characteristics differed between the two groups. The mean age of older adults without disabilities (mean 66.54, SD 0.18 years) was higher than that of older adults with disabilities (mean 62.60, SD 0.14 years), and there were more males in the group of older adults with disabilities. In terms of education, a higher proportion of older adults without disabilities had a high educational level (81.43%; ie, high school education and above) compared with those with disabilities (50.54%). As for living arrangements, 16.86% of those with disabilities lived alone, while 8.58% of those without disabilities lived alone.

Regarding independent variables, operational skills regarding mobile devices ($P<.001$), internet use skills ($P<.001$), motivation to use digital devices ($P=.04$), and attitude toward new technology ($P<.001$) were higher among older adults without disabilities compared to those with disabilities. Mobile internet use was also higher in the group without disabilities than in the group with disabilities ($P<.001$).

Table 1. Demographic characteristics and variables of the participants (N=3629).

Characteristic or variable	Participants with disabilities (n=1386)	Participants without disabilities (n=2243)	*t* test[a] (*df*=4498)	*F* test[a,b] (*df*=1, 3628)	*P* value[a]
Type of disability, n (%)[c]					
Physical disability	862 (61.85)	N/A[d]	N/A	N/A	N/A
Brain lesion	151 (10.78)	N/A	N/A	N/A	N/A
Visual impairment	153 (11.36)	N/A	N/A	N/A	N/A
Hearing impairment	133 (9.49)	N/A	N/A	N/A	N/A
Speech disability	30 (2.35)	N/A	N/A	N/A	N/A
Not reported	57 (4.17)	N/A	N/A	N/A	N/A
Age (years), mean (SD)	62.60 (0.14)	66.54 (0.18)	17.196	N/A	<.001
Gender, n (%)[c]					
Female	494 (35.27)	1199 (53.43)	N/A	95.07	<.001
Male	892 (64.73)	1044 (46.57)			
Educational level, n (%)[c]					
Below high school	724 (49.46)	414 (18.57)	N/A	333.32	<.001
High school and above	662 (50.54)	1829 (81.43)			
Living arrangements, n (%)[c]					
Living alone	231 (16.86)	192 (8.58)	N/A	50.34	<.001
Living with others	1155 (83.14)	2051 (91.42)			
Household income (₩[e]/month), n (%)[c]					
<4,000,000	1269 (90.84)	1263 (56.35)	N/A	326.85	<.001
≥4,000,000	117 (9.16)	980 (43.65)			
Living area, n (%)[c]					
Urban	1255 (92.27)	2083 (92.85)	N/A	0.45	.50
Rural	131 (7.73)	160 (7.15)			
Operational skills regarding mobile devices[f], mean (SE)	14.37 (0.18)	15.56 (0.12)	5.44	N/A	<.001
Internet use skills[g], mean (SE)	7.02 (0.08)	7.76 (0.06)	7.01	N/A	<.001
Motivation to use digital devices[h], mean (SE)	12.05 (0.11)	12.33 (0.07)	2.05	N/A	.04
Attitude toward new technology[i], mean (SE)	12.73 (0.11)	13.56 (0.09)	5.78	N/A	<.001
Mobile internet use[j], mean (SE)	40.05 (0.93)	44.08 (0.61)	3.64	N/A	<.001

[a]The *t* test (2-tailed), *F* test, and *P* values for a group are reported in the top row of that group.

[b]*P* values related to values in this column were determined using the Rao-Scott chi-square test.

[c]Percentages are weighted.

[d]N/A: not applicable; this measure did not apply to this group.

[e]A currency exchange rate of ₩1=US $0.00082 is applicable.

[f]Scores for operational skills ranged from 7 to 28.

[g]Scores for internet use skills ranged from 4 to 16.

[h]Motivation scores ranged from 5 to 20.

[i]Attitude scores ranged from 6 to 24.

[j]Mobile internet use scores ranged from 0 to 100.

Factors Associated With Mobile Internet Use

Multiple regression analysis was performed to examine the factors associated with mobile internet use among older adults. As Table 2 shows, age, household income, living areas, operational skills regarding mobile devices, internet use skills, motivation to use digital devices, and attitude toward new technology were positively related to mobile internet use. Disability status was not significantly associated with mobile internet use in model 1. The R^2 value of model 1 is 0.705.

Table 2. Factors associated with mobile internet use (N=3629).

Factor	Model 1 values		
	B (SE)	*t* test[a] (*df*=4498)	*P* value
Constant	12.28 (4.16)	2.95	.003
Age	–0.58 (0.05)	–12.46	<.001
Gender (reference: female)	–0.83 (0.61)	–1.36	.17
Educational level (reference: below high school)	0.83 (0.81)	1.03	.30
Living arrangements (reference: living alone)	0.68 (0.91)	0.75	.46
Household income (reference: <₩4,000,000[b]/month)	–1.38 (0.68)	–2.04	.04
Living area (reference: rural)	–1.77 (0.90)	–1.97	.049
Disability (reference: no disability)	–1.31 (0.78)	–1.67	.10
Operational skills regarding mobile devices	2.49 (0.10)	25.60	<.001
Internet use skills	1.85 (0.17)	10.91	<.001
Motivation to use digital devices	0.89 (0.13)	6.62	<.001
Attitude toward new technology	0.56 (0.12)	4.88	<.001

[a]The *t* test was 2-tailed.

[b]A currency exchange rate of ₩1=US $0.00082 is applicable.

Moderating Effects of Disability on Mobile Internet Use

The regression model to examine the moderating effect of disability indicated that disability status enhanced the relationship between independent variables and mobile internet use. The positive relationship between operational skills regarding mobile devices and mobile internet use was stronger among older adults with disabilities than among those without disabilities (B=0.31, *P*=.004). These moderating effects were also found in the relationship between mobile internet use and internet use skills (B=1.46, *P*<.001), motivation to use digital devices (B=0.46, *P*=.01), and attitude toward new technology (B=0.50, *P*=.002; Table 3).

The simple slopes of these relationships are presented in Figures 1 to 4.

Table 3. Moderating effects of disability (N=3629).

Model Number[a]	Model details	B (SE)	*t* test[b] (*df*=4498)	*P* value	R^2 ($\triangle R^2$)
2	Operational skills regarding mobile devices × disability	0.31 (0.11)	2.89	.004	0.706 (0.001)
3	Internet use skills × disability	1.46 (0.22)	6.64	<.001	0.710 (0.005)
4	Motivation to use digital devices × disability	0.46 (0.18)	2.53	.01	0.706 (0.001)
5	Attitude toward new technology × disability	0.50 (0.17)	3.03	.002	0.706 (0.001)

[a]Models 2 to 5 include all variables of model 1 from Table 2 (ie, age, gender, educational level, living arrangements, household income, living areas, disability, operational skills regarding mobile devices, internet use skills, motivation to use digital devices, and attitude toward new technology).

[b]The *t* test was 2-tailed.

Figure 1. Simple slopes for the relationship between operational skills and mobile internet use. Respective scores are listed on the axes.

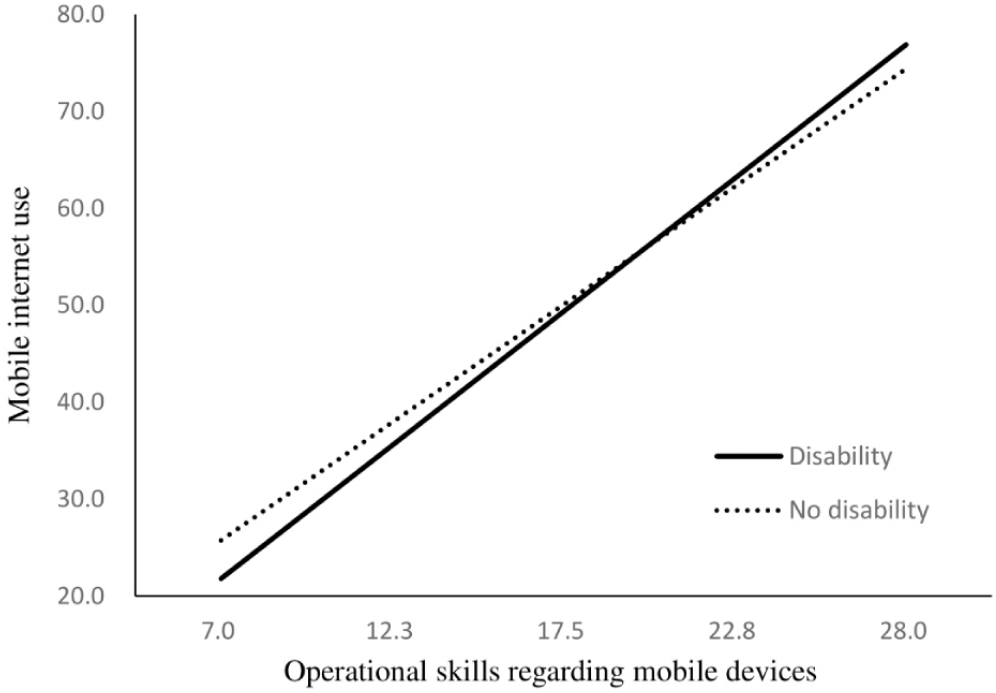


Figure 2. Simple slopes for the relationship between internet use skills and mobile internet use. Respective scores are listed on the axes.

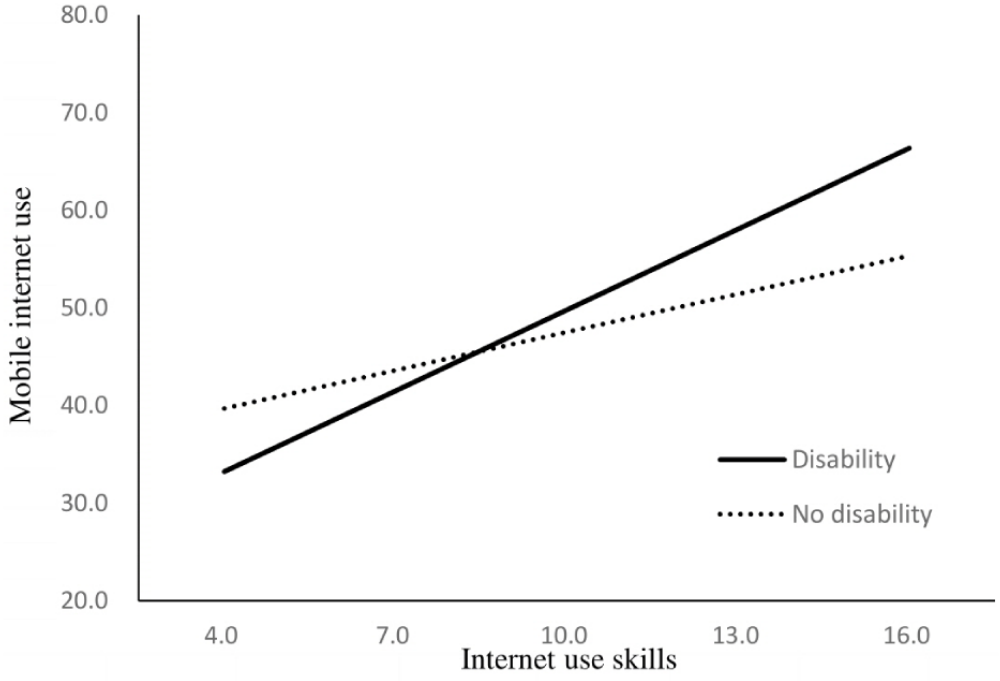


Figure 3. Simple slopes for the relationship between motivation and mobile internet use. Respective scores are listed on the axes.

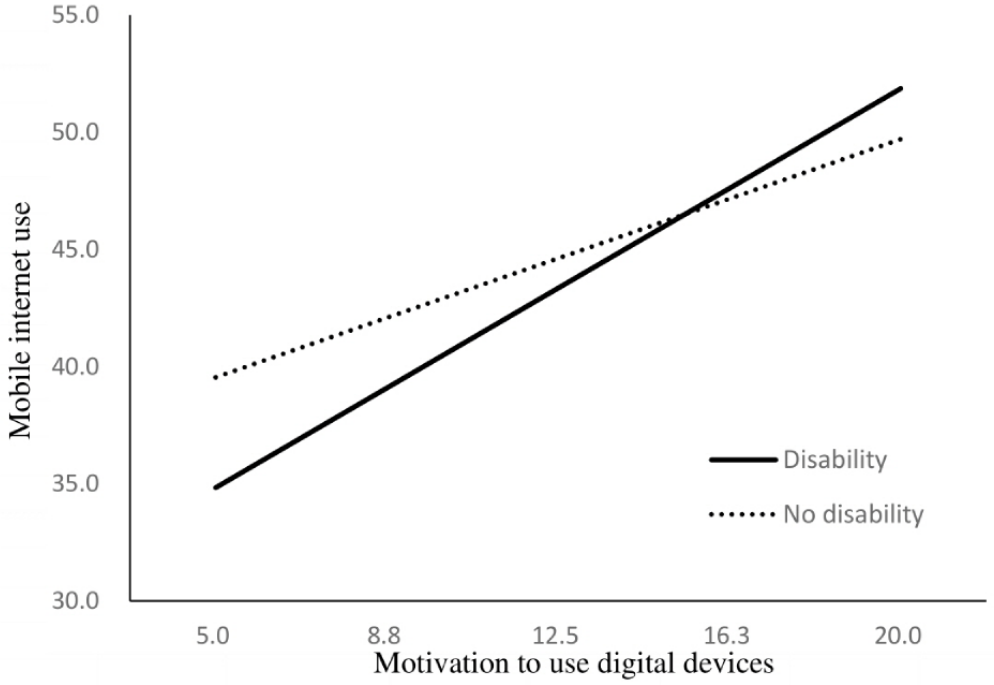
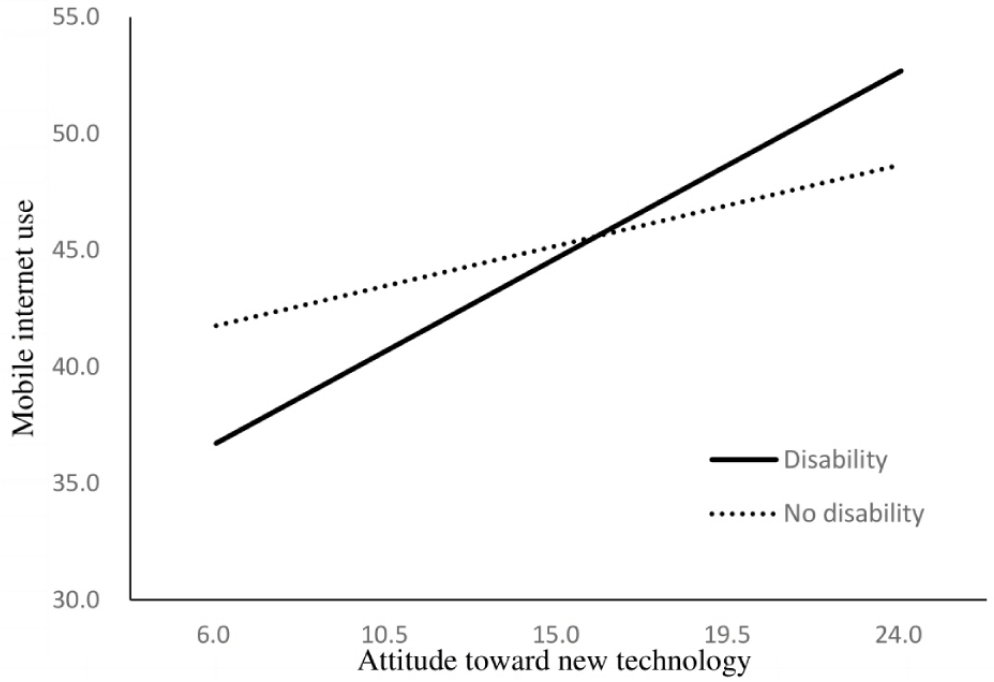


Figure 4. Simple slopes for the relationship between attitude and mobile internet use. Respective scores are listed on the axes.



Discussion

Principal Findings

This study examined the factors related to mobile internet use among older adults through a national representative sample in South Korea. The moderating effects of older adults' disability status on their mobile internet use was also investigated. The bivariate analysis showed gaps in both sociodemographic characteristics and level of mobile internet use among older adults with disabilities and those without disabilities. The level of internet use among older adults with disabilities was lower than that among their counterparts. In addition, operational skills regarding mobile devices, internet use skills, motivation to use digital devices, and attitude toward new technology were also low among the group with disabilities.

Previous research has frequently reported differences in the level of internet use between people with disabilities and the general population [15,16]. Our study confirmed these gaps by focusing on mobile devices and the older adult population. The reason for these results may be that sociodemographic factors contributing toward access to and use of the internet [28], such as education, income, and living arrangements, among older adults with disabilities were poorer than those among older

adults without disabilities. In addition, most mobile service and app designs and functions are less intuitive for older adults with disabilities, causing them to use the internet less than other populations [29,30]. The South Korean government releases web accessibility evaluation reports annually to ensure that technologically vulnerable populations, such as older adults and people with disabilities, can use the internet comfortably [31]. Although the government has developed standards for web accessibility and tries to bridge the digital divide, the standards and efforts to improve mobile web accessibility need to be reinforced beyond personal computer web accessibility.

In terms of factors related to mobile internet use, the sociodemographic characteristics of age, household income, and living area were statistically significant in this study, and our results were consistent with previous reports [20], except for disability. Unlike previous studies, which reported that disability is a determinant of internet use [12,15,21,28], this study found that disability is not statistically significant when it comes to using mobile internet. However, another study that focused on smartphones, and not personal computers, reported similar results as this study; it found that disability was not a significant factor related to mobile internet use when using the multiple regression model [16], which may be because of the types of devices targeted in each study. These results should be interpreted cautiously, as mobile devices are more appropriate than other devices for internet use by older adults, regardless of their disability status.

Therefore, mobile devices could provide many opportunities to older adults. Mobile device ownership, including smartphones, substantially increased among groups of people with disabilities and those without disabilities, and they became core tools for both groups. However, many information-literacy education courses for older adults provided at senior centers in South Korea focused on traditional personal computers for internet use [32]. From 2020 onward, the Korean government created digital competency centers as part of the digital inclusion policy, aiming for a rapid digital transformation. These are positive changes that can meet the need for mobile-based information-literacy education reflecting the trend of mobile device proliferation. Although these centers provide many courses [33], tailored interventions for older adults with disabilities and those without disabilities are still not enough and need to be expanded.

Other factors, such as operational skills, internet use skills, motivation to use digital devices, and attitude toward new technology, were also statistically significant in this study. All of these factors were reported as determinants of internet use [20], and our study confirmed that these relationships exist similarly among older adults with disabilities and those without disabilities. Many studies regarding the digital divide have verified that the predictors of internet use among people with disabilities are similar as those for people without disabilities [12,16]. This means that an integrated approach to increase mobile internet use and technology adoption among older adults, regardless of their disability status, is needed; this approach should consider common factors, such as skills, motivation, and attitude toward technology. Many countries, including Australia, Israel, New Zealand, Singapore, and the United Kingdom, have presented visions such as a digital readiness blueprint, digital initiative, and digital inclusion blueprint [34]. New Zealand's digital inclusion blueprint identifies motivation, access, skills, and trust as important factors and raises a concern that people who do not possess these factors can be excluded from the digital world [35]. Therefore, these common crucial factors should be considered by policy makers and practitioners to ensure that older adults with or without disabilities become more engaged in mobile internet use and mobile technology.

It was interesting to find that although disability status was not statistically significant in the multiple regression model, all interaction effects were statistically significant when we entered the interaction between variables and disability in each model. Although older adults with disabilities use mobile internet less as compared with those without disabilities, if skills, motivation, and attitude are enhanced through appropriate technology education, their actual use of mobile internet could be facilitated. Today, mobile health care via mobile devices has become prevalent; however, these services have not usually been not targeted toward people with disabilities [36]. Our results thus shed light on the possibility that older adults with various health care needs could be valuable consumers and active users through an increase in their skills, attitudes, and motivation. These are positive signals regarding the disability digital divide; however, mobile accessibility issues and technical barriers should be resolved first [13,37]. Even though older adults have high levels of skills, motivation, or positive attitudes toward technology, if the universal design is not optimized for senior classes, older adults will not be able to use the technology appropriately. Older adults, especially those with disabilities, have many health information needs, and mobile internet use for health purposes could be helpful in not only managing their health but also improving their well-being and health-related outcomes [38]. Further, their mobile internet use can be advantageous, such as by increasing life satisfaction and ensuring social inclusion [39]. Therefore, apart from the common concerns regarding the digital divide of people with disabilities, policy makers and practitioners should pay attention to delivering mobile services and information education to older adults with disabilities, because older adults with disabilities could be the main beneficiaries of mobile services and new technology.

Strengths and Limitations

A strength of this study is that it used nationally representative data from the 2020 Digital Divide Survey in South Korea, a country that has been ranked high in advanced ICT infrastructure among the Organisation for Economic Co-operation and Development countries. Further, this survey was conducted and managed by reliable professional survey firms and government institutions. Therefore, our findings are reliable and can be generalized to other older adults in South Korea. In addition, the findings will be helpful to other countries that have similar ICT infrastructure or technology environments.

The study also has several limitations. First, as the survey was conducted for only five types of disability, it does not encompass other types, such as developmental disabilities, internal organ impairments, and mental illnesses. Since internet use levels, barriers to internet use, and patterns of use could differ according

to the various types of disability and severity [40], further studies that include all types of disability are needed. In addition, variables associated with mobile internet use were limited because this study involved secondary analysis.

Conclusions

This study examined the levels of mobile internet use, factors associated with mobile internet use, and moderating effects of disability on mobile internet use among older adults. The findings have implications for policy makers, social workers, and health care providers who intend to deliver services using mobile internet and technology to improve older adults' quality of life. Although older adults and people with disabilities are considered a vulnerable population, disability creates a stronger association between several determinants and actual mobile internet use. Therefore, appropriate information-literacy education to ensure that older adults engage in mobile internet use and mobile technology is needed. Accordingly, older adults with disabilities could be the main beneficiaries of mobile services and new technology.

Acknowledgments

We thank the Ministry of Science and Information and Communications Technology and the National Information Society Agency for sharing the data with the researchers. This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (grant NRF-2020 R1A6A1A03041989).

Authors' Contributions

EY and KHL contributed to the study concept, designed the study, interpreted the data, wrote original drafts, and reviewed and edited the manuscript. EY was responsible for data acquisition and analysis. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Adedoyin OB, Soykan E. Covid-19 pandemic and online learning: The challenges and opportunities. Int Learn Environ 2020 Sep 02:1-13. [doi: 10.1080/10494820.2020.1813180]
2. Lee SM, Lee D. Opportunities and challenges for contactless healthcare services in the post-COVID-19 era. Technol Forecast Soc Change 2021 Jun;167:120712 [FREE Full text] [doi: 10.1016/j.techfore.2021.120712] [Medline: 33654330]
3. Paek H, Hove T. Information communication technologies (ICTs), crisis communication principles and the COVID-19 response in South Korea. J Creat Commun 2021 Jan 18;16(2):213-221 [FREE Full text] [doi: 10.1177/0973258620981170]
4. Framework Act on Intelligent Informatization. Enforcement date 10 December 2020. Korean Law Information Center. 2020. URL: https://www.law.go.kr/LSW//lsInfoP.do?lsiSeq=218737&chrClsCd=010203&urlMode=engLsInfoR&viewCls=engLsInfoR#0000 [accessed 2022-03-28]
5. Hargittai E. Second-level digital divide: Mapping differences in people's online skills. In: Proceedings of the 29th Research Conference on Communications, Information and Internet Policy. 2001 Presented at: The 29th Research Conference on Communications, Information and Internet Policy; October 27-29, 2001; Alexandria, VA p. 1-19 URL: https://arxiv.org/ftp/cs/papers/0109/0109068.pdf [doi: 10.48550/arXiv.cs/0109068]
6. van Dijk JAGM. The Deepening Divide: Inequality in the Information Society. Thousand Oaks, CA: SAGE Publications, Inc; 2005.
7. Park S, Kim GJ. Lessons from South Korea's digital divide index (DDI). Info 2014 May 06;16(3):72-84. [doi: 10.1108/info-07-2013-0042]
8. Seifert A. The digital exclusion of older adults during the COVID-19 pandemic. J Gerontol Soc Work 2020;63(6-7):674-676. [doi: 10.1080/01634372.2020.1764687] [Medline: 32401181]
9. Wang J, Katz I, Li J, Wu Q, Dai C. Mobile digital divide and older people's access to 'internet plus social work': Implications from the COVID-19 help-seeking cases. Asia Pac J Soc Work Dev 2020 Nov 26;31(1-2):52-58. [doi: 10.1080/02185385.2020.1850332]
10. Seifert A, Cotten SR, Xie B. A double burden of exclusion? Digital and social exclusion of older adults in times of COVID-19. J Gerontol B Psychol Sci Soc Sci 2021 Feb 17;76(3):e99-e103 [FREE Full text] [doi: 10.1093/geronb/gbaa098] [Medline: 32672332]
11. Kim HN. Characteristics of technology adoption by older adults with visual disabilities. Int J Hum Comput Interact 2021 Feb 04;37(13):1256-1268. [doi: 10.1080/10447318.2021.1876359]
12. Duplaga M. Digital divide among people with disabilities: Analysis of data from a nationwide study for determinants of internet use and activities performed online. PLoS One 2017 Jun 29;12(6):e0179825 [FREE Full text] [doi: 10.1371/journal.pone.0179825] [Medline: 28662125]

13. Dobransky K, Hargittai E. Unrealized potential: Exploring the digital disability divide. Poetics 2016 Oct;58:18-28. [doi: 10.1016/J.POETIC.2016.08.003]
14. Khetarpal A. Information and communication technology (ICT) and disability. Rev Mark Integr 2015 Apr 07;6(1):96-113. [doi: 10.1177/0974929214560117]
15. Vicente MR, López AJ. A multidimensional analysis of the disability digital divide: Some evidence for internet use. Inf Soc 2010 Jan 12;26(1):48-64. [doi: 10.1080/01615440903423245]
16. Kim KM, Hwang JH. Factors affecting smartphone online activity use in South Korea: With a focus on the moderating effect of disability status. Univers Access Inf Soc 2020 Sep 25;21(1):109-119. [doi: 10.1007/s10209-020-00758-z]
17. Kakulla B. Personal tech and the pandemic: Older adults are upgrading for a better online experience. AARP. 2021 Sep. URL: https://www.aarp.org/research/topics/technology/info-2021/2021-technology-trends-older-americans.html [accessed 2022-03-28]
18. Ministry of Science and ICT. Final Report on Internet Use Survey 2020. Daegu, South Korea: National Information Society Agency; 2021. URL: https://www.nia.or.kr/site/nia_kor/ex/bbs/View.do?cbIdx=99870&bcIdx=23213&parentSeq=23213 [accessed 2022-03-28]
19. Lee S. Smart divide and the paradigm shift of social capital. J Inf Sci Theory Pract 2020 Sep;9(3):1-13. [doi: 10.1633/JISTAP.2021.9.3.1]
20. Scheerder A, van Deursen A, van Dijk J. Determinants of internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide. Telemat Inform 2017 Dec;34(8):1607-1624. [doi: 10.1016/j.tele.2017.07.007]
21. Cho M, Kim KM. Exploring the disparity in tangible outcomes of internet use between persons with disabilities and persons without disabilities in South Korea. Disabil Health J 2021 Jul;14(3):101101. [doi: 10.1016/j.dhjo.2021.101101] [Medline: 33824093]
22. Cho M, Kim KM. Effect of digital divide on people with disabilities during the COVID-19 pandemic. Disabil Health J 2022 Jan;15(1):101214 [FREE Full text] [doi: 10.1016/j.dhjo.2021.101214] [Medline: 34518115]
23. Gell NM, Rosenberg DE, Demiris G, LaCroix AZ, Patel KV. Patterns of technology use among older adults with and without disabilities. Gerontologist 2015 Jun;55(3):412-421 [FREE Full text] [doi: 10.1093/geront/gnt166] [Medline: 24379019]
24. Enforcement Decree of the Act on Welfare of Persons with Disabilities. Enforcement date 13 April 2021. Korean Law Information Center. 2021. URL: https://www.law.go.kr/LSW//lsInfoP.do?lsiSeq=231515&chrClsCd=010203&urlMode=engLsInfoR&viewCls=engLsInfoR#0000 [accessed 2022-03-28]
25. van Deursen A, van Dijk J. Internet skills and the digital divide. New Media Soc 2010 Dec 06;13(6):893-911. [doi: 10.1177/1461444810386774]
26. Changki J, WookJoon S. The effect of digital divide on the use of online services in the severely disabled people: Focusing on online accessibility. Informatization Policy 2020 Sep;27(3):56-81. [doi: 10.22693/NIAIP.2020.27.3.056]
27. Lee IH. EasyFlow Statistics macro: Excel macro ver 1.8. StatEdu. 2020. URL: http://doi.org/10.22934/StatEdu.2020.01 [accessed 2022-03-28]
28. Scholz F, Yalcin B, Priestley M. Internet access for disabled people: Understanding socio-relational factors in Europe. Cyberpsychology 2017 May 31;11(1):4 [FREE Full text] [doi: 10.5817/CP2017-1-4]
29. Damaceno RJP, Braga JC, Mena-Chalco JP. Mobile device accessibility for the visually impaired: Problems mapping and recommendations. Univers Access Inf Soc 2017 May 4;17(2):421-435. [doi: 10.1007/s10209-017-0540-1]
30. Holzinger A, Searle G, Nischelwitzer A. On some aspects of improving mobile applications for the elderly. In: Proceedings of the 4th International Conference on Universal Access in Human-Computer Interaction. Berlin, Germany: Springer; 2007 Presented at: 4th International Conference on Universal Access in Human-Computer Interaction; July 22-27, 2007; Beijing, China p. 923-932. [doi: 10.1007/978-3-540-73279-2_103]
31. Ministry of Science and ICT. 2020 Web Accessibility Report. Daegu, South Korea: National Information Society Agency; 2021 Mar. URL: https://www.msit.go.kr/bbs/view.do?sCode=user&mId=99&mPid=74&pageIndex=&bbsSeqNo=79&nttSeqNo=3173397&searchOpt=ALL&searchTxt= [accessed 2022-03-28]
32. Kim HS, Kim JW, Chung KS. The future direction of information literacy for the aged in the lifelong learning: Centering on Gumi Senior Welfare Center. J Digit Contents Soc 2014 Aug 31;15(4):491-500. [doi: 10.9728/dcs.2014.15.4.491]
33. Digital competency centers. National Information Society Agency. 2021. URL: https://www.xn--2z1bw8k1pjz5ccumkb.kr/main.do [accessed 2022-03-28]
34. Ng IH, Sun LS, Pang N, Lim D, Soh G, Pakianathan PS, et al. From Digital Exclusion to Universal Digital Access in Singapore. Singapore: Faculty of Arts and Social Sciences, National University of Singapore; 2021 Jan. URL: https://fass.nus.edu.sg/ssr/wp-content/uploads/sites/8/2021/01/Digital-Access-20210118.pdf [accessed 2022-03-28]
35. Hartnett M, Fields A. Digital inclusion in New Zealand. J Open Flex Distance Learn 2019 Jan;23(2):1-4 [FREE Full text]
36. Jones M, Morris J, Deruyter F. Mobile healthcare and people with disabilities: Current state and future needs. Int J Environ Res Public Health 2018 Mar 14;15(3):515 [FREE Full text] [doi: 10.3390/ijerph15030515] [Medline: 29538292]
37. Agree EM. The potential for technology to enhance independence for those aging with a disability. Disabil Health J 2014 Jan;7(1 Suppl):S33-S39 [FREE Full text] [doi: 10.1016/j.dhjo.2013.09.004] [Medline: 24456682]

38. Duplaga M, Szulc K. The association of internet use with wellbeing, mental health and health behaviours of persons with disabilities. Int J Environ Res Public Health 2019 Sep 04;16(18):3252 [FREE Full text] [doi: 10.3390/ijerph16183252] [Medline: 31487949]
39. Drainoni M, Houlihan B, Williams S, Vedrani M, Esch D, Lee-Hood E, et al. Patterns of internet use by persons with spinal cord injuries and relationship to health-related quality of life. Arch Phys Med Rehabil 2004 Nov;85(11):1872-1879. [doi: 10.1016/j.apmr.2004.07.350] [Medline: 15520984]
40. Johansson S, Gulliksen J, Gustavsson C. Disability digital divide: The use of the internet, smartphones, computers and tablets among people with disabilities in Sweden. Univ Access Inf Soc 2020 Mar 07;20(1):105-120. [doi: 10.1007/s10209-020-00714-x]

Abbreviations

ICT: information and communications technology
IRB: Institutional Review Board
NRF: National Research Foundation of Korea

Edited by A Mavragani; submitted 11.02.22; peer-reviewed by K Colorafi, A Garcia; comments to author 04.03.22; revised version received 14.03.22; accepted 15.03.22; published 04.04.22.

Please cite as:
Yang E, Lee KH
The Moderating Effects of Disability on Mobile Internet Use Among Older Adults: Population-Based Cross-sectional Study
J Med Internet Res 2022;24(4):e37127
URL: https://www.jmir.org/2022/4/e37127
doi:10.2196/37127
PMID:35377329

Original Paper

Associations Between the Digital Clock Drawing Test and Brain Volume: Large Community-Based Prospective Cohort (Framingham Heart Study)

Jing Yuan[1,2], MD; Rhoda Au[2,3,4,5,6,7], MBA, PhD; Cody Karjadi[2,3], MSc; Ting Fang Ang[2,3,6], MD, MPH; Sherral Devine[2,3], PhD; Sanford Auerbach[3,4], MD; Charles DeCarli[8], MD, FAAN, FAHA; David J Libon[9], PhD; Jesse Mez[3,4,7], MSc, MD; Honghuang Lin[10], PhD

[1]Department of Neurology, Peking Union Medical College Hospital, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing, China

[2]Department of Anatomy and Neurobiology, Boston University School of Medicine, Boston University, Boston, MA, United States

[3]Framingham Heart Study, Boston University School of Medicine, Boston University, Boston, MA, United States

[4]Department of Neurology, Boston University School of Medicine, Boston University, Boston, MA, United States

[5]Department of Epidemiology, Boston University School of Public Health, Boston University, Boston, MA, United States

[6]Slone Epidemiology Center, Boston University School of Medicine, Boston University, Boston, MA, United States

[7]Alzheimer's Disease Research Center, Boston University, Boston, MA, United States

[8]Department of Neurology and Center for Neuroscience, University of California, Davis, Sacramento, CA, United States

[9]Department of Geriatrics and Gerontology and Department of Psychology, New Jersey Institute for Successful Aging, Rowan University, School of Osteopathic Medicine, Stratford, NJ, United States

[10]Division of Clinical Informatics, Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA, United States

Corresponding Author:
Honghuang Lin, PhD
Division of Clinical Informatics
Department of Medicine
University of Massachusetts Chan Medical School
55 Lake Ave North, S7-749
Worcester, MA, 01655
United States
Phone: 1 774 455 4881
Email: Honghuang.Lin3@umassmed.edu

Abstract

Background: The digital Clock Drawing Test (dCDT) has been recently used as a more objective tool to assess cognition. However, the association between digitally obtained clock drawing features and structural neuroimaging measures has not been assessed in large population-based studies.

Objective: We aimed to investigate the association between dCDT features and brain volume.

Methods: This study included participants from the Framingham Heart Study who had both a dCDT and magnetic resonance imaging (MRI) scan, and were free of dementia or stroke. Linear regression models were used to assess the association between 18 dCDT composite scores (derived from 105 dCDT raw features) and brain MRI measures, including total cerebral brain volume (TCBV), cerebral white matter volume, cerebral gray matter volume, hippocampal volume, and white matter hyperintensity (WMH) volume. Classification models were also built from clinical risk factors, dCDT composite scores, and MRI measures to distinguish people with mild cognitive impairment (MCI) from those whose cognition was intact.

Results: A total of 1656 participants were included in this study (mean age 61 years, SD 13 years; 50.9% women), with 23 participants diagnosed with MCI. All dCDT composite scores were associated with TCBV after adjusting for multiple testing (*P* value <.05/18). Eleven dCDT composite scores were associated with cerebral white matter volume, but only 1 dCDT composite score was associated with cerebral gray matter volume. None of the dCDT composite scores was associated with hippocampal volume or WMH volume. The classification model for differentiating MCI and normal cognition participants, which incorporated age, sex, education, MRI measures, and dCDT composite scores, showed an area under the curve of 0.897.

Conclusions: dCDT composite scores were significantly associated with multiple brain MRI measures in a large community-based cohort. The dCDT has the potential to be used as a cognitive assessment tool in the clinical diagnosis of MCI.

(J Med Internet Res 2022;24(4):e34513) doi:10.2196/34513

KEYWORDS

Clock Drawing Test; digital; neuropsychological test; cognitive; technology; Boston Process Approach; neurology; Framingham Heart Study; dementia; Alzheimer

Introduction

The Clock Drawing Test (CDT) is a widely used assessment tool able to screen for impairment associated with mild cognitive impairment (MCI) [1] and dementia [2-5]. The most common instruction for clock drawing asks participants to draw the face of a clock, put in all of the numbers, and set the hands for 10 past 11. This is followed by asking participants to copy a model of a clock [3,6]. Multiple cognitive domains are involved in completion of the test, including graphomotor ability, attention, syntactic comprehension, visual and semantic memory, executive function, and visuoconstructional ability [6]. Previous studies found that CDT performance evaluated by various manual scoring systems was correlated with a variety of cortical and subcortical areas on magnetic resonance imaging (MRI), without showing consistent and specific brain localization [7-9].

A significant innovation is the introduction of a digital CDT (dCDT), which was jointly developed by the Massachusetts Institute of Technology and Lahey Hospital & Medical Center, in collaboration with the Clock Sketch Consortium [10-13]. In addition to retaining the classic features of the traditional CDT, including easy, inexpensive, noninvasive administration, the dCDT captures more than 100 latencies and graphomotor features for subtle cognitive changes, which would be difficult or impossible without this technology [12-14]. A recent study reported that the DCTClock test (a commercial version of the dCDT) showed excellent discrimination between diagnostic groups of normal cognition (NC) and MCI or early Alzheimer dementia [14]. The test was also associated with amyloid β and tau burden in positron emission tomography scans among NC participants [14].

In our prior study [15], we validated the psychometric characteristics of the dCDT against standard paper-pencil neuropsychological (NP) tests in the community-based Framingham Heart Study (FHS) [15]. The composite scores derived from dCDT features were significantly associated with both NP test results and mild cognitive impairment [15]. Nonetheless, it is unclear whether dCDT composite scores are associated with MRI measures of atrophy, which have been used as important markers for cognitive impairment and cerebral small vessel disease injuries [16-20].

The objectives of this study were to investigate the associations between dCDT features and MRI measures in the FHS, and to assess the diagnostic potential of dCDT features as a surrogate for standard NP tests in the differentiation of MCI from NC that is validated against MRI measures of neurodegeneration.

Methods

Study Sample

The FHS is a community-based prospective cohort study that was established in 1948. Three generations of participants have been enrolled and followed up every 2 to 8 years. Details on the FHS cohorts can be found in previous publications [21-23]. As part of an ancillary study, participants were invited to undergo NP assessment and brain MRI scans regularly. The dCDT has been administered as part of the NP assessment along with other NP tests since 2011. This study included participants who completed at least one dCDT and had a contemporary MRI within 6 months (98.2% had MRI scans on the same day as the dCDT). Participants with prevalent dementia and stroke were excluded (n=84). We further excluded participants who were flagged as having possible cognitive impairment, but were not reviewed by an expert panel (n=139).

Ethics Approval

The Boston University Medical Campus Institutional Review Board approved the study procedures and protocols. Written informed consent was obtained from all participants. All the data used for this study could be requested through the FHS research website [24].

Standard NP Tests and the dCDT

In the FHS, standard NP tests were administered and scored to produce traditional quantitative measures of cognitive performance [25]. These tests included the Wechsler Memory Scale [26] Logical Memory—Immediate Recall, Delayed Recall, and Recognition; Visual Reproduction—Immediate Recall, Delayed Recall, and Recognition; and Paired Associate Learning—Immediate Recall, Delayed Recall, and Recognition; the Wechsler Adult Intelligence Scale [27] Digit Span—Forward and Backward and Similarities; the Boston Naming Test—30 item version [28]; the Trail Making Test A and B [29]; the Hooper Visual Organization Test [30]; and Verbal Fluency (FAS & FAS-Animal) [31,32]. These NP tests measure multiple cognitive domains of verbal memory, visual memory, attention and concentration, executive function, abstract reasoning, language, and visuoperceptual organization. For all tests, higher scores indicate better performance with the exception of the Trail Making Test A and B, where a shorter completion time indicates better cognitive performance.

Since October 2011, FHS participants have been administered the dCDT with a digital pen during their regular NP test visit by trained examiners [10-13]. The digital pen measures the pen position 80 times per second at a spatial resolution of ±0.002 inches [10]. After the dCDT test, the pen was connected to a

computer with preinstalled software that automatically extracted the drawing information and classified each pen stroke and associated latencies. An external rater manually examined each drawing to ensure appropriate classification of strokes using a user-friendly drag-and-drop interface to correct any classification error. A total of 105 dCDT features were derived and used from the entire drawing process, which captured the strokes, latencies, and spatial relationships as the measures of drawing efficiency, simple motor functioning, information processing, and spatial reasoning (Multimedia Appendix 1). As reported in our previous study [15], we further derived composite scores based on dCDT features significantly associated with 18 NP tests, which represented a weighted combination of dCDT features that were previously associated with a specific NP test.

Acquisition and Measurement of MRI Variables

The MRI protocol in the FHS has been described previously [33]. Briefly, participants were imaged using a Siemens 1.5T field strength machine (Siemens Medical) and a 3-dimensional T1-weighted coronal spoiled gradient-recalled echo sequence. All images were transferred to and processed by the University of California Davis Medical Center for centralized processing. Segmentation of brain structural MRI was performed by semiautomated procedures, and the complete information can be obtained elsewhere [33]. In brief, gray matter, white matter, and cerebrospinal fluid segmentation were performed using an expectation-maximization algorithm after skull stripping and intensity inhomogeneity correction. The hippocampus was segmented by a multi-atlas hippocampal segmentation algorithm [34]. White matter hyperintensity (WMH) was segmented on a combination of FLAIR and 3D T1 images using a modified Bayesian probability structure [35]. Total cerebral cranial volume (TCV) was determined by outlining the intracranial vault lying above the tentorium and was used for correcting the head size. The primary MRI measure was total cerebral brain volume (TCBV). Secondary measures were cerebral white matter volume, cerebral gray matter volume, hippocampal volume, and WMH volume. All MRI measures were corrected for head size by calculating the percentage of these volumes over the TCV. The percent of WMH/TCV was log-transformed for normality. The WMH volume was used both as a continuous variable and as a dichotomous variable (large WMH volume versus no or minimal WMH volume), similar to a previous study [36]. A large WMH volume (WMH-Large) was defined as a volume more than one standard deviation higher than the age-specific mean value.

Case Ascertainment of MCI

The detailed cognitive ascertainment procedure and quality control in the FHS have been described previously [37,38]. Briefly, participants received NP assessments on average every 4 to 5 years, in addition to their regular research center examinations. For participants with possible cognitive impairment, more frequent NP tests were conducted on average every 1 to 2 years, and neurological examinations were performed on a subset of these participants. A clinical review was triggered when there was an indication of a potential cognitive impairment and/or decline, and was conducted by a panel consisting of at least one neurologist and one neuropsychologist. The review panel determined the MCI diagnosis, which required evidence of a decline in cognitive performance in one or more cognitive domains, no records indicating functional decline, and not meeting the criteria for dementia. Although the Clinical Dementia Rating (CDR) scale [39] was not formally applied, the review panel used the CDR scoring scale (0-3) to quantify the severity of impairment; all MCI cases were given a rating of 0.5.

Statistical Analysis

The dCDT composite scores were derived by linear regression models based on dCDT features significantly associated with 18 NP tests, which represented a weighted combination of dCDT features that were previously associated with a specific NP test [15]. The scores and all MRI measures were normalized with a mean of 0 and standard deviation of 1. Linear regression models were used to assess the associations of dCDT composite scores with MRI measures, adjusting for age, sex, and education. The same models were also used to test the association between individual dCDT features and MRI measures. In the sensitivity analysis, the models were additionally adjusted for vascular risk factors, including hypertension, diabetes, smoking, and prevalent atrial fibrillation. Bonferroni correction was used to adjust for multiple testing in the linear regression models. Significant associations were claimed if $P<.05/N$, where N was the number of tests performed.

Logistic regression models were also constructed to classify MCI using 4 different sets of predictors. The analysis was restricted to participants who were 65 years or older at the time of NP testing (only 1 case was diagnosed before 65 years). Model 1 included only age, sex, and education. Model 2 included age, sex, education, and MRI measures (TCBV, cerebral white matter volume, cerebral gray matter volume, and hippocampal volume). Model 3 included age, sex, education, and dCDT composite scores. Model 4 included age, sex, education, MRI measures, and dCDT composite scores. Model performance was assessed by the area under the curve (AUC). All statistical analyses were performed using R software version 4.0.3 (R Core Team) [40].

Results

As shown in Table 1, our study sample included 1656 participants (mean age 61 years, SD 13 years; 50.9% women; 42.9% received college-level education or higher) who were free of dementia or stroke when NP tests were conducted. Among them, 23 participants were diagnosed with MCI at the time of or before their NP tests but had not progressed to the threshold of clinical dementia diagnosis. As expected, participants with MCI were generally older and had worse cognitive performance, smaller cerebral volume measures, and larger WMH volume on MRI than those in the NC group.

Table 1. Clinical characteristics of the study sample.

Variable	Total (N=1656)	MCI[a] (n=23)	NC[b] (n=1633)
Age (years), mean (SD)	61 (13)	78 (7)	61 (13)
Women, n (%)	843 (50.9)	14 (60.9)	829 (50.8)
Education, n (%)			
No high school	187 (11.3)	3 (13.0)	184 (11.3)
High school	319 (19.3)	6 (26.1)	313 (19.2)
Some college	440 (26.6)	3 (13.0)	437 (26.8)
College and higher	710 (42.9)	11 (47.8)	699 (42.8)
MRI[c] measures[d], mean (SD)			
Total cerebral cranial volume (cm^3)	1251 (129)	1252 (104)	1251 (129)
Total cerebral brain volume (%)	77.0 (2.9)	72.8 (2.5)	77.0 (2.9)
Cerebral white matter volume (%)	36.3 (2.4)	34.3 (2.7)	36.3 (2.4)
Cerebral gray matter volume (%)	40.4 (2.1)	37.4 (2.2)	40.4 (2.1)
Hippocampal volume (%)	0.54 (0.05)	0.50 (0.05)	0.54 (0.05)
Log (white matter hyperintensity volume) (%)	−2.33 (1.48)	−0.59 (1.38)	−2.35 (1.47)
Neuropsychological test scores, mean (SD)			
LMi[e]	13 (3)	10 (4)	13 (3)
LMd[e]	12 (4)	8 (4)	12 (4)
LMr[e]	10 (1)	9 (2)	10 (1)
VRi[f]	9 (3)	5 (2)	9 (3)
VRd[f]	8 (3)	3 (3)	8 (3)
VRr[f]	3 (1)	2 (1)	3 (1)
PASi[g]	15 (3)	11 (3)	15 (3)
PASd[g]	9 (1)	7 (2)	9 (1)
PASr[g]	10 (1)	9 (2)	10 (1)
DSf[h]	7 (1)	6 (1)	7 (1)
DSb[h]	5 (1)	4 (1)	5 (1)
Trails A[i] (s)	31 (13)	44 (14)	31 (13)
Trails B[i] (s)	83 (65)	219 (159)	82 (61)
Similarities	17 (3)	14 (5)	17 (3)
Hooper Visual Organization Test	26 (3)	23 (4)	26 (3)
Boston Naming Test—30 item version	26 (7)	24 (7)	26 (7)
FAS[j]	41 (12)	32 (12)	41 (12)
FAS-Animal[j]	19 (6)	15 (5)	19 (6)

[a]MCI: mild cognitive impairment.

[b]NC: normal cognition.

[c]MRI: magnetic resonance imaging.

[d]All MRI measures were corrected for head size by calculating the percentage of the volumes over the total cerebral cranial volume above the tentorium. The percentage of white matter hyperintensity volume/total cerebral cranial volume was log transformed.

[e]Wechsler Memory Scale Logical Memory—Immediate Recall (LMi), Delayed Recall (LMd), and Recognition (LMr).

[f]Visual Reproduction—Immediate Recall (VRi), Delayed Recall (VRd), and Recognition (VRr).

[g]Paired Associate Learning—Immediate Recall (PASi), Delayed Recall (PASd), and Recognition (PASr).

[h]Wechsler Adult Intelligence Scale Digit Span—Forward (DSf) and Backward (DSb).

[i]Trail Making Test A (Trails A) and B (Trails B).

[j]Verbal fluency test (FAS and FAS-Animal).

As shown in Table 2, all dCDT composite scores (n=18) were significantly associated with TCBV after adjusting for multiple testing. Better dCDT performance was associated with larger TCBV. The effect size of composite scores for visual memory and visuoperceptual organization tended to be slightly higher than other dCDT composite scores.

Table 2. Association between digital Clock Drawing Test composite scores and total cerebral brain volume.

dCDT[a] composite score[b]	Effect size	Standard error	*P* value[c]
dCDT_LMi[d]	0.079	0.017	3.2×10^{-6}
dCDT_LMd[d]	0.083	0.017	1.1×10^{-6}
dCDT_LMr[d]	0.061	0.017	2.9×10^{-4}
dCDT_VRi[e]	0.098	0.017	1.4×10^{-8}
dCDT_VRd[e]	0.099	0.017	1.2×10^{-8}
dCDT_VRr[e]	0.093	0.017	6.4×10^{-8}
dCDT_PASi[f]	0.079	0.017	3.5×10^{-6}
dCDT_PASd[f]	0.084	0.017	8.8×10^{-7}
dCDT_PASr[f]	0.073	0.017	1.7×10^{-5}
dCDT_DSf[g]	0.071	0.017	2.5×10^{-5}
dCDT_DSb[g]	0.067	0.017	6.5×10^{-5}
dCDT_Trails A[h]	−0.074	0.017	1.2×10^{-5}
dCDT_Trails B[h]	−0.080	0.017	3.0×10^{-6}
dCDT_SIM[i]	0.084	0.017	9.4×10^{-7}
dCDT_HVOT[j]	0.094	0.017	6.1×10^{-8}
dCDT_BNT30[k]	0.075	0.017	9.9×10^{-6}
dCDT_FAS[l]	0.083	0.017	1.5×10^{-6}
dCDT_FAS-Animal[l]	0.062	0.017	2.2×10^{-4}

[a]dCDT: digital Clock Drawing Test.

[b]The models were adjusted for age, sex, and education.

[c]Bonferroni correction was used to adjust for multiple testing, and significant associations were claimed if $P<.05/18$ (2.8×10^{-3}), where 18 is the number of tests performed. All *P* values were significant.

[d]Wechsler Memory Scale Logical Memory—Immediate Recall (LMi), Delayed Recall (LMd), and Recognition (LMr).

[e]Visual Reproduction—Immediate Recall (VRi), Delayed Recall (VRd), and Recognition (VRr).

[f]Paired Associate Learning—Immediate Recall (PASi), Delayed Recall (PASd), and Recognition (PASr).

[g]Wechsler Adult Intelligence Scale Digit Span—Forward (DSf) and Backward (DSb).

[h]Trail Making Test A (Trails A) and B (Trails B).

[i]SIM: Similarities.

[j]HVOT: Hooper Visual Organization Test.

[k]BNT30: Boston Naming Test—30 item version.

[l]Verbal fluency test (FAS and FAS-Animal).

As shown in Table 3, cerebral white matter volume was significantly associated with 11 dCDT composite scores involving the comprehensive cognitive domains of verbal memory, visual memory, abstract reasoning, language, and visuoperceptual organization (all with $P<2.8\times10^{-3}$). In contrast, cerebral gray matter volume was only associated with the visual memory composite score (dCDT_Visual Reproduction—Immediate Recall, β=.059; $P=2.6\times10^{-3}$).

Hippocampal volume was not significantly associated with any dCDT composite scores. As shown in Table 4, none of the dCDT composite scores was significantly associated with WMH volume.

Table 3. Association of digital Clock Drawing Test composite scores with cerebral white matter, gray matter, and hippocampal volumes.

dCDT[a] composite score[b]	Cerebral white matter volume[c]			Cerebral gray matter volume[c]			Hippocampal volume[c]		
	Effect size	Standard error	*P* value[d]	Effect size	Standard error	*P* value[d]	Effect size	Standard error	*P* value[d]
dCDT_LMi[e]	0.071	0.023	1.9×10^{-3}[f]	0.042	0.019	2.8×10^{-2}	0.033	0.025	2.0×10^{-1}
dCDT_LMd[e]	0.075	0.023	1.1×10^{-3}[f]	0.045	0.019	2.0×10^{-2}	0.038	0.025	1.3×10^{-1}
dCDT_LMr[e]	0.052	0.022	2.0×10^{-2}	0.034	0.019	7.1×10^{-2}	0.019	0.025	4.6×10^{-1}
dCDT_VRi[g]	0.083	0.023	4.2×10^{-4}[f]	0.059	0.020	2.6×10^{-3}[f]	0.054	0.026	3.8×10^{-2}
dCDT_VRd[g]	0.084	0.023	3.0×10^{-4}[f]	0.058	0.020	3.1×10^{-3}	0.054	0.026	3.9×10^{-2}
dCDT_VRr[g]	0.080	0.023	5.6×10^{-4}[f]	0.054	0.020	5.5×10^{-3}	0.052	0.026	4.3×10^{-2}
dCDT_PASi[h]	0.070	0.023	2.3×10^{-3}[f]	0.044	0.019	2.2×10^{-2}	0.035	0.025	1.7×10^{-1}
dCDT_PASd[h]	0.078	0.023	7.3×10^{-4}[f]	0.044	0.019	2.4×10^{-2}	0.041	0.026	1.1×10^{-1}
dCDT_PASr[h]	0.067	0.023	3.2×10^{-3}	0.036	0.019	6.2×10^{-2}	0.047	0.025	6.5×10^{-2}
dCDT_DSf[i]	0.063	0.023	5.7×10^{-3}	0.041	0.019	3.4×10^{-2}	0.026	0.025	3.1×10^{-1}
dCDT_DSb[i]	0.059	0.023	9.0×10^{-3}	0.037	0.019	5.0×10^{-2}	0.023	0.025	3.7×10^{-1}
dCDT_Trails A[j]	−0.063	0.023	6.0×10^{-3}	−0.045	0.019	1.8×10^{-2}	−0.046	0.025	7.0×10^{-2}
dCDT_Trails B[j]	−0.069	0.023	2.9×10^{-3}	−0.047	0.019	1.4×10^{-2}	−0.047	0.025	6.7×10^{-2}
dCDT_SIM[k]	0.071	0.023	2.0×10^{-3}[f]	0.049	0.019	1.2×10^{-2}	0.034	0.026	1.8×10^{-1}
dCDT_HVOT[l]	0.083	0.023	4.0×10^{-4}[f]	0.052	0.020	8.0×10^{-3}	0.048	0.026	6.4×10^{-2}
dCDT_BNT30[m]	0.071	0.023	1.9×10^{-3}[f]	0.036	0.019	5.9×10^{-2}	0.028	0.025	2.7×10^{-1}
dCDT_FAS[n]	0.071	0.023	2.1×10^{-3}[f]	0.047	0.019	1.5×10^{-2}	0.040	0.026	1.2×10^{-1}
dCDT_FAS-Animal[n]	0.054	0.023	1.7×10^{-2}	0.034	0.019	7.1×10^{-2}	0.021	0.025	4.0×10^{-1}

[a]dCDT: digital Clock Drawing Test.

[b]The models were adjusted for age, sex, and education.

[c]All magnetic resonance imaging measures were the percentage of the volumes over the total cerebral cranial volume above the tentorium.

[d]Bonferroni correction was used to adjust for multiple testing, and significant associations were claimed if $P<.05/18$ (2.8×10^{-3}), where 18 is the number of tests performed.

[e]Wechsler Memory Scale Logical Memory—Immediate Recall (LMi), Delayed Recall (LMd), and Recognition (LMr).

[f]Significant.

[g]Visual Reproduction—Immediate Recall (VRi), Delayed Recall (VRd), and Recognition (VRr).

[h]Paired Associate Learning—Immediate Recall (PASi), Delayed Recall (PASd), and Recognition (PASr).

[i]Wechsler Adult Intelligence Scale Digit Span—Forward (DSf) and Backward (DSb).

[j]Trail Making Test A (Trails A) and B (Trails B).

[k]SIM: Similarities.

[l]HVOT: Hooper Visual Organization Test.

[m]BNT30: Boston Naming Test—30 item version.

[n]Verbal fluency test (FAS and FAS-Animal).

Table 4. Association between digital Clock Drawing Test composite scores and white matter hyperintensity volume.

dCDT[a] composite score[b]	White matter hyperintensity volume[c]			White matter hyperintensity-Large[d]		
	Effect size	Standard error	*P* value[e]	Effect size	Standard error	*P* value[e]
dCDT_LMi[f]	0.040	0.018	3.0×10^{-2}	−0.120	0.073	1.1×10^{-1}
dCDT_LMd[f]	−0.041	0.018	2.5×10^{-2}	−0.130	0.073	8.2×10^{-2}
dCDT_LMr[f]	−0.035	0.018	5.0×10^{-2}	−0.110	0.071	1.3×10^{-1}
dCDT_VRi[g]	−0.042	0.019	2.3×10^{-2}	−0.150	0.074	3.7×10^{-2}
dCDT_VRd[g]	−0.043	0.019	2.0×10^{-2}	−0.150	0.074	3.7×10^{-2}
dCDT_VRr[g]	−0.043	0.019	2.2×10^{-2}	−0.160	0.074	2.9×10^{-2}
dCDT_PASi[h]	−0.036	0.018	5.0×10^{-2}	−0.100	0.073	1.5×10^{-1}
dCDT_PASd[h]	−0.040	0.018	3.0×10^{-2}	−0.130	0.073	8.0×10^{-2}
dCDT_PASr[h]	−0.020	0.018	2.7×10^{-1}	−0.035	0.073	6.3×10^{-1}
dCDT_DSf[i]	−0.042	0.018	2.2×10^{-2}	−0.160	0.072	3.0×10^{-2}
dCDT_DSb[i]	−0.036	0.018	4.7×10^{-2}	−0.100	0.072	1.5×10^{-1}
dCDT_Trails A[j]	0.037	0.018	4.4×10^{-2}	0.160	0.073	3.1×10^{-2}
dCDT_Trails B[j]	0.039	0.018	3.4×10^{-2}	0.150	0.074	3.6×10^{-2}
dCDT_SIM[k]	−0.039	0.018	3.5×10^{-2}	−0.130	0.073	6.6×10^{-2}
dCDT_HVOT[l]	−0.040	0.019	3.4×10^{-2}	−0.140	0.074	5.7×10^{-2}
dCDT_BNT30[m]	−0.037	0.018	4.5×10^{-2}	−0.120	0.073	9.8×10^{-2}
dCDT_FAS[n]	−0.041	0.018	2.5×10^{-2}	−0.160	0.074	3.2×10^{-2}
dCDT_FAS-Animal[n]	−0.035	0.018	4.8×10^{-2}	−0.110	0.071	1.3×10^{-1}

[a]dCDT: digital Clock Drawing Test.

[b]The models were adjusted for age, sex, and education.

[c]The white matter hyperintensity volume was the percentage over the total cerebral cranial volume above the tentorium and was log transformed.

[d]The large white matter hyperintensity volume was defined as that more than one standard deviation higher than the age-specific mean value.

[e]Bonferroni correction was used to adjust for multiple testing, and significant associations were claimed if $P<.05/18$ (2.8×10^{-3}), where 18 is the number of tests performed.

[f]Wechsler Memory Scale Logical Memory—Immediate Recall (LMi), Delayed Recall (LMd), and Recognition (LMr).

[g]Visual Reproduction—Immediate Recall (VRi), Delayed Recall (VRd), and Recognition (VRr).

[h]Paired Associate Learning—Immediate Recall (PASi), Delayed Recall (PASd), and Recognition (PASr).

[i]Wechsler Adult Intelligence Scale Digit Span—Forward (DSf) and Backward (DSb).

[j]Trail Making Test A (Trails A) and B (Trails B).

[k]SIM: Similarities.

[l]HVOT: Hooper Visual Organization Test.

[m]BNT30: Boston Naming Test—30 item version.

[n]Verbal fluency test (FAS and FAS-Animal).

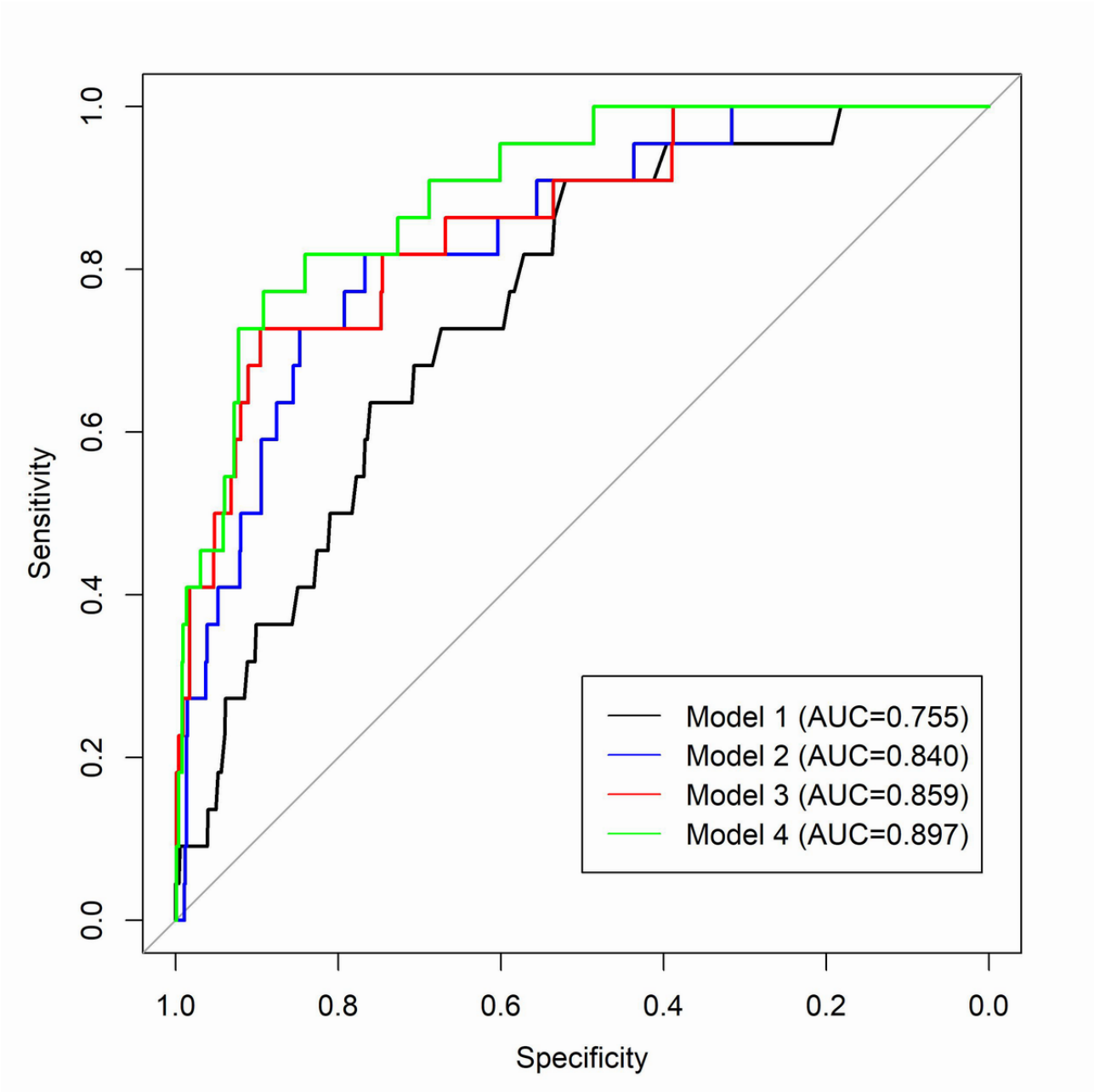
Further detailed analysis for cortical gray matter and specific brain areas (frontal, parietal, temporal, and occipital cortical gray matter volumes) is shown in Multimedia Appendix 2. We found that cortical gray matter volume was associated with visual memory composite scores ($P=1.4\times10^{-3}$ for dCDT_Visual Reproduction—Immediate Recall and 1.7×10^{-3} for dCDT_Visual Reproduction—Delayed Recall), which was consistent with the results for cerebral gray matter volume shown in Table 3. For specific cortical regions, parietal and temporal cortical gray matter volumes were associated with dCDT composite scores, whereas no association was found for frontal and occipital cortical gray matter volumes. The results remained largely the same after additionally adjusting for vascular risk factors (Multimedia Appendix 3, Multimedia Appendix 4, Multimedia Appendix 5, and Multimedia Appendix 6).

We also examined the associations between individual dCDT features and MRI measures. Among the 105 dCDT features, 17

of them were significantly associated with TCBV. However, only 2 features were associated with cerebral white matter volume and 4 features were associated with cerebral gray matter volume (Multimedia Appendix 7). The scattered associations suggest that, rather than individual features, the composite scores can better characterize the combinatory effects of multiple features.

We then built 4 classification models of MCI for participants over 65 years old, each with a different set of predictors. As shown in Figure 1, the AUCs were 0.755, 0.840, 0.859, and 0.897 for Models 1-4, respectively. Model 4 that incorporated traditional risk factors (age, sex, and education), MRI measures, and dCDT composite scores showed superior performance compared with models consisting of solely MRI or dCDT composite scores with traditional risk factors (Models 1-3).

Figure 1. Performance of classification models to distinguish mild cognitive impairment from normal cognition. AUC: area under the curve.



Discussion

In this study, we investigated the potential of the dCDT as a cognitive assessment tool by studying the association of dCDT features with brain MRI structural measures in more than 1600 participants. We observed significant associations between dCDT composite scores and multiple MRI measures, including TCBV, cerebral white matter volume, and cerebral gray matter volume. However, no association was observed between dCDT composite scores and hippocampal volume or WMH volume. The combination of dCDT composite scores with MRI measures and clinical risk factors reached an AUC of 0.897 to distinguish MCI from normal cognition. Given that standard NP tests are time burdensome and labor intensive, the dCDT might be used as an alternative tool to screen for MCI in a large population [12,13,41,42].

A previous study found that a smaller TCBV was associated with worse performance in Visual Reproduction—Immediate Recall, Visual Reproduction—Delayed Recall, Visual Reproduction—Recognition, the Trail Making Test A and B, and the Hooper Visual Organization Test [16]. Notably, this study found that global cerebrum atrophy was associated with several dCDT measures, suggesting a more extensive deficit in multiple cognitive domains, which might be used as an indicator of general "brain health." Interestingly, dCDT performance tended to be better correlated with cerebral white matter volume, and parietal and temporal cortical gray matter volumes, suggesting that they may be the primary neuroanatomical basis for cognitive processes in the dCDT. While the cerebral cortex has long been considered as the major neuroanatomical basis of cognitive function, white matter is increasingly recognized as equally critical for cognition [43-45]. Cerebral white matter contains neural fibers that are the extensions of neurons into subcortical regions [46]. The neural network composed of cortical neurons and subcortical neural fibers is essential to maintain normal neural signal transmission and functional connectivity for cognitive processing [46]. Nonetheless, although there was a significant association between the dCDT and cerebral white matter volume, we did not find an association between dCDT performance and WMH volume. One possible reason is the exclusion of dementia patients in our sample, who generally have a larger WMH volume compared to that in individuals with normal cognition or MCI. Another reason may be that the dCDT performance was more closely associated with white matter atrophy rather than localized white matter injuries.

Several limitations of our study, however, merit consideration. First, the cross-sectional design could not reveal a temporal relationship between dCDT features and MRI measures. Future longitudinal analysis will investigate whether dCDT features could serve as early cognitive markers for predicting incident brain structural changes or incident MCI or dementia. Second, only a moderate number of MCI cases were observed in this study, and the number of dementia cases was too small to study. The diagnostic value of the dCDT could be further validated when more incident cases are observed during longer follow-up periods in the FHS. In addition, besides the Harvard Aging Brain Study, future studies with larger sample sizes of MCI and early Alzheimer dementia in more diverse populations (population-based or hospital-based samples) are warranted to further test the diagnostic performance before extensive application of the dCDT as a substitute for classic NP tests in classification and selection criteria for clinical research and clinical trials. Third, standard NP test performance was used during the dementia diagnosis review process, whereas dCDT composite scores were built from standard NP tests. However, dCDT performance was not directly used during the review process. Finally, yet importantly, FHS participants were mostly non-Hispanic White and native English speakers. Therefore, the generalizability of our findings to other ethnic populations is unknown.

Aside from these limitations, our study had several strengths. This is the first study to investigate the association between dCDT features and MRI measures in a population-based evaluation. All the NP tests and MRI measures were collected consistently with rigorous quality control. The MCI diagnosis was made by an experienced review panel, which capitalized on all available relevant data sources for consistent diagnosis. Further, given the current relatively widespread use of the paper-pencil clock drawing test as a cognitive screening tool in clinical practice, results from this study do support the potential use of a digital version in place of the paper-pencil version. This application may be of particular use in clinical settings where neuropsychological and neurological expertise is unavailable.

In summary, we assessed the psychometric characteristics of the dCDT with brain volume measures by MRI. In combination with age, sex, education, and MRI measures, the dCDT improved the classification performance of MCI similar to standard NP tests. Our results suggest that the dCDT may be used as a cost-effective and easy-to-administer tool by general practitioners to assess subtle cognitive changes occurring in MCI or early dementia stages and in underdeveloped countries or regions where medical resources are limited.

Acknowledgments

We thank the Framingham Heart Study participants for their decades of dedication and the staff for their hard work in collecting and preparing the data. This work was supported by the Framingham Heart Study's National Heart, Lung, and Blood Institute contract (N01-HC-25195; HHSN268201500001I), and National Institutes of Health grants from the National Institute on Aging (AG008122, AG016495, AG033040, AG054156, AG049810, AG062109, and U01AG068221) and Pfizer. This work was also supported by the Alzheimer's Association Grant (AARG-NTF-20-643020) and American Heart Association Grant (20SFRN35360180). Support for JY was provided by the Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences (2016I2M1004). The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Institutes of Health or the US Department of Health and Human Services.

Authors' Contributions

Concept and design: HL, RA, and JY; Acquisition, analysis, or interpretation of data: HL, RA, JY, CK, AFAA, SD, and SA; Drafting of the manuscript: JY and HL; Statistical analysis: HL; Administrative, technical, or material support: AA; Supervision: HL and RA. All authors critically revised the manuscript for important intellectual content.

Conflicts of Interest

RA is a scientific advisor at Signant Health and a consultant at Biogen. There is no declaration from the other authors.

Multimedia Appendix 1
Raw feature names and descriptions.
[DOCX File , 29 KB - jmir_v24i4e34513_app1.docx]

Multimedia Appendix 2
Association between digital Clock Drawing Test composite scores and cortical gray matter volumes.
[DOCX File , 38 KB - jmir_v24i4e34513_app2.docx]

Multimedia Appendix 3
Association between digital Clock Drawing Test composite scores and total cerebral brain volume after additionally adjusting for vascular risk factors.
[DOCX File , 18 KB - jmir_v24i4e34513_app3.docx]

Multimedia Appendix 4
Association of digital Clock Drawing Test composite scores with cerebral white matter, gray matter, and hippocampal volumes after additionally adjusting for vascular risk factors.
[DOCX File , 21 KB - jmir_v24i4e34513_app4.docx]

Multimedia Appendix 5
Association between digital Clock Drawing Test composite scores and white matter hyperintensity volume after additionally adjusting for vascular risk factors.
[DOCX File , 19 KB - jmir_v24i4e34513_app5.docx]

Multimedia Appendix 6
Association between digital Clock Drawing Test composite scores and cortical gray matter after additionally adjusting for vascular risk factors.
[DOCX File , 24 KB - jmir_v24i4e34513_app6.docx]

Multimedia Appendix 7
Association between individual digital Clock Drawing Test features and magnetic resonance imaging measures.
[DOCX File , 212 KB - jmir_v24i4e34513_app7.docx]

References

1. Ahmed S, Brennan L, Eppig J, Price CC, Lamar M, Delano-Wood L, et al. Visuoconstructional Impairment in Subtypes of Mild Cognitive Impairment. Appl Neuropsychol Adult 2016 Mar;23(1):43-52 [FREE Full text] [doi: 10.1080/23279095.2014.1003067] [Medline: 26397732]
2. Cosentino S, Jefferson A, Chute DL, Kaplan E, Libon DJ. Clock drawing errors in dementia: neuropsychological and neuroanatomical considerations. Cogn Behav Neurol 2004 Jun;17(2):74-84. [doi: 10.1097/01.wnn.0000119564.08162.46] [Medline: 15453515]
3. Libon DJ, Malamut BL, Swenson R, Sands LP, Cloud BS. Further analyses of clock drawings among demented and nondemented older subjects. Arch Clin Neuropsychol 1996;11(3):193-205. [Medline: 14588923]
4. Agrell B, Dehlin O. The clock-drawing test. Age Ageing 1998;27(3):399-403. [doi: 10.1093/ageing/27.3.399]
5. Pinto E, Peters R. Literature review of the Clock Drawing Test as a tool for cognitive screening. Dement Geriatr Cogn Disord 2009 Feb 19;27(3):201-213 [FREE Full text] [doi: 10.1159/000203344] [Medline: 19225234]
6. Libon D, Swenson R, Barnoski E, Sands L. Clock drawing as an assessment tool for dementia. Archives of Clinical Neuropsychology 1993 Sep;8(5):405-415. [doi: 10.1016/0887-6177(93)90004-k]
7. Supasitthumrong T, Herrmann N, Tunvirachaisakul C, Shulman K. Clock drawing and neuroanatomical correlates: A systematic review. Int J Geriatr Psychiatry 2019 Feb;34(2):223-232. [doi: 10.1002/gps.5013] [Medline: 30370637]
8. Matsuoka T, Narumoto J, Shibata K, Okamura A, Nakamura K, Nakamae T, et al. Neural correlates of performance on the different scoring systems of the clock drawing test. Neurosci Lett 2011 Jan 10;487(3):421-425. [doi: 10.1016/j.neulet.2010.10.069] [Medline: 21055445]
9. Thomann PA, Toro P, Dos Santos V, Essig M, Schröder J. Clock drawing performance and brain morphology in mild cognitive impairment and Alzheimer's disease. Brain Cogn 2008 Jun;67(1):88-93. [doi: 10.1016/j.bandc.2007.11.008] [Medline: 18215449]
10. Davis R, Penney D, Pittman D, Libon D, Swenson R, Kaplan E. The Digital Clock Drawing Test (dCDT) - I: Development of a new computerized quantitative system. 2011 Presented at: 39th Annual Meeting of the International Neuropsychological

Society; February 2-5, 2011; Boston, MA URL: http://rationale.csail.mit.edu/clocksketch/Publications/INS2011-dCDTI.pdf
11. Penney D, Davis R, Libon D, Lamar M, Price C, Swenson R, et al. The Digital Clock Drawing Test (dCDT) - II: A new computerized quantitative system. 2011 Presented at: 39th Annual Meeting of the International Neuropsychological Society; February 2-5, 2011; Boston, MA URL: http://rationale.csail.mit.edu/clocksketch/Publications/INS2011-dCDTII.pdf
12. Davis R, Libon D, Au R, Pitman D, Penney D. THink: Inferring Cognitive Status from Subtle Behaviors. AIMag 2015 Sep 28;36(3):49-60. [doi: 10.1609/aimag.v36i3.2602]
13. Souillard-Mandar W, Davis R, Rudin C, Au R, Libon DJ, Swenson R, et al. Learning Classification Models of Cognitive Conditions from Subtle Behaviors in the Digital Clock Drawing Test. Mach Learn 2016 Mar;102(3):393-441 [FREE Full text] [doi: 10.1007/s10994-015-5529-5] [Medline: 27057085]
14. Rentz DM, Papp KV, Mayblyum DV, Sanchez JS, Klein H, Souillard-Mandar W, et al. Association of Digital Clock Drawing With PET Amyloid and Tau Pathology in Normal Older Adults. Neurology 2021 Apr 06;96(14):e1844-e1854 [FREE Full text] [doi: 10.1212/WNL.0000000000011697] [Medline: 33589537]
15. Yuan J, Libon DJ, Karjadi C, Ang AFA, Devine S, Auerbach SH, et al. Association Between the Digital Clock Drawing Test and Neuropsychological Test Performance: Large Community-Based Prospective Cohort (Framingham Heart Study). J Med Internet Res 2021 Jun 08;23(6):e27407 [FREE Full text] [doi: 10.2196/27407] [Medline: 34100766]
16. Seshadri S, Wolf PA, Beiser A, Elias MF, Au R, Kase CS, et al. Stroke risk profile, brain volume, and cognitive function: the Framingham Offspring Study. Neurology 2004 Nov 09;63(9):1591-1599. [doi: 10.1212/01.wnl.0000142968.22691.70] [Medline: 15534241]
17. Debette S, Seshadri S, Beiser A, Au R, Himali JJ, Palumbo C, et al. Midlife vascular risk factor exposure accelerates structural brain aging and cognitive decline. Neurology 2011 Aug 02;77(5):461-468 [FREE Full text] [doi: 10.1212/WNL.0b013e318227b227] [Medline: 21810696]
18. Libon DJ, Preis SR, Beiser AS, Devine S, Seshadri S, Wolf PA, et al. Verbal memory and brain aging: an exploratory analysis of the role of error responses in the Framingham Study. Am J Alzheimers Dis Other Demen 2015 Sep;30(6):622-628 [FREE Full text] [doi: 10.1177/1533317515577184] [Medline: 25788434]
19. Fletcher E, Raman M, Huebner P, Liu A, Mungas D, Carmichael O, et al. Loss of fornix white matter volume as a predictor of cognitive impairment in cognitively normal elderly individuals. JAMA Neurol 2013 Nov 01;70(11):1389-1395 [FREE Full text] [doi: 10.1001/jamaneurol.2013.3263] [Medline: 24018960]
20. Hugenschmidt CE, Peiffer AM, Kraft RA, Casanova R, Deibler AR, Burdette JH, et al. Relating imaging indices of white matter integrity and volume in healthy older adults. Cereb Cortex 2008 Feb;18(2):433-442. [doi: 10.1093/cercor/bhm080] [Medline: 17575289]
21. Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. The Lancet 2014 Mar 15;383(9921):999-1008 [FREE Full text] [doi: 10.1016/S0140-6736(13)61752-3] [Medline: 24084292]
22. Tsao CW, Vasan RS. Cohort Profile: The Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. Int J Epidemiol 2015 Dec;44(6):1800-1813 [FREE Full text] [doi: 10.1093/ije/dyv337] [Medline: 26705418]
23. Wolf PA. Contributions of the Framingham Heart Study to stroke and dementia epidemiologic research at 60 years. Arch Neurol 2012 May;69(5):567-571 [FREE Full text] [doi: 10.1001/archneurol.2011.977] [Medline: 22213410]
24. For Researchers. The Framingham Heart Study. URL: https://framinghamheartstudy.org/fhs-for-researchers/ [accessed 2022-04-02]
25. Au R, Seshadri S, Wolf PA, Elias MF, Elias PK, Sullivan L, et al. New norms for a new generation: cognitive performance in the framingham offspring cohort. Exp Aging Res 2004 Oct;30(4):333-358. [doi: 10.1080/03610730490484380] [Medline: 15371099]
26. Wechsler D, Stone C. Wechsler Memory Scale (WMS). New York, NY: The Psychological Corporation; 1948.
27. Wechsler D. Wechsler Adult Intelligence Scale (WAIS). New York, NY: The Psychological Corporation; 1955.
28. Kaplan E, Goodglass H, Weintraub S, Segal O. Boston Naming Test. Philadephia, PA: Lea & Febiger; 1983.
29. Reitan RM. Validity of the Trail Making Test as an indicator of organic brain damage. PMS 1958 Dec;8(7):271. [doi: 10.2466/pms.8.7.271-276]
30. Hooper H. The Hooper Visual Organization Test: Manual. Los Angeles, CA: Western Psychological Service; 1958.
31. Benton A, Hamsher K, Sivan A. Multilingual Aphasia Examination. Iowa City, IA: University of Iowa; 1976.
32. Goodglass H, Kaplan E. The Assessment of Aphasia and Related Disorders. Philadelphia, PA: Lea & Febiger; 1983.
33. DeCarli C, Massaro J, Harvey D, Hald J, Tullberg M, Au R, et al. Measures of brain morphology and infarction in the framingham heart study: establishing what is normal. Neurobiol Aging 2005 Apr;26(4):491-510. [doi: 10.1016/j.neurobiolaging.2004.05.004] [Medline: 15653178]
34. Sanborn V, Preis SR, Ang A, Devine S, Mez J, DeCarli C, et al. Association Between Leptin, Cognition, and Structural Brain Measures Among "Early" Middle-Aged Adults: Results from the Framingham Heart Study Third Generation Cohort. J Alzheimers Dis 2020;77(3):1279-1289 [FREE Full text] [doi: 10.3233/JAD-191247] [Medline: 32831199]

35. DeCarli C, Fletcher E, Ramey V, Harvey D, Jagust WJ. Anatomical mapping of white matter hyperintensities (WMH): exploring the relationships between periventricular WMH, deep WMH, and total WMH burden. Stroke 2005 Jan;36(1):50-55 [FREE Full text] [doi: 10.1161/01.STR.0000150668.58689.f2] [Medline: 15576652]
36. Au R, Massaro JM, Wolf PA, Young ME, Beiser A, Seshadri S, et al. Association of white matter hyperintensity volume with decreased cognitive functioning: the Framingham Heart Study. Arch Neurol 2006 Feb;63(2):246-250. [doi: 10.1001/archneur.63.2.246] [Medline: 16476813]
37. Seshadri S, Wolf PA, Beiser A, Au R, McNulty K, White R, et al. Lifetime risk of dementia and Alzheimer's disease. The impact of mortality on risk estimates in the Framingham Study. Neurology 1997 Dec;49(6):1498-1504. [doi: 10.1212/wnl.49.6.1498] [Medline: 9409336]
38. Au R, Piers RJ, Devine S. How technology is reshaping cognitive assessment: Lessons from the Framingham Heart Study. Neuropsychology 2017 Nov;31(8):846-861 [FREE Full text] [doi: 10.1037/neu0000411] [Medline: 29376667]
39. Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. Br J Psychiatry 1982 Jun;140:566-572. [doi: 10.1192/bjp.140.6.566] [Medline: 7104545]
40. R: A language and environment for statistical computing. URL: https://www.R-project.org/ [accessed 2022-04-02]
41. Dion C, Arias F, Amini S, Davis R, Penney D, Libon DJ, et al. Cognitive Correlates of Digital Clock Drawing Metrics in Older Adults with and without Mild Cognitive Impairment. J Alzheimers Dis 2020;75(1):73-83 [FREE Full text] [doi: 10.3233/JAD-191089] [Medline: 32250300]
42. Müller S, Herde L, Preische O, Zeller A, Heymann P, Robens S, et al. Diagnostic value of digital clock drawing test in comparison with CERAD neuropsychological battery total score for discrimination of patients in the early course of Alzheimer's disease from healthy individuals. Sci Rep 2019 Mar 05;9(1):3543 [FREE Full text] [doi: 10.1038/s41598-019-40010-0] [Medline: 30837580]
43. Price CC, Jefferson AL, Merino JG, Heilman KM, Libon DJ. Subcortical vascular dementia: integrating neuropsychological and neuroradiologic data. Neurology 2005 Aug 09;65(3):376-382 [FREE Full text] [doi: 10.1212/01.wnl.0000168877.06011.15] [Medline: 16087901]
44. Libon DJ, Price CC, Giovannetti T, Swenson R, Bettcher BM, Heilman KM, et al. Linking MRI hyperintensities with patterns of neuropsychological impairment: evidence for a threshold effect. Stroke 2008 Mar;39(3):806-813. [doi: 10.1161/STROKEAHA.107.489997] [Medline: 18258842]
45. Filley CM, Fields RD. White matter and cognition: making the connection. J Neurophysiol 2016 Nov 01;116(5):2093-2104 [FREE Full text] [doi: 10.1152/jn.00221.2016] [Medline: 27512019]
46. Fields RD. White matter in learning, cognition and psychiatric disorders. Trends Neurosci 2008 Jul;31(7):361-370 [FREE Full text] [doi: 10.1016/j.tins.2008.04.001] [Medline: 18538868]

Abbreviations

AUC: area under the curve
CDR: Clinical Dementia Rating
CDT: Clock Drawing Test
dCDT: digital Clock Drawing Test
FHS: Framingham Heart Study
MCI: mild cognitive impairment
MRI: magnetic resonance imaging
NC: normal cognition
NP: neuropsychological
TCBV: total cerebral brain volume
TCV: total cerebral cranial volume
WMH: white matter hyperintensity

Edited by R Kukafka; submitted 28.10.21; peer-reviewed by MJ Brouillette, I Tarnanas; comments to author 20.12.21; revised version received 08.02.22; accepted 13.03.22; published 15.04.22.

Please cite as:
Yuan J, Au R, Karjadi C, Ang TF, Devine S, Auerbach S, DeCarli C, Libon DJ, Mez J, Lin H
Associations Between the Digital Clock Drawing Test and Brain Volume: Large Community-Based Prospective Cohort (Framingham Heart Study)
J Med Internet Res 2022;24(4):e34513
URL: https://www.jmir.org/2022/4/e34513
doi:10.2196/34513
PMID:35436225

Review

Methods to Establish Race or Ethnicity of Twitter Users: Scoping Review

Su Golder[1], PhD; Robin Stevens[2], PhD; Karen O'Connor[3], MSc; Richard James[4], MSLIS; Graciela Gonzalez-Hernandez[3], PhD

[1]Department of Health Sciences, University of York, York, United Kingdom

[2]School of Communication and Journalism, University of Southern California, Los Angeles, CA, United States

[3]Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[4]School of Nursing Liaison and Clinical Outreach Coordinator, University of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:
Su Golder, PhD
Department of Health Sciences
University of York
Heslington
York, YO10 5DD
United Kingdom
Phone: 44 01904321904
Email: su.golder@york.ac.uk

Abstract

Background: A growing amount of health research uses social media data. Those critical of social media research often cite that it may be unrepresentative of the population; however, the suitability of social media data in digital epidemiology is more nuanced. Identifying the demographics of social media users can help establish representativeness.

Objective: This study aims to identify the different approaches or combination of approaches to extract race or ethnicity from social media and report on the challenges of using these methods.

Methods: We present a scoping review to identify methods used to extract the race or ethnicity of Twitter users from Twitter data sets. We searched 17 electronic databases from the date of inception to May 15, 2021, and carried out reference checking and hand searching to identify relevant studies. Sifting of each record was performed independently by at least two researchers, with any disagreement discussed. Studies were required to extract the race or ethnicity of Twitter users using either manual or computational methods or a combination of both.

Results: Of the 1249 records sifted, we identified 67 (5.36%) that met our inclusion criteria. Most studies (51/67, 76%) have focused on US-based users and English language tweets (52/67, 78%). A range of data was used, including Twitter profile metadata, such as names, pictures, information from bios (including self-declarations), or location or content of the tweets. A range of methodologies was used, including manual inference, linkage to census data, commercial software, language or dialect recognition, or machine learning or natural language processing. However, not all studies have evaluated these methods. Those that evaluated these methods found accuracy to vary from 45% to 93% with significantly lower accuracy in identifying categories of people of color. The inference of race or ethnicity raises important ethical questions, which can be exacerbated by the data and methods used. The comparative accuracies of the different methods are also largely unknown.

Conclusions: There is no standard accepted approach or current guidelines for extracting or inferring the race or ethnicity of Twitter users. Social media researchers must carefully interpret race or ethnicity and not overpromise what can be achieved, as even manual screening is a subjective, imperfect method. Future research should establish the accuracy of methods to inform evidence-based best practice guidelines for social media researchers and be guided by concerns of equity and social justice.

(J Med Internet Res 2022;24(4):e35788) doi:10.2196/35788

KEYWORDS

twitter; social media; race; ethnicity

Introduction

Research Using Twitter Data

Twitter data are increasingly being used as a surveillance and data collection tool in health research. When millions of users post on Twitter, it translates into a vast amount of publicly accessible, timely data about a variety of attitudes, behaviors, and preferences in a given population. Although these data were not originally intended as a repository of individual information, Twitter data have been retrofitted in infodemiology to investigate population-level health trends [1-15]. Researchers often use Twitter data in consort with other sources to test the relationship between web-based discourse and offline health behavior, public opinion, and disease incidence.

The appeal of Twitter data is clear. Twitter is one of the largest public-facing social media platforms, with an ethnically diverse user base [16,17] of more than 68 million US Twitter users, with Black users accounting for 26% of that base [18]. This diverse user base gives researchers access to people they may have difficulty reaching using more traditional approaches [19]. However, promising insights that can be derived from Twitter data are often limited by what is missing, specifically the basic sociodemographic information of each Twitter user. The demographic attributes of users are often required in health research for subpopulation analyses, to explore differences, and to identify inequity. Without evidence of the distal and proximal factors that lead to racial and ethnic health disparities, it is impossible to address and correct these drivers. Insights from social media data can be used to inform service provision as well as to develop targeted health messaging by understanding public perspectives from diverse populations.

Extracting Demographics From Twitter

However, to use social media and digital health research to address disparities, we need to know not only what is said on Twitter but also who is saying what [20]. Although others have discussed extracting or estimating features, such as location, age, gender, language, occupation, and class, no comprehensive review of the methods used to extract race or ethnicity has been conducted [20]. Extracting the race and ethnicity of Twitter users is particularly important for identifying trends, experiences, and attitudes of racially and ethnically diverse populations [21]. As race is a social construction and not a genetic categorization [22,23], the practice of defining race and ethnicity in health research has been an ongoing, evolving challenge. Traditional research has the advantage of identifying the person in the study and allowing them to systematically identify their racial and ethnic identities. In digital health research [22,23], determining a user's race or ethnicity by extracting data from a user's Twitter profile, metadata, or tweets is a process that is inevitably challenging, complex, and not without ethical questions.

Furthermore, although Twitter is used for international research, an international comparative study of methods to determine race or ethnicity is difficult, practically impossible, given that societies use different standardized categories that describe their own populations [24]. A common approach in the United States is based on the US Census Bureau practice to allow participants to identify with as many as 5-6 large racial groupings (Black, White, Asian, Pacific Islander, Native, and other), while separately choosing one ethnicity (Hispanic) [25]. However, race and ethnicity variables continue to be misused in the study design or when drawing conclusions. For example, race or ethnicity is often incorrectly treated as a predictor of poor health rather than as a proxy for the impact of being a particular race or ethnicity has on that person's experience with the health system [26]. Simply put, health disparities are driven by racism, not race [27-29]. Although race or ethnicity affiliation is an important factor in understanding diverse populations, digital research must tread lightly and thoughtfully both the collection and assignment of race or ethnicity.

Objectives

The lack of basic sociodemographic data on Twitter users has led researchers to apply a variety of approaches to better understand the characteristics of the people behind each tweet. The breadth of the landscape of approaches to extracting race or ethnicity is currently unknown. Our overall aim was to summarize and assess the range of computational and manual methods used in research based on Twitter data to determine the race or ethnicity of Twitter users.

Methods

Overview

We conducted a comprehensive scoping review of extraction methods and offered recommendations and cautions related to these approaches [30]. We selected Twitter, as it is currently the most commonly used social media platform in health care research, and it has some unique intrinsic characteristics that drive the methods used for mining it. Thus, we felt that the methods, type of data, and social media platforms used are related in such a way that comparing methods for different social media would add too many variables and would not be truly comparing like with like. A detailed protocol was designed for the methods to be used in our scoping review, but we were unable to register scoping reviews on PROSPERO. We report our methods according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) scoping review statement [30].

Inclusion Criteria

Overview

We devised strict inclusion criteria for our review based on the Population, Intervention, Comparators, Outcomes, and Study design format. Although this was not a review of effectiveness, we felt that the Population, Intervention, Comparators, Outcomes, and Study design question breakdown [31] was still the most appropriate one available for our question format [31]. The inclusion criteria are described in the following sections.

Population

We included only data sets of Twitter users. Studies were eligible for inclusion if they collected information to extract or infer race or ethnicity directly from the users' tweets, their profile details (such as the users' photo or avatar, their name, location, and biography [bio]), or their followers. We excluded

studies that extracted race or ethnicity from social media platforms other than Twitter, from unspecified social media platforms, or those that used multiple social media platforms that included Twitter, but the data relating to Twitter were not presented separately.

Intervention

Studies were included where the methods to extract or infer the race or ethnicity data of Twitter users were stated. Articles that used machine learning (ML), natural language processing (NLP), human-in-the-loop, or other computationally assisted methods to predict race or ethnicity of users were included, as were manual or noncomputational methods, including photo recognition or linking to census data. We excluded studies for which we were unable to determine the methods used or for which we extracted data solely on other demographic characteristics, such as age, gender, or geographic location.

Comparator

The use of a comparison of the methods used was not required. A method could be compared with another (such as a gold standard), or no comparison could be undertaken.

Outcome

The extraction or inference of the race or ethnicity of Twitter users was the primary or secondary outcome of the study. As this was a scoping review in which we aimed to demonstrate the full landscape of the literature, no particular measurement of the performance of the method used was required in our included studies.

Study Design

Any type of research study design was considered relevant. Discussion papers, commentaries, and letters were excluded.

Limits

No restrictions on date, language, or publication type were applied to the inclusion criteria. However, no potentially relevant studies were identified in any non-English language, and the period by default was since 2006, the year of the inception of Twitter.

Search Strategy

A database search strategy was derived by combining three facets: facet 1 consisted of free-text terms related to Twitter (*Twitter* OR *Tweet** OR *Tweeting* OR *Retweet** OR *Tweep**); facet 2 consisted of terms for race or ethnicity; and facet 3 consisted of terms for methods of prediction, such as ML, NLP, and artificial intelligence–related terms (Table S1 in Multimedia Appendix 1 [3,10,12,18,20,21,32-96]). All ethnology-related subject terms were adapted for different database taxonomies and syntax, with standard methods for predicting subject terms in MEDLINE and other database indexing. The methods of predicting term facets were expanded using a comprehensive list of specific text analysis tools and software names extracted from the study by Hinds and Joinson [97], which included a comprehensive list of automated ML processes used in predicting demographic markers in social media. Additional terms have been added from a related study [98].

Sources Searched

A wide range of bibliographic and gray literature databases were selected to search for topics on computer science, health, and social sciences. The databases (Table 1) were last searched on May 15, 2021, with no date or other filter applied.

Reference checking of all included studies and any related systematic reviews identified by the searches were conducted. We browsed the Journal of Medical Internet Research, as this is a key journal in this field, and hand searched 2 relevant conferences, the International Conference on Weblogs and Social Media and Association for Computational Linguistics proceedings.

Citations were exported to a shared Endnote library, and duplicates were removed. The deduplicated records were then imported into Rayyan to facilitate independent blinded screening by the authors. Using the inclusion criteria, at least two screeners (SG, RS, KO, or RJ) from the research team independently screened each record, with disputes on inclusion discussed and a consensus decision reached.

Only the first 50 records from ACL and the first 100 records from a Google Scholar search were screened during two searches (March 11, 2020, and May 24, 2021) as these records are displayed in order of relevance, and it was felt that after this number no relevant studies were being identified [12,21,32-95,99].

Table 1. Databases searched with number of records retrieved.

Database	Total results, n
ACL Anthology	Screened first 50 records from 2 searches
ACM Digital Library	150
CINAHL	200
Conference Proceedings Citation Index—Science	84
Conference Proceedings Citation Index—Social Science	7
Emerging Sources Citation Index	41
Google Scholar	Screened first 100 records from 2 searches
IEEE Xplore	186
Library and Information Science Abstracts	120
LISTA	79
OpenGrey	0
ProQuest dissertations and theses—United Kingdom and Ireland	195
PsycINFO	72
PubMed	84
Science Citation Index	56
Social Science Citation Index	111
Zetoc	50

Data Extraction

For each included study, we extracted the following data on an excel spreadsheet:

year of publication, study country and language, race or ethnicity categories extracted (such as for race—Black, White, or Asian or for ethnicity—Hispanic or European), and paper type (journal, conference, or thesis). We also extracted details on extraction methods (such as classification models or software used), features and predictors used in extraction (tweets, profiles, and pictures), number of Twitter users, number of tweets or images used, performance measures to evaluate methods used (validation), and results of any evaluation (such as accuracy). All performance measure metrics were reported as stated in the included studies. All the extracted data were checked by 2 reviewers.

Quality Assessment

There was no formally approved quality assessment tool for this type of study. As this was a scoping review, we did not carry out any formal assessment. However, we assessed any validation performed and whether the methods were reproducible.

Data Analysis

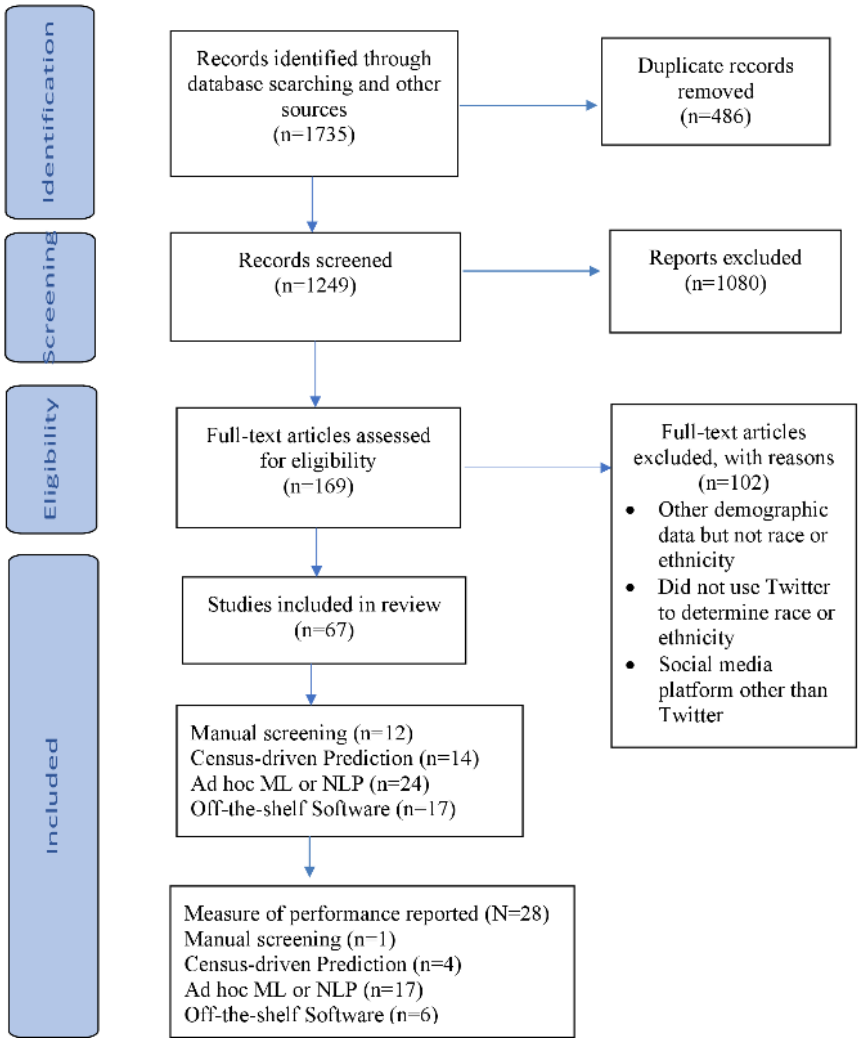
We have summarized the stated performance of the papers that included validation. However, we could not compare approaches using the stated performance, as the performance measures and validation approaches varied considerably. In addition, there is no recognized gold standard data set for comparison.

Results

Overview

A total of 1735 records were entered into an Endnote library (Clarivate), and duplicates were removed, leaving 1249 (72%) records for sifting (Figure 1). A total of 1080 records were excluded based on the title and abstract screening alone. A total of 169 references were deemed potentially relevant by one of the independent sifters (RS, GG, RJ, SG, and KO). The full text of these articles was screened independently, and 67 studies [12,21,32-95,99] met our inclusion criteria and 102 references were excluded [77,97,100-198]. The main reason for exclusion was that although the abstract indicated that demographic data were collected, it did not include race or ethnicity (most commonly, other demographic attributes such as gender, age, or location were collected). Other reasons for exclusion were that the researchers collected demographic data through surveys or questionnaires administered via Twitter (but not from data posted on Twitter) or that the researchers used a social media platform other than Twitter.

Figure 1. Flow diagram for included studies.



Characteristics of the Included Studies

Most of the studies (51/67, 76%) stated or implied that they were based solely or predominantly in the United States and were limited to English language bios or tweets. A total of 6 studies were multinational [38,41,56,66,83,86]; 1 was UK based (also in English) [59], another was based in Qatar [55], and 12% (8/67) of studies extracted data from tweets in multiple languages [32,38,52,55,56,66,83,86] (Table S2 in Multimedia Appendix 1).

The most common race examined was White (58/67, 87%), followed by Black or African American (56/67, 84%), Asian (45/67, 67%), and the most common ethnicity examined was Hispanic/Latino (43/67, 64%).

Some studies (12/67, 18%) treated race as a binary classification, such as African American or not or African American or White, whereas others created a multiclass classifier of 3 (15/67, 22%) or 4 classes (33/67, 49%) or a combination of classes. A total of 6 studies identified >4 classes; however, these often included ethnicity or nationality classifiers as well as race [38,48,54,66,83,95]. Wang and Chi [77] was a conference paper which did not report the race types extracted.

The data objects from Twitter used to extract race or ethnicity varied, with the use of profile pictures or Twitter users' names being the most common. Others have also used tweets in the users' timeline, information from Twitter bios, or Twitter users' locations. Most studies (39/67, 58%) used more than one data object from Twitter data. In addition, the data sets within the studies varied in size between 392 and 168,000,000, with those using manual methods having smaller data sets ranging from just 392 [50] to 4900 [65].

Unfortunately, although performance has been measured in 67% (45/67) of studies (this was inconsistently measured Table 2). The metrics used to report results were particularly varied for studies using ML or NLP and included the F_1 score (which combines precision and recall), accuracy, area under the curve, or mean average precision. Table 2 lists the methods, features, and reported performance of the top model from each study.

Table 2. Top system performance within studies using machine learning or natural language processing (result metrics are reflected here as reported in the original publications).

Study	Classifier	ML[a] model	Features	Results reported		
				Accuracy	F_1 score	Area under curve
Pennacchiotti and Popescu, 2011 [68]	Binary	GBDT[b]	Images, text, topics, and sentiment	N/A[c]	0.66	N/A
Pennacchiotti and Popescu, 2011 [67]	Binary	GBDT	Images, text, topics, sentiment, and network	N/A	0.70	N/A
Bergsma et al, 2013 [38]	Binary	SVM[d]	Names and name clusters	0.85	N/A	N/A
Ardehaly and Culotta, 2017 [35]	Binary	DLLP[e]	Text and images	N/A	0.95 (image); 0.92 (text)	N/A
Volkova and Backrach, 2018 [76]	Binary	LR[f]	Text, sentiment, and emotion	N/A	N/A	0.97
Wood-Doughtry et al, 2018 [79]	Binary	CNN[g]	Name	0.73	0.72	N/A
Saravanan, 2017 [72]	Ternary	CNN	Text	NR[h]	NR	NR
Ardehaly and Culotta, 2017 [33]	Ternary	DLLP	Text and images	N/A	0.84 (image); 0.83 (text)	N/A
Gunarathne et al, 2019 [94]	Ternary	CNN	Text	N/A	0.88	N/A
Wood-Doughtry et al, 2018 [79]	Ternary	CNN	Name	0.62	0.43	N/A
Culotta et al, 2016 [47]	Quaternary	Regression	Network and text	N/A	0.86	N/A
Chen et al, 2015 [46]	Quaternary	SVM	n-grams, topics, self-declarations, and image	0.79	0.79	0.72
Markson, 2017 [61]	Quaternary	CNN	Synonym expansion and topics	0.76	N/A	N/A
Wang et al, 2016 [189]	Quaternary	CNN	Images	0.84	N/A	N/A
Xu et al, 2016 [82]	Quaternary	SVM	Synonym expansion and topics	0.76	N/A	N/A
Ardehaly and Culotta, 2015 [34]	Quaternary	Multinomial logistic regression	Census, name, network, and tweet language	0.83	N/A	N/A
Ardehaly, 2014 [64]	Quaternary	LR	Census and image tweets	0.82	0.81	N/A
Barbera, 2016 [37]	Quaternary	LR with EN[i]	Tweets, emojis, and network	0.81	N/A	N/A
Wood-Doughty 2020 [81]	Quaternary	CNN	Name, profile metadata, and text	0.83	0.46	N/A
Preotiuc-Pietro and Ungar, 2018 [96]	Quaternary	LR with EN	Text, topics, sentiment, part-of-speech tagging, name, perceived race labels, and ensemble	N/A	N/A	0.88 (African American), 0.78 (Latino), 0.83 (Asian), and 0.83 (White)
Mueller et al, 2021 [91]	Quaternary	CNN	Text and accounts followed	N/A	0.25 (Asian), 0.63 (African American or Black), 0.28 (Hispanic), and 0.90 (White)	N/A
Bergsma et al, 2013 [38]	Multinomial (>4)	SVM	Name and name clusters	0.81	N/A	N/A
Nguyen et al, 2018 [66]	Multinomial (>4)	Neural network	Images	0.53	N/A	N/A

[a]ML: machine learning.
[b]GBDT: gradient-boosted decision tree.
[c]N/A: not applicable.
[d]SVM: support vector machine.
[e]DLLP: deep learning from label proportions.
[f]LR: logistic regression.
[g]CNN: convolutional neural network.
[h]NR: not reported.
[i]EN: elastic net.

Manual Screening

A total of 12 studies used manual techniques to classify Twitter users into race or ethnicity categories [21,36,40,49-51,57,65,87-90]. These studies generally combined qualitative interpretations of recent tweets, information in user bios making an affirmation of racial or ethnic identity, or photographs or images in the user timeline or profile.

In most cases, tweets were first identified by text matching based on terms of interest in the research topic, such as having a baby with a birth defect [50], commenting on a controversial topic [57,89], or using potentially gang- or drug-related language [40]. Researchers then identified the tweet authors and, in most cases, assigned race or ethnicity through hand coding based on profile and timeline content. Some studies coded primarily based on self-identifying statements of race used in a tweet or in users' bios, such as people stating that they are a *Black American* [49,50,88,90] or hashtags [36] (such as #BlackScientist). Others coded exclusively based on the research team's attribution of racial identity through the examination of profile photographs [21,57] or avatar [87]. Some authors coded primarily with self-declarations, with secondary indicators, such as profile pictures, language, usernames, or other content [40,51,65,88,89]. In most cases, it appears reasonable to infer that coding was performed by the study authors or members of their research teams, with the exception of those using the crowdsourcing marketplace, Amazon Mechanical Turk [21,90].

The agreement among coders was sometimes measured, but validity and accuracy measurements were not generally included. A study [65], however, documented 78% reliability for coding race compared with census demographics, with Black and White users being coded accurately 90% of the time and Hispanic or Asian users being accurately coded between 45% and 60% of the time. The high accuracy of Black users was based on the higher likelihood of Black users to self-identify.

Census-Driven Prediction

Another approach to predict race or ethnicity is to use demographic information from the national census and census-like data and transfer it to the social media cohort. The US-based studies largely used census-based race and ethnicity categories: Asian and Pacific Islander, Black or African American, Latino or Hispanic, Native American, and White. A UK-based study included the categories British and Irish, West European, East European, Greek or Turkish, Southeast Asian, other Asian, African and Caribbean, Jewish, Chinese, and other minorities [83].

We identified 14 studies [39,48,52,54,60,63,70,71,74,77,83-85,95] that used census geographic data, census surname classification, or a combination of both. A total of 6 studies incorporated geographic census data [39,52,63,74,83,84]. For example, Blodgett et al [39] created a simple probabilistic model to infer a user's ethnicity by matching geotagged tweets with census block information. They averaged the demographic values of all tweets by the user and assumed this to be a rough proxy for the user's demographics. Stewart [74] collected tweets tagged with geolocation information (longitude and latitude). The ZIP code of the user was derived from this geolocation information and matched with the demographic information found in the ZIP Code Tabulation Area defined by the Census Bureau. This information was used to find a correlation between ethnicity and African American vernacular English syntax [74].

Other studies have used the census-derived name classification system to determine race or ethnicity based on user names. We identified 12 studies that predicted user race or ethnicity using surnames [48,54,60,63,70,71,77,83-85,95,189]. Surnames were used to assign race or ethnicity using either a US census-based name classification system or, less commonly, an author in-house generated classification system. Of these 12 studies, 7 (58%) relied solely on the user's last names [48,54,60,63,70,71,85]. Of those that reported validating the system, validation methods of this name-based system alone were not reported, but 4 (33%) of the 12 studies reported an accuracy between 71.8% and 81.25% [63,70,71,83]. Of note, a study reported vastly different accuracies in predicting whiteness versus blackness (94% predicting White users vs 33% predicting African American or Black users) [83]. The remaining 2 studies augmented name-based predictions with aggregate demographic data from the American Community Survey or equivalent surveys. For example, statistical and text mining methods have been used to extract surnames from Twitter profiles, combining this information with census block information based on geolocated tweets to assess the probability of the user's race or ethnicity [60]. However, these studies did not report validation or accuracy.

Ad Hoc ML or NLP

A total of 24 papers [33-35,37,38,46,47,61,64,66-68,72,76,78-82,91-94,99] used ML or NLP to automatically classify users based on their race or ethnicity. ML and NLP methods were used to process the data made available by Twitter users, such as profile images, tweets, and location of residence. These studies almost invariably consisted of larger cohorts, with considerable variation in the specific methods used.

Supervised ML models (in which some annotated data were used to *train* the system) were used in 12 (50%) of the 24 studies. The models used include support vector machine [38,46,61], gradient-boosted decision trees [67,68], and regression models [33,34,37,76,96].

Semisupervised (where a large set of unannotated data is also used for training the system, in addition to annotated data) or fully unsupervised models using neural networks or regression were used for classification in 10 (42%) of the 24 studies [33,35,66,72,78,79,81,92-94].

A total of 2 studies used an ensemble of previously published race or ethnicity classifiers by processing the data through 4 extant models and using a majority rule approach to classify users based on the output of each classifier [80,91].

ML models use features or data inputs to predict desired outputs. Features derived from textual information in the user's profile description, such as name or location, have been used in some studies [34,35,38,60,67,68,79,81,92,93]. Other studies included features related to images, including but not exclusively profile images [46,67,68,189], and facial features in those images [66]. Some studies have used linguistic features to classify a user's race or ethnicity [37,38,46,47,61,67,68,72,76,78,81,92-94,96]. Specific linguistic features used in the models include n-grams [38,46,72,91-94], topic modeling [46,61,78], sentiment and emotion [76], and self-reports [67,68,81]. Information about a user's followers or network of friends was included as a feature in some studies under the assumption that members of these networks have similar traits [34,37,46,47,91].

Labeled data sets are used to train and test supervised and semisupervised ML models and to validate the output of unsupervised learning methods. Some of the studies used previously created data sets that contained demographic information, such as the MORPH longitudinal face database of images [189], a database of mugshots [38], or manually annotated data from previous studies [79,81]. Others created ground truth data sets from surveys [96] or by semiautomatic means, such as matching Twitter users to voter registrations [37], using extracted self-identification from user profiles or tweets [67,68,81], or using celebrities with known ethnicities [66]. Manual annotation of Twitter users was also used based on profile metadata [34,35,46,76], self-declarations in the timeline [61,82], or user images [35,94]. Table 2 summarizes the best performing ML approach, features used, and the reported results for each study that used automatic classification methods. In the table, the classifier is the number of race or ethnicity classification groups, ML model is the top performing algorithm reported, and features are the variables used in the predictions.

Data from Twitter are inherently imbalanced in terms of race and ethnicity. In ML, it is important to attempt to mitigate the effects of the imbalance, as the models have difficulty learning from a few examples and will tend to classify to the majority class and ignore the minority class. Few studies (12/67, 18%) have directly addressed this imbalance. Some opted to make the task binary, focusing only on their group of interest versus all others [67,68,94] or only on the majority classes [38,76]. Others choose modified performance metrics that account for imbalance when reporting their results [33,61,82]. A group, which was classified based on images, supplemented their training set from an additional data source for the minority classes [33,35]. Only 2 studies have experimented with comparator models trained on balanced data sets. In a study by Wood-Doughty et al [81], the majority class was undersampled in their training sets and [96] the minority classes were oversampled. In both cases, the overall performance of the models decreased in accuracy from 0.83 to 0.41 (on their best performing unbalanced model) and 0.84 to 0.68. [96], as the performance boost from the models, the superior performance on the majority class was eradicated.

Off-the-shelf Software

A total of 17 studies [12,32,41-45,53,55,56,58,59,62,69,73,75,86] used off-the-shelf software packages to derive race or ethnicity. Moreover, 10 studies [32,44,45,53,55,56,58,62,69,75] used Face++ [199], 5 studies [12,41-43,73] used Demographics Pro [200], and 2 studies used Onomap [201] software to determine ethnicity [59,86]. Face++ is a validated ML face detection service that analyzes features with confidence levels for inferred race attributes. Specifically, it uses deep learning to identify whether profile pictures contain a single face and then the race of the face (limited to Asian, Black, and White) and does not infer ethnicity (eg, Hispanic) [199]. Demographics Pro estimates the demographic characteristics based on Twitter behavior or use using NLP, entity identification, image analyses, and network theory [200]. Onomap is a software tool used for classifying names [201]. A total of 3 studies that used Face++ used the same baseline data set [45,62,75], and one used a partial subset of the same data set [69].

In total, 2 studies that used Face++ [32,58] did not measure its performance. Another study [44] stated that Face++ could identify race with 99% confidence or higher for 9% of total users. In addition, 2 studies [53,55] used Face++ along with other methods. One of these studies used Face++ in conjunction with demographics, using a given name or full name from a database that contains US census data for demographics. This study simply measured the percentage of Twitter users for which race data could be extracted (46% college students and 92% role models) but did not measure the performance of Face++ [53]. Another study [55] built a classifier model on top of using Face++ and recorded an accuracy of 83.8% when compared with users who stated their nationality.

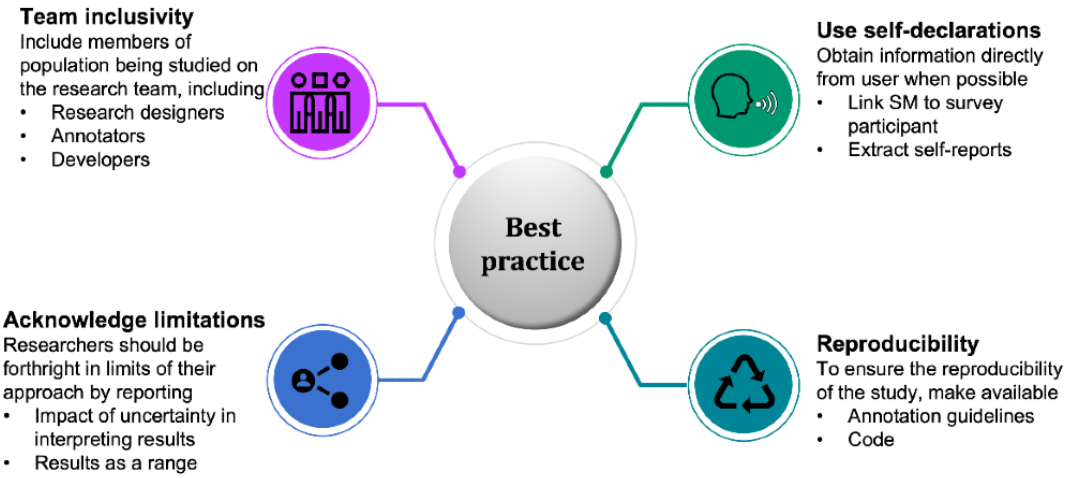
A total of 4 studies [45,62,69,75] (with the same data set in full or in part) used the average confidence level reported by Face++ for race which was 85.97 (SD 0.024%), 85.99 (SD 0.03%), 86.12 (SD 0.032%), respectively, with a CI of 95%. When one of these studies [45] carried out its own accuracy assessment, they found an accuracy score of 79% for race when compared with 100 manually annotated pictures. Huang et al [56] also carried out an accuracy assessment and found that Face++ achieved an averaged accuracy score of 88.4% for race when compared with 250 manually annotated pictures.

A total of 5 studies [12,41-43,73] used Demographics Pro, and although they reported on Demographics Pro success in general, they did not directly report any metrics of its success. The 2

studies using Onomap provided no validation of the software [59,86].

In light of our results, we have compiled our recommendations for best practice, which are summarized in Figure 2 and further examined in the Discussion section.

Figure 2. Summary of our best practice recommendations.



Discussion

Principal Findings

As there are no currently published guidelines or even best practice guidance, it is no surprise that researchers have used a variety of methods for estimating the race or ethnicity of Twitter users. We identified four categories for the methods used: manual screening, census-based prediction, ad hoc ML or NLP, and off-the-shelf software. All these methods exhibit particular strengths, as well as inherent biases and limitations.

Comparing the validity of methods for the purpose of deriving race or ethnicity is difficult as classification models differ not only in approach but also in the definition of the classification of race or ethnicity itself [112,202,203]. There is also a distinct lack of evaluation or validation of the methods used. Those that measured the performance of the methods used found accuracy to vary from 45% to 93%, with significantly lower accuracy in identifying categories of people of color.

This review sheds little light on the performance of commercial software packages. Previous empirical comparisons of facial recognition application programming interfaces have found that Face++ achieves 93% accuracy [204] and works comparatively better for men with lighter skins [205]. The studies included in our review suggested a lower accuracy. However, data on accuracy were not forthcoming in any of the included studies using Demographics Pro [200]. Even when performance is assessed, the methodology used may be biased if there are issues with the *gold standard* used to train the model.

In addition to the 4 overarching methods used, the studies varied in terms of the features used to determine or define race or ethnicity. Furthermore, the reliability of the features used to determine or define race or ethnicity for this purpose is questionable. Specifically, the use of Twitter users' profile pictures, names, and locations, the use of unvalidated linguistic features attributed to racial groups (such as slang words, African American vernacular English, Spanglish, or Multicultural London English), and the use of training data that are prone to perpetuate biases (eg, police booking photos or mug shots) were all of particular concern.

Issues Related to the Methods Used

Approaches that include or rely solely on profile pictures to determine race or ethnicity can introduce bias. First, not all users have a photograph as their profile picture, nor is it easy to determine whether the picture used is that of the user. A study on the feasibility of using Face++ found that only 30.8% of Twitter users had a detectable single face in their profile. A manual review of automatically detected faces determined that 80% could potentially be of the user (ie, not a celebrity) [206]. Human annotation may introduce additional bias, and studies have found systematic biases in the classification of people into racial or ethnic groups based on photographs [207,208]. Furthermore, humans tend to perceive their own race more readily than others [209,210]. Thus, race or ethnicity in the annotation team has an impact on the accuracy of their race or ethnicity labels, potentially skewing the sample labels toward the race or ethnicity of the annotators [211,212]. Given ML and NLP methods are trained on these data sets, the human biases transfer to automated methods, leading to poorly supervised ML and training, which has been shown to result in discrimination by the algorithm [213-215]. These concerns did not appear to be interrogated by the study designers. Without exception, they present categorization of persons into race or ethnicity, assuming that a subjective reading of facial features or idiomatic speech is the gold standard both for coding of race or ethnicity and for training and evaluation of automated methods.

Other methods, such as using geography or names as indicators of race, may also be unreliable. One could argue that the demographic profile for a geographic region is a better representation of race or ethnicity in the demographic environment than an individual's race or ethnicity. Problems in using postcodes or locations to decipher individual social determinants are well documented [216]. The use of census data

from an area that is too large may skew the results. Among the studies reviewed, some used census block data, which are granular, whereas others extrapolated from larger areas, such as city- or county-level data. For example, Saravanan [72] inferred the demographics of users in a city as a certain ethnic group based on a city with a large population of that group; however, no fine-grained analysis was performed either for the city chosen or for geolocation of the Twitter user. Thus, the validity of their assumption that a user in Los Angeles County is of Mexican descent [72] is questionable. As these data were then used to create a *race or ethnicity* dictionary of terms used by that group to train their model, the questionable assumption further taints downstream applications and results. The models also do not consider the differences between the demographics of Twitter users and the general demographics of the population.

In addition, census demographic data that uses names are also questionable because of name-taking in marriage and indiscernible names.

The practice of using a Twitter user's self-reported race or ethnicity would provide a label with high confidence but restrict the amount of usable data and introduce a margin of error depending on the method used to extract such self-reports. For example, in a sample of 14 million users, >0.1% matched precise regular expressions created to detect self-reported race or ethnic identity [128]. Another study used mentions of keywords related to race or ethnicity in a user's bio; however, limited validation was conducted to ensure that the mention was actually related to the user's race or ethnicity [67,68]. This lack of information gathered from the profile information leads to sampling bias in the training of the models [152].

Some models trained on manually annotated data did not have high interannotator agreement; for example, Chen et al [46] crowdsourced annotation agreement measured at 0.45. This can be interpreted as weak agreement, with the percentage of reliable data being 15% to 35% [217]. Training a model on such weakly labeled data produces uncertain results.

It is not possible to assume the accuracy of black box proprietary tools and algorithms. The only race or ethnicity measure that seems empirically reliable is self-report, but this has considerable limitations. Thus, faulty methods continue to underpin digital health research, and researchers are likely to become increasingly dependent on them. The *gold standard* data required to know the demographic characteristics of the Twitter user is difficult to ascertain.

The methods that we highlight as best practices include directly asking the Twitter users. This can be achieved, for example, by asking respondents of a traditional survey for both their demographic data and their Twitter handles so that the data can be linked [96]. This was undertaken in the NatCen Social Research British Social Attitudes Survey 2015, which has the added benefit of allowing the study of the accuracy of further methods for deriving demographic data [20]. Contacting Twitter users may also provide a gold standard but is impractical, given the current terms of use of Twitter that might consider such contact a form of spamming [72,204,205,216]. A limitation of extracting race or ethnicity from social media is the necessity to oversimplify the complexity of racial identity. The categories were often limited to Black, White, Hispanic, or Asian. Note that *Hispanic* is considered *ethnicity* by the US census, but most studies in ML used it as a *race* category, more so than Asian (because of low numbers in this category). Multiple racial identities exist, particularly from an international perspective, which overlooks multiracial or primary and secondary identities. In addition, inferred identities may differ from self-identity, raising further issues.

Given the sensitive nature of the data, it is important as a *best practice* for the results of studies that derive race or ethnicity from Twitter data to be reproducible for validation and future use. The reproducibility of most of the studies in this review would be difficult or impossible, as only 5 studies were linked to available code or data [38,47,79,81,108]. Furthermore, there is limited information regarding the coding of the training data. None of the studies detailed their annotation schemas or made available annotation guidelines. Detailed guidelines as a *best practice* may allow recreation or extension of data sets in situations where the original data may not be shared or where there is data loss over time. This is particularly true of data collected from Twitter, where the terms of use require that shared data sets consist of only tweet IDs, not tweets, and that best efforts to delete IDs from the data set if the original tweet is removed or made private by the user be in place. Additional restrictions are placed on special use cases for sensitive information, prohibiting the storage of such sensitive information if detected or inferred from the user. Twitter explicitly states that information on racial or ethnic origin cannot be *derived or inferred* for an individual Twitter user and allows academic research studies to use only aggregate-level data for analysis [218]. It may be argued that this policy is more likely to be targeted at commercial activities.

Strengths and Limitations

We did not limit our database searches and other methods by study design; however, we were unable to identify any previous reviews on the subject. To the best of our knowledge, this is the first review of methods used to extract race or ethnicity from social media. We identified studies from a range of disciplines and sources and categorized and summarized the methods used. However, we were unable to obtain information on the methodologies used by private-sector companies that created software for this purpose. Marketing and targeted advertising are common on social media and are likely to use race as a part of their algorithms to derive target users.

We did not limit our included papers to those in which the extraction of race or ethnicity was the primary focus. Although this can be conceived as a strength, it also meant that reporting of the methods used was often poor. The accurate recreation of the data lost was hampered by not knowing how decisions were made in the original studies, including what demographic definitions of race or ethnicity were used, or how accuracy was determined. This limited the assessment of the included studies. Few studies have validated the methods or conducted an error analysis to assess how often race is misapplied and those that did, rarely used the most appropriate gold standard. This makes it difficult to directly compare the results of the different approaches.

Future Directions

Future studies should investigate their methodological approaches to estimate race or ethnicity, offering careful interpretations that acknowledge the significant limits of these approaches and their impact on the interpretation of the results. This may include reporting the results as a range that communicates the inherent uncertainty of the classification model. Social media data may best be used in combination with other information. In addition, we must always be mindful that race is a proxy measure for the much larger impact of being a particular race or ethnicity in a society. As a result, the variability associated with race and ethnicity might reveal more about the effects of racism and social stratification than about individual user attributes. To conduct this study ethically and rigorously, we recommend several practices that can help reduce bias and increase reproducibility.

We recommend acknowledging the researchers' bias that can influence the conceptualization of the implementation of the study. Incorporating this reflexivity, as is common in qualitative research, allows for the identification of potential blind spots that weaken the research. One way to address homogenous research teams is through the inclusion of experts in race or ethnicity or in those communities being examined. These biases can also be reduced by including members of the study population in the research process as experts and advisers [219]. Although big data from social media can be collected without ever connecting with the people who contributed the data, it does not eliminate the ethical need for researchers to include representative perspectives in research processes. Examples of patient-engaged research and patient-centered outcomes research, community-based participatory research, and citizen science (public participation in scientific research) within the health and social sciences amply demonstrate the instrumental value and ethical obligation of intentional efforts to involve nonscientist partners in cocreation of research [219]. The quality of data science can be improved by seriously heeding the imperative, *Nothing about us without us* [219]. Documenting and establishing the diverse competence attributes of a research team should become a standard. Emphasizing the importance of diverse teams within the research process will contribute to social and racial justice in ways other than improving the reliability of research.

In terms of the retrieved data, the most reliable (though imperfect) method for ascertaining race was when users self-identified their racial affiliation. Further research on overcoming the limitations of availability and sample size may be warranted. Indeed, a hybrid model with automated methods and manual extraction may be preferred. For example, automation methods could be developed to identify potential self-declarations in a user profile or timeline, which can then be manually interpreted.

Finally, we call for greater reporting of the validation by our colleagues. Without error analysis, computational techniques would not be able to detect bias. Further research is needed to establish whether any bias is systematic or random, that is, whether inaccuracies favor one direction or another.

Conclusions

We identified major concerns that affect the reliability of the methods and bias the results. There are also ethical concerns throughout the process, particularly regarding the inference of race or ethnicity, as opposed to the extraction of self-identity. However, the potential usefulness of social media research requires thoughtful consideration of the best ways to estimate demographic characteristics such as race and ethnicity [112]. This is particularly important, given the increased access to Twitter data [202,203].

Therefore, we propose several approaches to improve the extraction of race or ethnicity from social media, including representative research teams and a mixture of manual and computational methods, as well as future research on methods to reduce bias.

Acknowledgments

This work was supported by the National Institutes of Health (NIH) National Library of Medicine under grant NIH-NLM 1R01 (principal investigator: GG, with coapplicants KO and SG) and NIH National Institute of Drug Abuse grant R21 DA049572-02 to RS. NIH National Library of Medicine funded this research but was not involved in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Data Availability

The included studies are available on the web, and the extracted data are presented in Table S2 in Multimedia Appendix 1. A preprint of this paper is also available: Golder S, Stevens R, O'Connor K, James R, Gonzalez-Hernandez G. 2021. Who Is Tweeting? A Scoping Review of Methods to Establish Race and Ethnicity from Twitter Datasets. SocArXiv. February 14. doi:10.31235/osf.io/wru5q.

Authors' Contributions

SG, RS, KO, RJ, and GG contributed equally to the study. RS and GG proposed the topic and the main idea. SG and RJ were responsible for literature search. SG, RS, KO, RJ, and GG were responsible for study selection and data extraction. SG drafted the manuscript. SG, RS, KO, RJ, and GG commented on and revised the manuscript. SG provided the final version of this manuscript. All authors contributed to the final draft of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies and characteristics of included studies.
[DOCX File , 59 KB - jmir_v24i4e35788_app1.docx]

References

1. Golder S, Norman G, Loke YK. Systematic review on the prevalence, frequency and comparative value of adverse events data in social media. Br J Clin Pharmacol 2015 Oct;80(4):878-888. [doi: 10.1111/bcp.12746] [Medline: 26271492]
2. Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, et al. Utilizing social media data for pharmacovigilance: a review. J Biomed Inform 2015 Apr;54:202-212 [FREE Full text] [doi: 10.1016/j.jbi.2015.02.004] [Medline: 25720841]
3. Bhattacharya M, Snyder S, Malin M, Truffa MM, Marinic S, Engelmann R, et al. Using social media data in routine pharmacovigilance: a pilot study to identify safety signals and patient perspectives. Pharm Med 2017 Apr 17;31(3):167-174. [doi: 10.1007/s40290-017-0186-6]
4. Convertino I, Ferraro S, Blandizzi C, Tuccori M. The usefulness of listening social media for pharmacovigilance purposes: a systematic review. Expert Opin Drug Saf 2018 Nov;17(11):1081-1093. [doi: 10.1080/14740338.2018.1531847] [Medline: 30285501]
5. Golder S, Smith K, O'Connor K, Gross R, Hennessy S, Gonzalez-Hernandez G. A comparative view of reported adverse effects of statins in social media, regulatory data, drug information databases and systematic reviews. Drug Saf 2021 Feb 01;44(2):167-179 [FREE Full text] [doi: 10.1007/s40264-020-00998-1] [Medline: 33001380]
6. Bychkov D, Young S. Social media as a tool to monitor adherence to HIV antiretroviral therapy. J Clin Transl Res 2018 Dec 17;3(Suppl 3):407-410 [FREE Full text] [Medline: 30873489]
7. Kalf RR, Makady A, Ten HR, Meijboom K, Goettsch WG, IMI-GetReal Workpackage 1. Use of social media in the assessment of relative effectiveness: explorative review with examples from oncology. JMIR Cancer 2018 Jun 08;4(1):e11 [FREE Full text] [doi: 10.2196/cancer.7952] [Medline: 29884607]
8. Golder S, O'Connor K, Hennessy S, Gross R, Gonzalez-Hernandez G. Assessment of beliefs and attitudes about statins posted on Twitter: a qualitative study. JAMA Netw Open 2020 Jun 01;3(6):e208953 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.8953] [Medline: 32584408]
9. Golder S, Bach M, O'Connor K, Gross R, Hennessy S, Gonzalez Hernandez G. Public perspectives on anti-diabetic drugs: exploratory analysis of Twitter posts. JMIR Diabetes 2021 Jan 26;6(1):e24681 [FREE Full text] [doi: 10.2196/24681] [Medline: 33496671]
10. Hswen Y, Naslund JA, Brownstein JS, Hawkins JB. Monitoring online discussions about suicide among Twitter users with schizophrenia: exploratory study. JMIR Ment Health 2018 Dec 13;5(4):e11483 [FREE Full text] [doi: 10.2196/11483] [Medline: 30545811]
11. Howie L, Hirsch B, Locklear T, Abernethy AP. Assessing the value of patient-generated data to comparative effectiveness research. Health Aff (Millwood) 2014 Jul;33(7):1220-1228. [doi: 10.1377/hlthaff.2014.0225] [Medline: 25006149]
12. Cavazos-Rehg PA, Krauss MJ, Costello SJ, Kaiser N, Cahn ES, Fitzsimmons-Craft EE, et al. "I just want to be skinny.": a content analysis of tweets expressing eating disorder symptoms. PLoS One 2019;14(1):e0207506 [FREE Full text] [doi: 10.1371/journal.pone.0207506] [Medline: 30650072]
13. Ahmed W, Bath PA, Sbaffi L, Demartini G. Novel insights into views towards H1N1 during the 2009 Pandemic: a thematic analysis of Twitter data. Health Info Libr J 2019 Mar;36(1):60-72 [FREE Full text] [doi: 10.1111/hir.12247] [Medline: 30663232]
14. Cook N, Mullins A, Gautam R, Medi S, Prince C, Tyagi N, et al. Evaluating patient experiences in dry eye disease through social media listening research. Ophthalmol Ther 2019 Sep;8(3):407-420 [FREE Full text] [doi: 10.1007/s40123-019-0188-4] [Medline: 31161531]
15. Roccetti M, Salomoni P, Prandi C, Marfia G, Mirri S. On the interpretation of the effects of the Infliximab treatment on Crohn's disease patients from Facebook posts: a human vs. machine comparison. Netw Model Anal Health Inform Bioinforma 2017 Jun 26;6(1):10.1007/s13721-017-0152-y. [doi: 10.1007/s13721-017-0152-y]
16. Madden ML, Cortesi S, Gasser U, Duggan M, Smith A, Beaton M. Teens, social media, and privacy. Pew Internet & American Life Project. 2013. URL: http://www.pewinternet.org/2013/05/21/teens-social-media-and-privacy/ [accessed 2022-04-19]
17. Chou WS, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res 2009;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]
18. Social media use in 2018. Pew Research Center. URL: https://www.pewresearch.org/internet/2018/03/01/social-media-use-in-2018/ [accessed 2022-04-19]

19. Bowleg L, Teti M, Malebranche DJ, Tschann JM. "It's an Uphill Battle Everyday": intersectionality, low-income black heterosexual men, and implications for hiv prevention research and interventions. Psychol Men Masc 2013 Jan 1;14(1):25-34 [FREE Full text] [doi: 10.1037/a0028392] [Medline: 23482810]
20. Sloan L. Social Science 'Lite'? Deriving demographic proxies from Twitter. In: The SAGE Handbook of Social Media Research Methods. Thousand Oaks, California: SAGE Publications; 2016.
21. McCormick TH, Lee H, Cesare N, Shojaie A, Spiro ES. Using Twitter for demographic and social science research: tools for data collection and processing. Sociol Methods Res 2017 Aug;46(3):390-421 [FREE Full text] [doi: 10.1177/0049124115605339] [Medline: 29033471]
22. Smedley A, Smedley BD. Race as biology is fiction, racism as a social problem is real: anthropological and historical perspectives on the social construction of race. Am Psychol 2005;60(1):16-26. [doi: 10.1037/0003-066x.60.1.16]
23. Yudell M, Roberts D, DeSalle R, Tishkoff S. NIH must confront the use of race in science. Science 2020 Sep 10;369(6509):1313-1314. [doi: 10.1126/science.abd4842]
24. Davenport L. The fluidity of racial classifications. Annu Rev Polit Sci 2020 May 11;23(1):221-240. [doi: 10.1146/annurev-polisci-060418-042801]
25. Resident population and net change. U.S. Census Bureau. URL: https://www.census.gov/quickfacts/fact/note/US/RHI625219 [accessed 2022-04-19]
26. Zuberi T. Thicker Than Blood How Racial Statistics Lie. Minneapolis: University of Minnesota Press; 2001.
27. Hardeman RR, Karbeah J. Examining racism in health services research: a disciplinary self‐critique. Health Serv Res 2020 Sep 25;55(S2):777-780. [doi: 10.1111/1475-6773.13558]
28. Jenkins W, Schoenbach V, Rowley D, Ford C. 2. Overcoming the impact of racism on the health of communities: what we have learned and what we have not. In: Racism: Science & Tools for the Public Health Professional. Washington, D.C: American Public Health Association; 2019.
29. Jones CP. Toward the science and practice of anti-racism: launching a national campaign against racism. Ethn Dis 2018 Aug 08;28(Supp 1):231. [doi: 10.18865/ed.28.s1.231]
30. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
31. Updated guidance for trusted systematic reviews: a new edition of the Cochrane Handbook for Systematic Reviews of Interventions. Cochrane Database of Systematic Reviews. URL: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.ED000142/full [accessed 2022-04-21]
32. An J, Weber I. # greysanatomy vs # yankees: demographics and hashtag use on Twitter. arXiv 2016 [FREE Full text] [doi: 10.48550/arXiv.1603.01973]
33. Lightly supervised machine learning for classifying online social data. ProQuest. 2017. URL: https://www.proquest.com/openview/25ccbef5caa83249e9dc363bdb196827/1?pq-origsite=gscholar&cbl=18750 [accessed 2022-04-21]
34. Ardehaly E, Culotta A. Inferring latent attributes of Twitter users with label regularization. In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2015 Presented at: 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; 2015; Denver, Colorado. [doi: 10.3115/v1/n15-1019]
35. Ardehaly E, Culotta A. Co-training for demographic classification using deep learning from label proportions. In: Proceedings of the IEEE International Conference on Data Mining Workshops, ICDMW. 2017 Presented at: IEEE International Conference on Data Mining Workshops, ICDMW; Nov 18-21, 2017; Sorrento, Italy URL: https://doi.org/10.1109/ICDMW.2017.144 [doi: 10.1109/icdmw.2017.144]
36. Auguste D, Polman J, Miller S. A data science approach to STEM (science, technology, engineering and math) identity research for African American communities. ProQuest. URL: https://www.proquest.com/openview/ee3643a744b7c01262cb9a917611f812/1.pdf?pq-origsite=gscholar&cbl=18750&diss=y [accessed 2022-04-21]
37. Barbera P. Less is more? How demographic sample weights can improve public opinion estimates based on Twitter data. Work Paper NYU. 2017. URL: http://pablobarberacom/static/less-is-morepdf [accessed 2022-04-19]
38. Bergsma S, Dredze M, Van Durme B, Wilson T, Yarowsky D. Broadly improving user classification via communication-based name and location clustering on Twitter. In: Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2013 Presented at: 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; Jun, 2013; Atlanta, Georgia.
39. Blodgett S, Wei J, O'Connor B. Twitter Universal Dependency Parsing for African-American and Mainstream American English. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2018 Presented at: 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); Jul 15 - 20, 2018; Melbourne, Australia. [doi: 10.18653/v1/p18-1131]
40. Borradaile G, Burkhardt B, LeClerc A. Whose tweets are surveilled for the police: an audit of a social-media monitoring tool via log files. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. 2020 Presented

at: FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency; Jan 27 - 30, 2020; Barcelona Spain. [doi: 10.1145/3351095.3372841]
41. Cavazos-Rehg P, Krauss M, Grucza R, Bierut L. Characterizing the followers and tweets of a marijuana-focused Twitter handle. J Med Internet Res 2014;16(6):e157 [FREE Full text] [doi: 10.2196/jmir.3247] [Medline: 24974893]
42. Cavazos-Rehg PA, Krauss M, Fisher SL, Salyer P, Grucza RA, Bierut LJ. Twitter chatter about marijuana. J Adolesc Health 2015 Feb;56(2):139-145. [doi: 10.1016/j.jadohealth.2014.10.270] [Medline: 25620299]
43. Cavazos-Rehg PA, Zewdie K, Krauss MJ, Sowles SJ. "No high like a brownie high": a content analysis of edible marijuana Tweets. Am J Health Promot 2018 May;32(4):880-886. [doi: 10.1177/0890117116686574] [Medline: 29214836]
44. United we tweet?: a quantitative analysis of racial differences in twitter use. ResearchWorks Archive. URL: https://digital.lib.washington.edu/researchworks/handle/1773/40971 [accessed 2022-04-21]
45. Chakraborty A, Messiaso J, Benevenutoo F, Ghosh S, Ganguly N, Gummadi K. Who makes trends? Understanding demographic biases in crowdsourced recommendations. In: Proceedings of the 11th AAAI International Conference on Web and Social Media (ICWSM). 2017 Presented at: 11th AAAI International Conference on Web and Social Media (ICWSM); May 15-18, 2017; Montreal, Quebec, Canada.
46. Chen X, Wang Y, Agichtein E, Wang F. A comparative study of demographic attribute inference in twitter. Proc Int AAAI Conf Web Social Media 2021;9(1):590-593 [FREE Full text]
47. Culotta A, Ravi NK, Cutler J. Predicting Twitter user demographics using distant supervision from website traffic data. J Artificial Intell Res 2016 Feb 19;55:389-408. [doi: 10.1613/jair.4935]
48. De Choudhury M. Tie formation on Twitter: homophily and structure of egocentric networks. In: Proceedings of the 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing. 2011 Presented at: 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing; Oct 9-11, 2011; Boston, MA. [doi: 10.1109/passat/socialcom.2011.177]
49. Firmansyah F, Jones J. Did the black panther movie make blacks blacker? Examining black racial identity on twitter before and after the black panther movie release. Lecture Notes Comput Sci 2019;11864:66-78. [doi: 10.1007/978-3-030-34971-4_5]
50. Golder S, Chiuve S, Weissenbacher D, Klein A, O'Connor K, Bland M, et al. Pharmacoepidemiologic evaluation of birth defects from health-related postings in social media during pregnancy. Drug Saf 2019 Mar;42(3):389-400 [FREE Full text] [doi: 10.1007/s40264-018-0731-6] [Medline: 30284214]
51. González Y, Cutter S. Leveraging geotagged social media to monitor spatial behavior during population movements triggered by hurricanes. Scholar Common. URL: https://scholarcommons.sc.edu/etd/5367/ [accessed 2022-04-21]
52. Haffner M. A spatial analysis of non-English Twitter activity in Houston, TX. Transact GIS 2018 Apr 11;22(4):913-929. [doi: 10.1111/tgis.12335]
53. He L, Murphy L, Luo J. Using social media to promote STEM education: matching college students with role models. In: Machine Learning and Knowledge Discovery in Databases. Cham: Springer; 2016.
54. Hswen Y, Hawkins JB, Sewalk K, Tuli G, Williams DR, Viswanath K, et al. Racial and ethnic disparities in patient experiences in the United States: 4-year content analysis of Twitter. J Med Internet Res 2020 Aug 21;22(8):e17048 [FREE Full text] [doi: 10.2196/17048] [Medline: 32821062]
55. Huang W, Weber I, Vieweg S. Inferring nationalities of Twitter users and studying inter-national linking. In: Proceedings of the 25th ACM conference on hypertext and social media. 2014 Presented at: HT '14: Proceedings of the 25th ACM conference on hypertext and social media; Sep 1-4, 2014; New York, NY, United States. [doi: 10.1145/2631775.2631825]
56. Huang X, Xing L, Dernoncourt F, Paul M. Multilingual Twitter corpus and baselines for evaluating demographic bias in hate speech recognition. In: 12th Language Resources and Evaluation Conference, European Language Resources Association. 2020 Presented at: Proceedings of the 12th Language Resources and Evaluation Conference, European Language Resources Association; May 11-16, 2020; Marseille, France.
57. Karlsen AS, Scott KD. Making sense of Starbucks' anti-bias training and the arrests of two African American men: a thematic analysis of Whites' Facebook and Twitter comments. Discourse Context Media 2019 Dec;32:100332. [doi: 10.1016/j.dcm.2019.100332]
58. Kteily NS, Rocklage MD, McClanahan K, Ho AK. Political ideology shapes the amplification of the accomplishments of disadvantaged vs. advantaged group members. Proc Natl Acad Sci U S A 2019 Jan 29;116(5):1559-1568 [FREE Full text] [doi: 10.1073/pnas.1818545116] [Medline: 30642960]
59. Longley PA, Adnan M. Geo-temporal Twitter demographics. Int J Geographical Inf Sci 2015 Sep 24;30(2):369-389. [doi: 10.1080/13658816.2015.1089441]
60. Luo F, Cao G, Mulligan K, Li X. Explore spatiotemporal and demographic characteristics of human mobility via Twitter: a case study of Chicago. Applied Geography 2016 May;70:11-25. [doi: 10.1016/j.apgeog.2016.03.001]
61. Markson C. Detecting user demographics in twitter to inform health trends in social media. New Jersey Institute of Technology. URL: https://digitalcommons.njit.edu/dissertations/36/ [accessed 2022-04-21]
62. Messias J, Vikatos P, Benevenuto F. White, man, and highly followed: gender and race inequalities in Twitter. In: IEEE/WIC/ACM International Conference on Web Intelligence (WI'17). 2017 Presented at: Proceedings of the

IEEE/WIC/ACM International Conference on Web Intelligence (WI'17); Aug 23-26, 2017; Leipzig, Germany. [doi: 10.1145/3106426.3106472]
63. Mislove A, Lehmann S, Ahn Y, Onnela JP, Rosenquist J. Understanding the Demographics of Twitter Users. In: Proceedings of the International AAAI Conference on Web and Social Media. 2011 Presented at: Fifth International AAAI Conference on Weblogs and Social Media; 2011; Barcelona Spain.
64. Mohammady E. Using county demographics to infer attributes of Twitter users. In: Proceedings of the ACL Joint Workshop on Social Dynamics and Personal Attributes in Social Media. 2014 Presented at: ACL Joint Workshop on Social Dynamics and Personal Attributes in Social Media; Jun 27, 2014; Baltimore, Maryland. [doi: 10.3115/v1/w14-2702]
65. Murthy D, Gross A, Pensavalle A. Urban social media demographics: an exploration of Twitter use in major American cities. J Comput Mediat Commun 2015 Nov 19;21(1):33-49. [doi: 10.1111/jcc4.12144]
66. Nguyen V, Tran M, Luo J. Are French really that different? Recognizing Europeans from faces using data-driven learning. In: Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR). 2018 Presented at: 24th International Conference on Pattern Recognition (ICPR); Aug 20-24, 2018; Beijing, China. [doi: 10.1109/icpr.2018.8545887]
67. Pennacchiotti M, Popescu AM. Democrats, republicans and starbucks afficionados: user classification in twitter. In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. 2011 Presented at: KDD '11: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining; Aug 21- 24, 2011; KDD '11: San Diego, CA, USA. [doi: 10.1145/2020408.2020477]
68. Pennacchiotti M, Popescu AM. A machine learning approach to twitter user classification. In: Proc International AAAI Conference Web Social Media. 2011 Jul Presented at: Proceedings of the Fifth International Conference on Weblogs and Social Media; Jul 17-21, 2011; Barcelona, Catalonia, Spain.
69. Reis J, Kwak H, An J, Messias J, Benevenuto F. Demographics of news sharing in the U.S. Twittersphere. In: Proceedings of the 28th ACM Conference on Hypertext and Social Media. 2017 Presented at: HT '17: Proceedings of the 28th ACM Conference on Hypertext and Social Media; Jul, 2017; Prague, Czech Republic. [doi: 10.1145/3078714.3078734]
70. Sadah SA, Shahbazi M, Wiley MT, Hristidis V. A study of the demographics of web-based health-related social media users. J Med Internet Res 2015;17(8):e194 [FREE Full text] [doi: 10.2196/jmir.4308] [Medline: 26250986]
71. Sadah SA, Shahbazi M, Wiley MT, Hristidis V. Demographic-based content analysis of web-based health-related social media. J Med Internet Res 2016 Jun 13;18(6):e148 [FREE Full text] [doi: 10.2196/jmir.5327] [Medline: 27296242]
72. Saravanan M. Determining Ethnicity of Immigrants using Twitter Data. In: Proceedings of the 4th Multidisciplinary International Social Networks Conference. 2017 Presented at: MISNC '17: Proceedings of the 4th Multidisciplinary International Social Networks Conference : Association for Computing Machinery; Jul 17 - 19, 2017; Bangkok, Thailand. [doi: 10.1145/3092090.3092100]
73. Sowles SJ, Krauss MJ, Connolly S, Cavazos-Rehg PA. A content analysis of vaping advertisements on Twitter, November 2014. Prev Chronic Dis 2016 Sep 29;13:E139 [FREE Full text] [doi: 10.5888/pcd13.160274] [Medline: 27685432]
74. Stewart I. Now we stronger than ever: African-American English syntax in twitter. In: Student Research Workshop at the 14th Conference of the European Chapter of the Association for Computational Linguistics. 2014 Presented at: Proceedings of the Student Research Workshop at the 14th Conference of the European Chapter of the Association for Computational Linguistics; 2014; Gothenburg, Sweden. [doi: 10.3115/v1/e14-3004]
75. Vikatos P, Messias J, Manoel M, Benevenuto F. Linguistic diversities of demographic groups in twitter. In: Proceedings of the 28th ACM Conference on Hypertext and Social Media. 2017 Presented at: HT '17: Proceedings of the 28th ACM Conference on Hypertext and Social Media; Jul 4-7, 2017; Prague, Czech Republic. [doi: 10.1145/3078714.3078742]
76. Volkova S, Backrach Y. Inferring perceived demographics from user emotional tone and user-environment emotional contrast. In: 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2018 Presented at: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); Aug, 2018; Berlin, Germany. [doi: 10.18653/v1/p16-1148]
77. Wang W, Chi G. Who are you? Estimating demographics of twitter users. PAA. URL: http:///C:/Users/user/Downloads/PAA_2017_Twitter_edited%20(1).pdf [accessed 2022-04-21]
78. Wang Y, Li Y, Luo J. Deciphering the 2016 U.S. Presidential Campaign in the Twitter sphere: a comparison of the Trumpists and Clintonists. In: Proceedings of the 10th International AAAI Conference on Web and Social Media. 2016 Presented at: 10th International AAAI Conference on Web and Social Media; May 17-20, 2016; Cologne, Germany.
79. Wood-Doughty Z, Andrews N, Marvin R, Dredze M. Predicting Twitter User Demographics from Names Alone. In: Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media, Association for Computational Linguistics. 2018 Presented at: Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media, Association for Computational Linguistics; Jun 6, 2018; New Orleans, Louisiana. [doi: 10.18653/v1/w18-1114]
80. Wood-Doughty Z, Smith M, Broniatowski D, Dredze M. How Does Twitter User Behavior Vary Across Demographic Groups? In: Proceedings of the Second Workshop on Natural Language Processing and Computational Social Science, Association for Computational Linguistics. 2017 Presented at: Second Workshop on Natural Language Processing and Computational Social Science, Association for Computational Linguistics; Aug 3, 2017; Vancouver, Canada. [doi: 10.18653/v1/w17-2912]

81. Wood-Doughty Z, Xu P, Liu X, Dredze M. Using noisy self-reports to predict twitter user demographics. In: Proceedings of the Ninth International Workshop on Natural Language Processing for Social Media, Association for Computational Linguistics. 2021 Presented at: Ninth International Workshop on Natural Language Processing for Social Media, Association for Computational Linguistics; 2021; Online. [doi: 10.18653/v1/2021.socialnlp-1.11]
82. Xu S, Markson C, Costello KL, Xing CY, Demissie K, Llanos AA. Leveraging social media to promote public health knowledge: example of cancer awareness via Twitter. JMIR Public Health Surveill 2016;2(1):e17 [FREE Full text] [doi: 10.2196/publichealth.5205] [Medline: 27227152]
83. Ye J, Han S, Hu Y, Coskun B, Liu M, Qin H. Nationality classification using name embeddings. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. 2017 Presented at: CIKM '17: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management; Nov, 2017; Singapore p. 1897-1906. [doi: 10.1145/3132847.3133008]
84. Yin J, Chi G, Hook J. Evaluating the representativeness in the geographic distribution of twitter user population. In: Proceedings of the 12th Workshop on Geographic Information Retrieval. 2018 Presented at: GIR'18: Proceedings of the 12th Workshop on Geographic Information Retrieval; Nov 6, 2018; WA, USA. [doi: 10.1145/3281354.3281360]
85. Automated analysis of user-generated content on the web. ProQuest. 2021. URL: https://www.proquest.com/openview/d035cbe40f3a459a9aa347273139233f/1?pq-origsite=gscholar&cbl=18750&diss=y [accessed 2022-04-21]
86. Adnan M, Longley PA, Khan SM. Social dynamics of Twitter usage in London, Paris, and New York City. First Monday 2014;19(5). [doi: 10.5210/fm.v19i5.4820]
87. Coleman LS. “we’re a part of this city, too”: an examination of the politics of representation of D.C. Native via #dcnativesday. Social Media Soc 2021 Jan 21;7(1):205630512098444. [doi: 10.1177/2056305120984446]
88. Saha K, Yousuf A, Hickman L, Gupta P, Tay L, De Choudhury M. A social media study on demographic differences in perceived job satisfaction. Proc ACM Human Comput Interaction (HCI) 2021 Apr 13;5(CSCW1):1-29. [doi: 10.1145/3449241]
89. Hong T, Wu J, Wijaya D, Xuan Z, Fetterman J. JUUL the heartbreaker: Twitter analysis of cardiovascularhealth perceptions of vaping. Tobacco Induced Diseases 2021 Jan 8;19(January):1-6. [doi: 10.18332/tid/130961]
90. Jiang J, Vosoughi S. Not judging a user by their cover: understanding harm in multi-modal processing within social media research. In: Proceedings of the 2nd International Workshop on Fairness, Accountability, Transparency and Ethics in Multimedia: ACM. 2020 Presented at: 2nd International Workshop on Fairness, Accountability, Transparency and Ethics in Multimedia: ACM; Oct 12, 2020; Seattle WA USA. [doi: 10.1145/3422841.3423534]
91. Mueller A, Wood-Doughty Z, Amir S, Dredze M, Nobles AL. Demographic representation and collective storytelling in the me too twitter hashtag activism movement. Proc ACM Human Comput Interaction 2020;5(CSCW1):1-28. [doi: 10.1145/3449181]
92. Aguirre C, Harrigian K, Dredze M. Gender and racial fairness in depression research using social media. In: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics. 2021 Presented at: 16th Conference of the European Chapter of the Association for Computational Linguistics; Apr 19-23, 2021; Online. [doi: 10.18653/v1/2021.eacl-main.256]
93. Aguirre C, Dredze M. Qualitative analysis of depression models by demographics. In: Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access. 2021 Presented at: Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access; 2021; Online. [doi: 10.18653/v1/2021.clpsych-1.19]
94. Gunarathne P, Rui H, Seidmann A. Racial Discrimination in Social Media Customer Service: Evidence from a Popular Microblogging Platform. Manoa: University of Hawaii; 2019.
95. Ye J, Skiena S. The Secret Lives of Names? Name Embeddings from Social Media. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2019 Presented at: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 4 - 8, 2019; Anchorage, AK. [doi: 10.1145/3292500.3330751]
96. Preotiuc-Pietro D, Ungar L. User-level race and ethnicity predictors from twitter text. In: Proceedings of the The 27th International Conference on Computational Linguistics; Association for Computational Linguistics. 2018 Presented at: The 27th International Conference on Computational Linguistics; Association for Computational Linguistics; Aug 20-26, 2018; Santa Fe, New Mexico, USA URL: https://wwwaclweborg/anthology/C18-1130
97. Hinds J, Joinson AN. What demographic attributes do our digital footprints reveal? A systematic review. PLoS One 2018;13(11):e0207112 [FREE Full text] [doi: 10.1371/journal.pone.0207112] [Medline: 30485305]
98. Abubakar U, Bashir SA, Abdullahi MB, Adebayo OS. Comparative study of various machine learning algorithms for tweet classification. J Comput Sci 2019;6(4):12-24. [doi: 10.26634/jcom.6.4.15722]
99. Ardehaly E, Culotta A, Raghavan V, Aluru S, Karypis G, Miele L. Mining the demographics of political sentiment from twitter using learning from label proportions. In: Proceedings of the IEEE International Conference on Data Mining Workshops, ICDMW. 2017 Presented at: IEEE International Conference on Data Mining Workshops, ICDMW; Nov 18-21, 2017; New Orleans, Louisiana, USA. [doi: 10.1109/icdm.2017.84]

100. An J, Ciampaglia GL, Grinberg N, Joseph K, Mantzarlis A, Maus G, et al. Reports of the workshops held at the 2017 international AAAI conference on web and social media. AI Magazine 2017 Dec 28;38(4):93-98. [doi: 10.1609/aimag.v38i4.2772]
101. Anindya I. Understanding and mitigating privacy risks raised by record linkage. The University of Texas at Dallas. 2020. URL: https://utd-ir.tdl.org/handle/10735.1/9373 [accessed 2022-04-21]
102. Bardier C. Detecting electronic cigarette user disparity behaviors: an infovelliance study on twitter. ProQuest. URL: https://www.proquest.com/openview/dbcad596abc1e82eb6718d504134ec17/1?pq-origsite=gscholar&cbl=18750&diss=y [accessed 2022-04-21]
103. Basterra L, Worthington T, Rogol J, Brown D. Socio-Temporal Trends in Urban Cultural Subpopulations through Social Media. New York: IEEE; 2017.
104. Beretta V, Maccagnola D, Cribbin T, Messina E. An interactive method for inferring demographic attributes in twitter. In: Proceedings of the 26th ACM Conference on Hypertext & Social Media. 2015 Presented at: HT '15: Proceedings of the 26th ACM Conference on Hypertext & Social Media; Sep 1-4, 2015; Guzelyurt Northern Cyprus. [doi: 10.1145/2700171.2791031]
105. Bergsma S, Van Durme B. Using conceptual class attributes to characterize social media users. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2013 Presented at: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); Aug 4-9, 2013; Sofia, Bulgaria.
106. Bi B, Shokouhi M, Kosinski M, Graepel T. Inferring the demographics of search user: social data meets search queries. In: Proceedings of the International World Wide Web Conference Committee (IW3C2). 2013 Presented at: International World Wide Web Conference Committee (IW3C2); May 13-17, 2013; Rio de Janeiro Brazil. [doi: 10.1145/2488388.2488401]
107. Blevins T, Kwiatkowski R, Macbeth J, McKeown K, Patton D, Rambow O. Automatically processing tweets from gang-involved youth: towards detecting loss and aggression. In: COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. 2016 Presented at: COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers; Dec 11-16, 2016; Osaka, Japan.
108. Blodgett S, Wei J, O'Connor B. A dataset and classifier for recognizing social media english. In: Proceedings of the 3rd Workshop on Noisy User-generated Text: Association for Computational Linguistics. 2017 Presented at: 3rd Workshop on Noisy User-generated Text: Association for Computational Linguistics; Sep, 2017; Copenhagen, Denmark p. 56-61. [doi: 10.18653/v1/w17-4408]
109. Bokányi E, Kondor D, Dobos L, Sebők T, Stéger J, Csabai I, et al. Race, religion and the city: twitter word frequency patterns reveal dominant demographic dimensions in the United States. Palgrave Commun 2016 Apr 26;2(1):1-9. [doi: 10.1057/palcomms.2016.10]
110. Racial identities on social media: projecting racial identities on Facebook, Instagram, and Twitter. Minnesota State University. URL: https://cornerstone.lib.mnsu.edu/etds/781/ [accessed 2022-04-21]
111. Burnap P, Colombo G, Amery R, Hodorog A, Scourfield J. Multi-class machine classification of suicide-related communication on Twitter. Online Soc Netw Media 2017 Aug;2:32-44 [FREE Full text] [doi: 10.1016/j.osnem.2017.08.001] [Medline: 29278258]
112. Cesare N, Grant C, Nguyen Q, Lee H, Nsoesie E. How well can machine learning predict demographics of social media users? arXiv 2017:1-24.
113. Chan MS, Winneg K, Hawkins L, Farhadloo M, Jamieson KH, Albarracín D. Legacy and social media respectively influence risk perceptions and protective behaviors during emerging health threats: a multi-wave analysis of communications on Zika virus cases. Soc Sci Med 2018 Sep;212:50-59 [FREE Full text] [doi: 10.1016/j.socscimed.2018.07.007] [Medline: 30005224]
114. Chenworth M, Perrone J, Love JS, Greller HA, Sarker A, Chai PR. Buprenorphine initiation in the emergency department: a thematic content analysis of a #firesidetox Tweetchat. J Med Toxicol 2020 Jul;16(3):262-268. [doi: 10.1007/s13181-019-00754-7] [Medline: 31898154]
115. Cheong M, Lee V. Integrating web-based intelligence retrieval and decision-making from the twitter trends knowledge base. In: Proceedings of the 2nd ACM workshop on Social web search and mining. 2009 Presented at: SWSM '09: Proceedings of the 2nd ACM workshop on Social web search and mining; Nov 2, 2009; Hong Kong, China p. 1-8. [doi: 10.1145/1651437.1651439]
116. Chi G, Giles L, Kifer D, Van Hook J, Yin J. Predicting twitter user demographics as a first step in big data for population research: developing unsupervised, scalable methods using real-time, large-scale twitter data. In: Proceedings of the 2017 International Population Conference. 2017 Presented at: 2017 International Population Conference; Oct 29 -Nov 3, 2017; Cape Town, South Africa.
117. Claude F, Konow R, Ladra S. Fast compressed-based strategies for author profiling of social media texts. In: Proceedings of the 4th Spanish Conference on Information Retrieval. 2016 Presented at: CERI '16: Proceedings of the 4th Spanish Conference on Information Retrieval; Jun 14-16, 2016; Granada Spain. [doi: 10.1145/2934732.2934744]
118. Compton R, Lee C, Lu T, De Silva L, Macy M. Detecting future social unrest in unprocessed Twitter data: “emerging phenomena and big data”. In: Proceedings of the 2013 IEEE International Conference on Intelligence and Security Informatics.

2013 Presented at: 2013 IEEE International Conference on Intelligence and Security Informatics; Jun 4-7, 2013; Seattle, WA, USA. [doi: 10.1109/isi.2013.6578786]
119. Augmenting household travel survey and travel behavior analysis using large-scale social media data and smartphone GPS data. ProQuest. URL: https://www.proquest.com/openview/30e2c6f084eb32378522fe9929604037/1?pq-origsite=gscholar&cbl=18750&diss=y [accessed 2022-04-21]
120. Dai H, Hao J. Mining social media data for opinion polarities about electronic cigarettes. Tob Control 2017 Mar;26(2):175-180. [doi: 10.1136/tobaccocontrol-2015-052818] [Medline: 26980151]
121. Daughton AR, Paul MJ. Identifying protective health behaviors on Twitter: observational study of travel advisories and Zika virus. J Med Internet Res 2019 May 13;21(5):e13090 [FREE Full text] [doi: 10.2196/13090] [Medline: 31094347]
122. DeJohn AD, Schulz EE, Pearson AL, Lachmar EM, Wittenborn AK. Identifying and understanding communities using twitter to connect about depression: cross-sectional study. JMIR Ment Health 2018 Nov 05;5(4):e61 [FREE Full text] [doi: 10.2196/mental.9533] [Medline: 30401662]
123. Diaz F, Gamon M, Hofman JM, Kıcıman E, Rothschild D. Online and social media data as an imperfect continuous panel survey. PLoS One 2016;11(1):e0145406 [FREE Full text] [doi: 10.1371/journal.pone.0145406] [Medline: 26730933]
124. Using social media to evaluate public acceptance of infrastructure projects. Digital Repository at University of Maryland. 2018. URL: https://drum.lib.umd.edu/handle/1903/20848 [accessed 2022-04-21]
125. Eisenstein J. Phonological factors in social media writing. In: Proceedings of the Workshop on Language in Social Media (LASM 2013). 2013 Presented at: Proceedings of the Workshop on Language Analysis in Social Media; Jun 13, 2013; Atlanta, Georgia URL: https://aclanthology.org/W13-1102.pdf
126. A case study on Black Twitter's reactions to the framing of blacks in Dove's 2017 Facebook advertisement. Digital Communs @ University of South Florida. URL: https://digitalcommons.usf.edu/etd/8446/ [accessed 2022-04-21]
127. Filho R, Almeida J, Pappa G. Twitter Population Sample Bias and its impact on predictive outcomes: a case study on elections. In: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015. 2015 Presented at: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM); Aug 25-28, 2015; Paris France. [doi: 10.1145/2808797.2809328]
128. Filippova K. User demographics and language in an implicit social network. In: Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. 2012 Presented at: Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning; Jul 12-14, 2012; Jeju Island, Korea URL: https://aclanthology.org/D12-1135/
129. A world of one-way and two-way streetsxploring the nuances of fan-athlete interaction on Twitter. Indiana University. URL: https://scholarworks.uark.edu/cgi/viewcontent.cgi?article=3185&context=etd [accessed 2022-04-21]
130. Georgiou T, Abbadi A, Yan X. Privacy cyborg: towards protecting the privacy of social media users. In: Proceedings of the 2017 IEEE 33rd International Conference on Data Engineering (ICDE). 2017 Presented at: 2017 IEEE 33rd International Conference on Data Engineering (ICDE); Apr 19-22, 2017; San Diego, CA, USA. [doi: 10.1109/icde.2017.193]
131. Ghazouani D, Lancieri L, Ounelli H, Jebari C. Assessing socioeconomic status of Twitter users: a survey. In: Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP). 2019 Presented at: International Conference on Recent Advances in Natural Language Processing (RANLP); Sep 2-4, 2019; Varna, Bulgaria. [doi: 10.26615/978-954-452-056-4_046]
132. Gibbons J, Malouf R, Spitzberg B, Martinez L, Appleyard B, Thompson C, et al. Twitter-based measures of neighborhood sentiment as predictors of residential population health. PLoS One 2019;14(7):e0219550 [FREE Full text] [doi: 10.1371/journal.pone.0219550] [Medline: 31295294]
133. Gilchrist-Herring N. An analysis of attitudes towards transgender individuals utilizing social media usage, ethnicity, gender, age range, and level of education. ProQuest. 2020. URL: https://www.proquest.com/openview/1275cca853a493adc44dad6ad0f3d1ab/1?pq-origsite=gscholar&cbl=2026366&diss=yq [accessed 2022-04-21]
134. Giorgi S, Yaden DB, Eichstaedt JC, Ashford RD, Buffone AEK, Schwartz HA, et al. Cultural differences in Tweeting about drinking across the US. Int J Environ Res Public Health 2020 Feb 11;17(4):1125 [FREE Full text] [doi: 10.3390/ijerph17041125] [Medline: 32053866]
135. Profiling social media users with selective self-disclosure behavior. Singapore Management University (Singapore). URL: https://ink.library.smu.edu.sg/etd_coll_all/1/#:~:text=Among%20the%20social%20media%20user,is%20called%20selective%20self%2Ddisclosure [accessed 2022-04-21]
136. Towards secure and privacy-preserving online social networking services. University of California, Berkeley. URL: https://escholarship.org/uc/item/1b14t6kq [accessed 2022-04-21]
137. Gundecha P, Ranganath S, Feng Z, Liu H. A tool for collecting provenance data in social media. In: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. 2013 Presented at: KDD '13: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining; Aug 11-14, 2013; Chicago Illinois USA. [doi: 10.1145/2487575.2487713]
138. Guo G, Zhu F, Chen E, Liu Q, Wu L, Guan C. From footprint to evidence: an exploratory study of mining social data for credit scoring. ACM Trans Web 2016 Dec 27;10(4):1-38. [doi: 10.1145/2996465]

139. Gupta H, Lam T, Pettigrew S, Tait RJ. The association between exposure to social media alcohol marketing and youth alcohol use behaviors in India and Australia. BMC Public Health 2018 Jun 13;18(1):726 [FREE Full text] [doi: 10.1186/s12889-018-5645-9] [Medline: 29895264]
140. Haffner M. A place-based analysis of #BlackLivesMatter and counter-protest content on Twitter. GeoJournal 2018 Sep 1;84(5):1257-1280. [doi: 10.1007/s10708-018-9919-7]
141. Ikeda K, Hattori G, Ono C, Asoh H, Higashino T. Twitter user profiling based on text and community mining for market analysis. Knowledge Based Syst 2013 Oct;51:35-47. [doi: 10.1016/j.knosys.2013.06.020]
142. Ireland ME, Chen Q, Schwartz HA, Ungar LH, Albarracin D. Action Tweets linked to reduced county-level HIV prevalence in the United States: online messages and structural determinants. AIDS Behav 2016 Jun;20(6):1256-1264 [FREE Full text] [doi: 10.1007/s10461-015-1252-2] [Medline: 26650382]
143. Jha D, Singh R. SMARTS: the social media-based addiction recovery and intervention targeting server. Bioinformatics 2019 Oct 24;36(6):1970-1972. [doi: 10.1093/bioinformatics/btz800] [Medline: 31647520]
144. Jimenez S, Dueñas G, Gelbukh A, Rodriguez-Diaz C, Mancera S. Automatic detection of regional words for pan-hispanic spanish on twitter. In: Proceedings of the Ibero-American Conference on Artificial Intelligence. 2018 Presented at: Ibero-American Conference on Artificial Intelligence; Nov 13-14, 2018; Trujillo (Perú). [doi: 10.1007/978-3-030-03928-8_33]
145. Jones T. Toward a description of African American vernacular English dialect regions using "Black twitter". American Speech 2015;90:2015-2040. [doi: 10.1215/00031283-3442117]
146. Jørgensen A, Hovy D, Søgaard A. Learning a POS tagger for AAVE-like language. In: Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies. 2016 Presented at: 2016 conference of the North American chapter of the association for computational linguistics: human language technologies; Jun 12-17, 2016; San Diego, California. [doi: 10.18653/v1/n16-1130]
147. Kang Y, Zeng X, Zhang Z, Wang Y, Fei T. Who are happier? Spatio-temporal analysis of worldwide human emotion based on geo-crowdsourcing faces. In: Proceedings of the Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS). 2018 Presented at: Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS); Mar 22-23, 2018; Wuhan, China. [doi: 10.1109/upinlbs.2018.8559710]
148. Kent JD, Capello HT. Spatial patterns and demographic indicators of effective social media content during theHorsethief Canyon fire of 2012. Cartography Geographic Inf Sci 2013 Mar;40(2):78-89. [doi: 10.1080/15230406.2013.776727]
149. Debiasing 2016 Twitter election analysis via Multi-Level Regression and Poststratification (MRP). University of Illinois. 2019. URL: https://indigo.uic.edu/articles/thesis/Debiasing_2016_Twitter_Election_Analysis_via_Multi-Level_Regression_and_Poststratification_MRP_/10904234 [accessed 2022-04-21]
150. Kostakos P, Pandya A, Kyriakouli O, Oussalah M. Inferring demographic data of marginalized users in twitter with computer vision APIs. In: Proceedings of the 2018 European Intelligence and Security Informatics Conference (EISIC). 2018 Presented at: 2018 European Intelligence and Security Informatics Conference (EISIC); Oct 24-25, 2018; Karlskrona, Sweden. [doi: 10.1109/eisic.2018.00022]
151. Kotzé E, Senekal B. Employing sentiment analysis for gauging perceptions of minorities in multicultural societies: an analysis of Twitter feeds on the Afrikaner community of Orania in South Africa. J Transdisciplinary Res Southern Africa 2018 Nov 15;14(1):11. [doi: 10.4102/td.v14i1.564]
152. Kumar D, Ukkusuri SV. Enhancing demographic coverage of hurricane evacuation behavior modeling using social media. J Comput Sci 2020 Sep;45:101184. [doi: 10.1016/j.jocs.2020.101184]
153. Lachlan KA, Spence PR, Lin X. Expressions of risk awareness and concern through Twitter: on the utility of using the medium as an indication of audience needs. Comput Human Behav 2014 Jun;35:554-559. [doi: 10.1016/j.chb.2014.02.029]
154. Lama Y, Chen T, Dredze M, Jamison A, Quinn SC, Broniatowski DA. Discordance between human Papillomavirus Twitter images and disparities in human Papillomavirus risk and disease in the United States: mixed-methods analysis. J Med Internet Res 2018 Sep 14;20(9):e10244 [FREE Full text] [doi: 10.2196/10244] [Medline: 30217792]
155. Virtual Homespace (Re)constructing the Body and Identity Through Social Media. Binghamton: State University of New York; 2016.
156. Lee-Won RJ, White TN, Potocki B. The Black catalyst to tweet: the role of discrimination experience, group identification, and racial agency in Black Americans' instrumental use of Twitter. Inf Commun Soc 2017 Mar 23;21(8):1097-1115. [doi: 10.1080/1369118x.2017.1301516]
157. Li J, Ritter A, Hovy E. Weakly supervised user profile extraction from twitter. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2014 Presented at: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); Jun 22-27, 2014; Baltimore, Maryland p. 165-174. [doi: 10.3115/v1/p14-1016]
158. Lienemann BA, Unger JB, Cruz TB, Chu K. Methods for coding tobacco-related Twitter data: a systematic review. J Med Internet Res 2017 Mar 31;19(3):e91 [FREE Full text] [doi: 10.2196/jmir.7022] [Medline: 28363883]
159. Lin Y. Assessing sentiment segregation in urban communities. In: Proceedings of the International Conference on Social Computing. 2014 Presented at: SocialCom '14: Proceedings of the International Conference on Social Computing; Aug 4-7, 2014; Beijing China. [doi: 10.1145/2639968.2640066]

160. As seen on Twitter: African-American rhetorical traditions gone viral. Michigan University. 2012. URL: https://docgo.net/as-seen-on-twitter-african-american-rhetorical-traditions-gone-viral [accessed 2022-04-21]
161. Human activity recognition: a data-driven approach. UC Irvine. URL: https://escholarship.org/uc/item/4w98w1zd [accessed 2022-04-21]
162. Lwowski B, Rios A. The risk of racial bias while tracking influenza-related content on social media using machine learning. J Am Med Inform Assoc 2021 Mar 18;28(4):839-849 [FREE Full text] [doi: 10.1093/jamia/ocaa326] [Medline: 33484133]
163. Magdy A, Ghanem T, Musleh M, Mokbel M. Understanding language diversity in local twitter communities. In: Proceedings of the 27th ACM Conference on Hypertext and Social Media. 2016 Presented at: 27th ACM Conference on Hypertext and Social Media; Jul 10 - 13, 2016; Halifax Nova Scotia Canada. [doi: 10.1145/2914586.2914612]
164. Maheshwari T, Reganti A, Chakraborty T, Das A. Socio-ethnic ingredients of social network communities. In: Proceedings of the Companion of the ACM Conference on Computer Supported Cooperative Work and Social Computing. 2017 Presented at: CSCW '17 Companion: Companion of the ACM Conference on Computer Supported Cooperative Work and Social Computing; Feb 25- Mar 1, 2017; Portland, Oregon, USA. [doi: 10.1145/3022198.3026322]
165. Meng H, Kath S, Li D, Nguyen QC. National substance use patterns on Twitter. PLoS One 2017;12(11):e0187691 [FREE Full text] [doi: 10.1371/journal.pone.0187691] [Medline: 29107961]
166. Montasser OK. Predicting Demographics of High-Resolution Geographies with Geotagged Tweets. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence AAAI Press. 2017 Presented at: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence AAAI Press; Feb 4 - 9, 2017; Vancouver, Canada p. 1460-1466.
167. Nguyen TT, Adams N, Huang D, Glymour MM, Allen AM, Nguyen QC. The association between state-level racial attitudes assessed from Twitter data and adverse birth outcomes: observational study. JMIR Public Health Surveill 2020 Jul 06;6(3):e17103 [FREE Full text] [doi: 10.2196/17103] [Medline: 32298232]
168. Mulders D, Bodt C, Bjell J, Pentl A, Verleysen M, Montjoye Y. Improving individual predictions using social networks associativity. In: Proceedings of the 12th International Workshop on Self-Organizing Maps and Learning Vector Quantization, Clustering and Data Visualization (WSOM). 2017 Presented at: 12th International Workshop on Self-Organizing Maps and Learning Vector Quantization, Clustering and Data Visualization (WSOM); Jun 28-30, 2017; Nancy, France p. 1-8. [doi: 10.1109/wsom.2017.8020023]
169. Nelson J, Quinn S, Swedberg B, Chu W, MacEachren A. Geovisual analytics approach to exploring public political discourse on Twitter. Isprs Int J Geo Inf 2015 Mar 05;4(1):337-366. [doi: 10.3390/ijgi4010337]
170. Nguyen QC, Kath S, Meng H, Li D, Smith KR, VanDerslice JA, et al. Leveraging geotagged Twitter data to examine neighborhood happiness, diet, and physical activity. Appl Geogr 2016 Aug;73:77-88 [FREE Full text] [doi: 10.1016/j.apgeog.2016.06.003] [Medline: 28533568]
171. Nguyen QC, Li D, Meng H, Kath S, Nsoesie E, Li F, et al. Building a national neighborhood dataset from geotagged Twitter data for indicators of happiness, diet, and physical activity. JMIR Public Health Surveill 2016 Oct 17;2(2):e158 [FREE Full text] [doi: 10.2196/publichealth.5869] [Medline: 27751984]
172. Novak AN, Johnson K, Pontes M. LatinoTwitter: discourses of Latino civic engagement in social media. First Monday 2016 Jul 24;21(8). [doi: 10.5210/fm.v21i8.6752]
173. Odlum M, Cho H, Broadwell P, Davis N, Patrao M, Schauer D, et al. Application of topic modeling to Tweets as the foundation for health disparity research for COVID-19. Stud Health Technol Inform 2020 Jun 26;272:24-27 [FREE Full text] [doi: 10.3233/SHTI200484] [Medline: 32604591]
174. Oktay H, Firat A, Ertem Z. Demographic breakdown of Twitter users: an analysis based on names. In: Proceedings of the Academy of Science and Engineering (ASE). 2014 Presented at: Academy of Science and Engineering (ASE); Jan 11, 2014; Birmingham.
175. Orsolini L, Papanti GD, Francesconi G, Schifano F. Mind navigators of chemicals' experimenters? A web-based description of e-psychonauts. Cyberpsychol Behav Soc Netw 2015 May;18(5):296-300. [doi: 10.1089/cyber.2014.0486] [Medline: 25965863]
176. Pick J, Sarkar A, Rosales J. Social media use in American counties: geography and determinants. Isprs Int J Geo Inf 2019 Sep 19;8(9):424. [doi: 10.3390/ijgi8090424]
177. Developing computational approaches to investigate health inequalities. University of Washington. 2017. URL: https://soc.washington.edu/research/graduate/developing-computational-approaches-investigate-health-inequalities [accessed 2022-04-21]
178. Priante A, Hiemstra D, Saeed A, van den Broek T, Ehrenhard M, Need A. #WhoAmI in 160 characters? Classifying social identities based on twitter profile descriptions. In: Proceedings of the First Workshop on NLP and Computational Social Science. 2016 Presented at: Proceedings of the First Workshop on NLP and Computational Social Science: Association for Computational Linguistics; Nov 5, 2016; Austin, Texas. [doi: 10.18653/v1/w16-5608]
179. Riederer C, Zimmeck S, Phanord C, Chaintreau A, Bellovin S. I don't have a photograph, but you can have my footprints: revealing the demographics of location data. In: Proceedings of the ACM on Conference on Online Social Networks. 2015 Presented at: OSN '15: Proceedings of the ACM on Conference on Online Social Networks; Nov 2-3, 2015; Palo Alto California USA. [doi: 10.1145/2817946.2817968]

180. Roberts MJ, Perera M, Lawrentschuk N, Romanic D, Papa N, Bolton D. Globalization of continuing professional development by journal clubs via microblogging: a systematic review. J Med Internet Res 2015;17(4):e103 [FREE Full text] [doi: 10.2196/jmir.4194] [Medline: 25908092]
181. Roy S, Ghosh P. A comparative study on distancing, mask and vaccine adoption rates from global Twitter trends. Healthcare 2021 Apr 21;9(5):488. [doi: 10.3390/healthcare9050488]
182. Rummo PE, Cassidy O, Wells I, Coffino JA, Bragg MA. Examining the relationship between youth-targeted food marketing expenditures and the demographics of social media followers. Int J Environ Res Public Health 2020 Mar 03;17(5):1631 [FREE Full text] [doi: 10.3390/ijerph17051631] [Medline: 32138342]
183. Runge K. "Social" science, spider goats and American science audiences: investigating the effects of interpersonal networks on perceptions of emerging technologies. ProQuest. 2017. URL: https://www.proquest.com/openview/85b51025bda09f9a6941fd9b6e7cc054/1?pq-origsite=gscholar&cbl=18750&diss=y [accessed 2022-04-21]
184. Sijtsma B, Qvarfordt P, Chen F. Tweetviz: visualizing tweets for business intelligence. In: Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval. 2016 Presented at: 39th International ACM SIGIR conference on Research and Development in Information Retrieval; Jul 17 - 21, 2016; Pisa Italy. [doi: 10.1145/2911451.2911470]
185. Singh M, Singh A, Bansal D, Sofat S. An analytical model for identifying suspected users on Twitter. Cybernetics Syst 2019 Apr 02;50(4):383-404. [doi: 10.1080/01969722.2019.1588968]
186. Tomeny TS, Vargo CJ, El-Toukhy S. Geographic and demographic correlates of autism-related anti-vaccine beliefs on Twitter, 2009-15. Soc Sci Med 2017 Dec;191:168-175 [FREE Full text] [doi: 10.1016/j.socscimed.2017.08.041] [Medline: 28926775]
187. Tulloch J. An appraisal of health datasets to enhance the surveillance of Lyme disease in the United Kingdom. ProQuest. 2019. URL: https://www.proquest.com/openview/72acae0b592c9f8564d2204eaa84f5d1/1?pq-origsite=gscholar&cbl=44156 [accessed 2022-04-21]
188. Vydiswaran VG, Romero DM, Zhao X, Yu D, Gomez-Lopez I, Lu JX, et al. Uncovering the relationship between food-related discussion on Twitter and neighborhood characteristics. J Am Med Inform Assoc 2020 Feb 01;27(2):254-264 [FREE Full text] [doi: 10.1093/jamia/ocz181] [Medline: 31633756]
189. Wang Y, Feng Y, Luo J, Zhang X. Voting with feet: who are leaving Hillary Clinton and Donald Trump. In: Proceedings of the IEEE International Symposium on Multimedia (ISM). 2016 Presented at: IEEE International Symposium on Multimedia (ISM); Dec 11-13, 2016; San Jose, CA. [doi: 10.1109/ism.2016.0022]
190. Weeg C, Schwartz HA, Hill S, Merchant RM, Arango C, Ungar L. Using Twitter to measure public discussion of diseases: a case study. JMIR Public Health Surveill 2015;1(1):e6 [FREE Full text] [doi: 10.2196/publichealth.3953] [Medline: 26925459]
191. Wright M, Adams T. #KnowBetterDoBetter: an examination of Twitter impact on disaster literacy. ProQuest. 2019. URL: https://www.proquest.com/openview/1d20d09437e5921d1b996c6657c29011/1?pq-origsite=gscholar&cbl=18750&diss=y [accessed 2022-04-21]
192. Yazdavar AH, Mahdavinejad MS, Bajaj G, Romine W, Sheth A, Monadjemi AH, et al. Multimodal mental health analysis in social media. PLoS One 2020;15(4):e0226248 [FREE Full text] [doi: 10.1371/journal.pone.0226248] [Medline: 32275658]
193. Ying Q, Chiu D, Venkatramanan S, Zhang X. Profiling OSN users based on temporal posting patterns. In: Proceedings of The Web Conference. 2018 Presented at: WWW '18: Proceedings of The Web Conference; Apr 23-27, 2018; Lyon, France. [doi: 10.1145/3184558.3191592]
194. Yuan F, Li M, Zhai W, Qi B, Liu R. Social media based demographics analysis for understanding disaster response disparity. In: Proceedings of the Construction Research Congress 2020: Computer Applications. 2020 Presented at: Construction Research Congress 2020: Computer Applications; Mar 8–10, 2020; Tempe, Arizona. [doi: 10.1061/9780784482865.108]
195. Zhang Z, Bors G. “Less is more” : mining useful features from Twitter user profiles for Twitter user classification in the public health domain. Online Inf Rev 2019;44(1):213-237 [FREE Full text]
196. Zhao P, Jia J, An Y, Liang J, Xie L, Luo J. Analyzing and predicting emoji usages in social media. In: Proceedings of the The Web Conference 2018. 2018 Presented at: WWW '18: Companion Proceedings of the The Web Conference 2018; Apr 23 - 27, 2018; Lyon France. [doi: 10.1145/3184558.3186344]
197. Zhong Y, Yuan N, Zhong W, Zhang F, Xie X. You are where you go: inferring demographic attributes from location check-ins. In: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. 2015 Presented at: WSDM '15: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining; Feb 2-6, 2015; Shanghai China. [doi: 10.1145/2684822.2685287]
198. Jiang Y, Li Z, Ye X. Understanding demographic and socioeconomic biases of geotagged Twitter users at the county level. Cartography Geographic Inf Sci 2018 Feb 09;46(3):228-242. [doi: 10.1080/15230406.2018.1434834]
199. Face++ homepage. Face++. URL: https://www.faceplusplus.com/ [accessed 2022-04-21]
200. Powerful audience demographics. DemographicsPro. URL: https://www.demographicspro.com/ [accessed 2022-04-21]
201. Onomap is changing. Onomap. URL: https://www.onomap.org/ [accessed 2022-04-21]

202. Academic research: preparing for the academic research application: learn everything there is to know about applying for the academic research product track. Twitter Inc. URL: https://developer.twitter.com/en/solutions/academic-research/application-info [accessed 2022-04-19]
203. Twitter grants academics full access to public data, but not for suspended accounts. Reuters. URL: https://www.usnews.com/news/technology/articles/2021-01-26/twitter-grants-academics-full-access-to-public-data-but-not-for-suspended-accounts [accessed 2022-04-19]
204. Jung S, An J, Kwak H, Salminen J, Jansen B. Assessing the accuracy of four popular face recognition tools for inferring gender, age, and race. In: Proceedings of the Twelfth International AAAI Conference on Web and Social Media. 2018 Presented at: Twelfth International AAAI Conference on Web and Social Media; Jun 25-28, 2018; Palo Alto, California, USA.
205. Buolamwini J, Gebru T. Gender shades: intersectional accuracy disparities in commercial gender classification. In: 1st Conference on Fairness, Accountability and Transparency. 2018 Presented at: Proceedings of the 1st Conference on Fairness, Accountability and Transparency; 2018; New York.
206. Jung S, An J, Kwak H, Salminen J, Jansen B. Inferring social media users? Demographics from profile pictures: a face++ analysis on twitter users. In: Proceedings of The 17th International Conference on Electronic Business. 2017 Presented at: Proceedings of The 17th International Conference on Electronic Business (ICEB); 2017; Dubai, UAE.
207. Jointly de-biasing face recognition and demographic attribute estimationiasing Face Recognition and Demographic Attribute Estimation. In: Computer Vision – ECCV 2020. Cham: Springer; 2020.
208. Fu S, He H, Hou Z. Learning race from face: a survey. IEEE Trans Pattern Anal Mach Intell 2014 Dec 1;36(12):2483-2509. [doi: 10.1109/tpami.2014.2321570]
209. Goldinger SD, He Y, Papesh MH. Deficits in cross-race face learning: insights from eye movements and pupillometry. J Experimental Psychol Learn Memory Cognition 2009;35(5):1105-1122. [doi: 10.1037/a0016548]
210. Meissner CA, Brigham JC. Thirty years of investigating the own-race bias in memory for faces: a meta-analytic review. Psychol Public Policy Law 2001;7(1):3-35. [doi: 10.1037/1076-8971.7.1.3]
211. Jofre A, Berardi V, Brennan K, Cornejo A, Bennett C, Harlan J. Crowdsourcing image extraction and annotation: software development and case study. Digital Humanities Q 2020;14(2) [FREE Full text]
212. King RD, Johnson BD. A punishing look: skin tone and afrocentric features in the halls of justice. Am J Sociol 2016 Jul;122(1):90-124. [doi: 10.1086/686941]
213. Cavazos JG, Phillips PJ, Castillo CD, O'Toole AJ. Accuracy comparison across face recognition algorithms: where are we on measuring race bias? IEEE Trans Biom Behav Identity Sci 2021 Jan;3(1):101-111. [doi: 10.1109/tbiom.2020.3027269]
214. Buolamwini J, Gebru T. Gender shades: intersectional accuracy disparities in commercial gender classification. In: Proceedings of the 1st Conference on Fairness, Accountability and Transparency. 2018 Presented at: 1st Conference on Fairness, Accountability and Transparency; Feb 23-24, 2018; New York.
215. Torralba A, Efros A. Unbiased look at dataset bias. In: Proceedings of CVPR 2011. 2011 Presented at: CVPR 2011; Jun 20-25, 2011; Colorado Springs, CO, USA. [doi: 10.1109/cvpr.2011.5995347]
216. Moscrop A, Ziebland S, Bloch G, Iraola JR. If social determinants of health are so important, shouldn't we ask patients about them? BMJ 2020 Nov 24:m4150. [doi: 10.1136/bmj.m4150]
217. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
218. Developer terms: more about restricted uses of the Twitter APIs. Developer Platform. URL: https://developer.twitter.com/en/developer-terms/more-on-restricted-use-cases [accessed 2022-03-04]
219. Alsaied T, Allen KY, Anderson JB, Anixt JS, Brown DW, Cetta F, et al. The Fontan outcomes network: first steps towards building a lifespan registry for individuals with Fontan circulation in the United States. Cardiol Young 2020 Jul 08;30(8):1070-1075. [doi: 10.1017/s1047951120001869]

Abbreviations

ML: machine learning
NIH: National Institutes of Health
NLP: natural language processing
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by T Leung; submitted 17.12.21; peer-reviewed by JW Kwon, D Storman, A Joseph; comments to author 23.02.22; revised version received 08.03.22; accepted 23.03.22; published 29.04.22.

Please cite as:
Golder S, Stevens R, O'Connor K, James R, Gonzalez-Hernandez G
Methods to Establish Race or Ethnicity of Twitter Users: Scoping Review
J Med Internet Res 2022;24(4):e35788
URL: https://www.jmir.org/2022/4/e35788
doi:10.2196/35788
PMID:35486433

Original Paper

Blockchain-Based Architecture Design for Personal Health Record: Development and Usability Study

Thiago Bulhões da Silva Costa[1*], PhD; Lucas Shinoda[1*], MSc; Ramon Alfredo Moreno[1], PhD; Jose E Krieger[1], PhD; Marco Gutierrez[1], PhD

Instituto do Coração, Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil

[*]these authors contributed equally

Corresponding Author:
Thiago Bulhões da Silva Costa, PhD
Instituto do Coração
Hospital das Clínicas
Faculdade de Medicina, Universidade de São Paulo
Av. Dr. Enéas Carvalho de Aguiar, 44
Cerqueira César
São Paulo, 05403-900
Brazil
Phone: 55 (11) 2661 5333
Email: thiagobscosta@gmail.com

Abstract

Background: The importance of blockchain-based architectures for personal health record (PHR) lies in the fact that they are thought and developed to allow patients to control and at least partly collect their health data. Ideally, these systems should provide the full control of such data to the respective owner. In spite of this importance, most of the works focus more on describing how blockchain models can be used in a PHR scenario rather than whether these models are in fact feasible and robust enough to support a large number of users.

Objective: To achieve a consistent, reproducible, and comparable PHR system, we build a novel ledger-oriented architecture out of a permissioned distributed network, providing patients with a manner to securely collect, store, share, and manage their health data. We also emphasize the importance of suitable ledgers and smart contracts to operate the blockchain network as well as discuss the necessity of standardizing evaluation metrics to compare related (net)works.

Methods: We adopted the Hyperledger Fabric platform to implement our blockchain-based architecture design and the Hyperledger Caliper framework to provide a detailed assessment of our system: first, under workload, ranging from 100 to 2500 simultaneous record submissions, and second, increasing the network size from 3 to 13 peers. In both experiments, we used throughput and average latency as the primary metrics. We also created a health database, a cryptographic unit, and a server to complement the blockchain network.

Results: With a 3-peer network, smart contracts that write on the ledger have throughputs, measured in transactions per second (tps) in an order of magnitude close to 10^2 tps, while those contracts that only read have rates close to 10^3 tps. Smart contracts that write also have latencies, measured in seconds, in an order of magnitude close to 10^1 seconds, while that only read have delays close to 10^0 seconds. In particular, smart contracts that retrieve, list, and view history have throughputs varying, respectively, from 1100 tps to 1300 tps, 650 tps to 750 tps, and 850 tps to 950 tps, impacting the overall system response if they are equally requested under the same workload. Varying the network size and applying an equal fixed load, in turn, writing throughputs go from 10^2 tps to 10^1 tps and latencies go from 10^1 seconds to 10^2 seconds, while reading ones maintain similar values.

Conclusions: To the best of our knowledge, we are the first to evaluate, using Hyperledger Caliper, the performance of a PHR blockchain architecture and the first to evaluate each smart contract separately. Nevertheless, blockchain systems achieve performances far below what the traditional distributed databases achieve, indicating that the assessment of blockchain solutions for PHR is a major concern to be addressed before putting them into a real production.

(J Med Internet Res 2022;24(4):e35013) doi:10.2196/35013

KEYWORDS

electronic health record; personal health record; blockchain; smart contract

Introduction

Background

Two closely related concepts have been drawing the attention of the biomedical and health informatics community: electronic health record (EHR) and health information exchange (HIE). The former, broadly speaking, covers all the repositories of digital data concerning retrospective, concurrent, and prospective information for ongoing support for patient health care [1,2]. Some examples of these digital repositories are electronic medical record [3,4], electronic patient record [5,6], and the personal health record (PHR). In particular, PHR systems are thought and developed to allow health data to be controlled and at least partly collected by the patient [7-9]. The latter, in turn, covers all electronic protocols for transferring data among hospitals, clinics, and other health organizations in order to share standard information regarding patient's treatment [10]. The Office of the National Coordinator for Health Information Technology defines 3 strategies for HIE: direct, query-based, and consumer-mediated. In particular, consumer-mediated HIE allows patients to retrieve their health information, share it with health care providers and stakeholders they trust, and then make better decisions in partnership [11]. Even though it is a contentious issue yet, patients should ideally have full control of their own health data—authorizing access, sharing, and use—to reach an actual patient-centered HIE [12,13].

Despite having been separately presented, an EHR repository and an HIE protocol can be incorporated into the same system as a matter of fact. In general, they comprise systems to store, retrieve, and share health data and, invariably, lead to interoperability, scalability, reliability, privacy, and security issues regarding those data. Interoperability can reduce or even eliminate handmade administrative tasks, avoid duplicate clinical services, and facilitate access to relevant information, thereby decreasing cost and waste and improving coordinate and unplanned care [14]. Scalability can impact the scale and the transmission of health data, limiting the overall latency and throughput [15]. Reliability can increase confidence in health organizations and contribute to the total testing process, thereby reducing diagnostic errors and supporting malpractice litigation [15,16].

In particular, privacy and security relating to EHRs have been especially important issues because health data are undoubtedly sensitive. Patients must have their personal information guaranteed by civil rights, that is, only used and disclosed under their consent to indeed have privacy. In this sense, health care providers and regulators should be previously authorized before they are able to examine such information. Furthermore, patients must be protected from unauthorized access, modification, and exclusion of their stored data to really be safe. In general, lack of security can result in data theft and leakage [17]. According to the US Department of Health and Human Services Office for Civil Rights, millions of people have had sensitive information stolen and exposed owing to recurrent database attacks on the health care industry [18]. Although traditional cloud-assisted EHR has been a promising paradigm developed to address these issues, cloud environments rely on trusted and centralized third entities, which do not take full responsibility for privacy and security protection and only ensure it as much as possible [19,20]. However, blockchain-based systems, originally created to replace the trusted third party of the financial transactions [21,22], have been spreading to other fields, arousing the interest of the biomedical and health informatics community because they are tightly related to privacy and security concerns over EHR and HIE [19]. Maintaining a distributed, tamper-resistant, and continuously growing ledger, blockchain networks are systems designed to have decentralized storage and management, avoiding the single point of failure and encouraging health care providers and patients to mutually collaborate without the control of a central intermediary. They are also systems created to have a permanent audit trail and a well-defined and consensual set of transaction rules (smart contracts), supplying and certifying health data provenance and establishing formal criteria to handle sensitive information [23-26].

In view thereof, the aforesaid community has already provided an increasing number of blockchain uses: a decentralized record management to handle electronic medical records [27], a PHR smartphone app to empower patients to take control of their own health data [28], an architecture model to provide a PHR in which patients maintain a unified register of their health history even from different organizations [29], a mobile health system to remotely perform cognitive behavioral therapy for insomnia [30], a teledermatology platform to support diagnosis of skin diseases [31], a privacy-preserving location sharing for telecare medical information systems [32], an authentication service to seal biomedical database requests and the respective responses [33], a pharmaceutical supply chain management to prevent counterfeit medicines [34], a framework to share medical images [35], a platform to remotely watch patient vital signs [36], and an EHR to manage and share data from cancer treatment [37], indicating a wide range of promising applications.

Related Works and Our Contribution

There are several contributions proposing blockchain-based architecture designs to address existing problems with EHR. However, most of them have targeted electronic medical records and electronic patient records, and only few approached PHR [38,39]. Combining traditional database storage, blockchain framework, and smartphone app, Yue et al [28] were among the first to suggest an architecture model to empower the patient's ability to control and share health data. Despite adopting access control policies in different usage scenarios, the authors did not provide a detailed description of the blockchain infrastructure or perform a system assessment.

Roehrs et al [29] presented a distributed and interoperable model, named the OmniPHR, in which patients can gather their

health data to optimally manage their health history and in which health care providers, with the patient's consent, can access such data, regardless of the institutional source. Although the work pointed out several relevant concepts about the PHR, it only simulated a peer-to-peer network infrastructure using OverSim [40] and did not, in fact, implement a blockchain routine with the timestamped hashing blocks and the smart contracts. To remotely apply cognitive behavioral therapy for insomnia, Ichikawa et al [30] developed a mobile health system based on a Hyperledger Fabric blockchain infrastructure [41] to store the collected data. With a 4-node network, the authors evaluated the tamper resistance under simulated fault by taking 1 node down and subsequently, uploading new data and verifying the information recovery by lifting that node up and, from this, querying the update of the previous data [30]. Even though the work had proposed a PHR system and tested its failure resilience, it did not provide performance indicators—throughput and latency under workload [42-44]—to assess the distributed network infrastructure.

Liang et al [45] developed a mobile app for users to store their personal health data in a cloud database, from wearable or medical devices and manual inputs as well and to share it with health care providers and health insurance companies they trust. Similar to [30], Hyperledger Fabric was the blockchain framework used to implement a permissioned distributed network. Besides Fabric, to improve scalability and integrity, Merkle tree protocol, via Chainpoint [46], was the tree-based data structure used to aggregate hashed records into leaf nodes until reaching a single root—the final hash to be saved in the blockchain. To evaluate performance, the work measured the average time cost during simultaneous recording. In another work, Liang et al [47] elaborated a web application for PHR. The authors built a patient-centered architecture out of a trusted environment, supplied by Intel Software Guard Extension [48] to maintain health data and control access logs regarding these data, and out of a permanent blockchain network supplied by Tierion [49] to record both hashes of that data, certifying integrity and raw copies of that logs, thereby ensuring traceability. To evaluate performance and estimate overload, the work adopted 2 measures: the average time cost to handle a concurrent number of records and the average time cost to handle a large number of access tokens.

Uddin et al [50] proposed an end-to-end eHealthcare architecture for continuous patient monitoring, including a patient-centered component to oversee access control policies, coordinate sensors and devices, and ultimately, decide which data stream should be stored on a blockchain. Inspired by Bitcoin and Ethereum environments [21,51], the authors designed a customized blockchain infrastructure by using Java programming language, with which they implemented a selection of only trusted mining nodes to perform proof of work as consensus protocol. They compared their customized system with Bitcoin's algorithm performance, analyzing surviving generations value and central processing unit and memory monitoring as metrics [50].

Using an Ethereum-based blockchain network [51], Omar et al [52] developed a privacy-preserving platform in which patients control all health data stored on and retrieved from a blockchain, while having their identity protected by cryptographic functions. Besides that, the authors suggested specific protocols to attain pseudonymity, privacy, integrity, accountability, and security throughout platform transactions. To analyze performance, they evaluated the transaction and execution costs of smart contracts by varying the string length of the data block and employing Ethereum's crypto-fuel as a metric [52].

Roehrs et al [53] extended the OmniPHR model devised in their prior work to a production scenario, considering a private blockchain network in which only verified and authenticated participants can access and manage it. Notwithstanding Ethereum and Hyperledger Fabric had been pondered as suitable blockchain platforms, the authors preferred to develop their own infrastructure by using open application programming interfaces such as Apache Kafka [54], Apache Zookeeper [55], and others. To evaluate performance over many queries, the work observed how throughput and latency varied from 50 to 500, from 1000 to 10,000, and from 13,000 to 40,000 concurrent requests.

Through an Ethereum-based blockchain architecture, Lee et al [56] proposed an international cross-area platform to arrange data from different health care services and manage authorizations for HIE among patients, health care providers, and stakeholders. By considering a test scenario in which a person had traveled from her/his home country to a foreign one and suddenly needed medical attention, the patient, registered on the platform, successfully granted a physician authorization to access her/his PHR. The physician, in turn, also registered on the platform, searched the requested PHR, and according to it and the current patient condition, provided a diagnosis and ordered treatment and medication [56].

Alongside the preceding papers, our work builds a blockchain-based architecture out of a permissioned distributed network in order to supply a PHR system for patients to securely collect, store, share, and manage their health data. Despite the similarities, it brings a novel ledger-oriented architecture model using Hyperledger Fabric, emphasizing the importance of suitable ledgers and smart contracts to operate the overall blockchain. In addition, it provides a detailed assessment of a 3-peer network—applying throughput and latency—under workload, ranging from 100 to 2500 simultaneous record submissions, and analyses, in this case for a fixed load, the impact of increasing the network, ranging from 3 to 13 peers. At the end, our work discusses the necessity of standardizing evaluation metrics to facilitate the comparison between related works.

Methods

Blockchain and Smart Contracts

Blockchain is a distributed, tamper-resistant, and continuously growing ledger for recording desirable assets and transactions in cryptographically chained blocks. It results from a protocol to add data blocks, using public-key cryptography and hash functions, and from a protocol to validate them, using a consensus algorithm on a peer-to-peer network [21]. In this sense, each new block contains the timestamp, the hash of the previous block, and the list of the retrospective and current

digitally signed assets and transactions. Each new one is also verified by the majority of the peers in order to provide a reliable full history of the register. Once the assets and transactions are validated by consensus, the new block is recorded in the chain and becomes immutable. Subsequently, the updated ledger is shared by all peers and, thenceforth, can be attested without the need of a central authority [57,58].

Blockchain networks can be arranged either into a permissionless or a permissioned mechanism for selecting participants, to ensure the honest majority assumption, that is, the conjecture that the majority of the peers will be honest and run the consensus protocol correctly [59]. On the one hand, a permissionless blockchain network—a domain of the cryptocurrencies and financial markets [60]—does not have administrators managing membership or banning illegitimate peers; it is literally open to anyone who wants to be part of it [58,61]. In these circumstances, the network maintains incentive alignments as long as participants self-select but must expend computational resources, as in the proof of work, or even money, as in the proof of stake, to run the consensus protocol [59]. On the other hand, a permissioned blockchain network—a domain of the business and institutional practices [60]—has external administrators managing membership and defining which peers have read and write permission on the blockchain [58,61]. Although choosing the participants is outside the scope of the consensus protocol, the network establishes a consortium whereby members obey publicly documented policies to achieve group decision-making [59].

Smart contracts, in turn, are prespecified rules that allow a blockchain to be conducted in a consensual manner by all network participants. In practice, these rules represent transactions, which automatically operate digital assets and can be constructively used to state a bylaw among parties with common goals, attaining a decentralized autonomous organization [51]. Encoding state transition functions, smart contracts are logically and effectively implemented as executable programs in both domain-specific and general-purpose languages and owe their security to the accomplishment of the consensus protocol [41]. Despite opening a way to make digital codes into laws or official statements, blockchain and smart contracts are emerging technologies still. Therefore, they neither are legally binding documents nor have a jurisprudential agreement to be interpreted [61].

As already suggested in the introduction, Ethereum and Hyperledger Fabric have been the main open-source platforms used to develop blockchain frameworks into EHR and HIE [23-26,38,39]. Providing a built-in, Turing-complete, and domain-specific language (Solidity) to write smart contracts and distributed applications, Ethereum is an alternative to the first-generation scripting systems without full programming capabilities [51]. In the beginning, it was launched to create permissionless networks [62], implementing a consensus protocol (Ethash) based on the proof of work, in which a hash puzzle needs to be solved by a prover and validated by a set of verifiers [22]. To mediate this computation and avoid network abuse, Ethereum has an internal cryptocurrency (Ether) to charge transaction fees and reward nodes competing to append new blocks to the chain [63]. By the advent of the permissioned networks, Ethereum was also adapted to support general purpose languages such as Go and C++ [23] and run a consensus protocol based on the proof of authority, in which only a set of known verifiers can be selected to validate a new block [22].

Hosted by the Linux Foundation, Hyperledger Fabric, in turn, is a decentralized operating system to create permissioned networks. It allows smart contracts (chaincodes) and distributed applications to be written in Go, Java, and Node. Using an ordering service implementation based on a crash-tolerance consensus [22], it has an endorsement policy in which the smart contracts themselves, via chaincode lifecycle and private communication mechanisms (channels), specify a set of nodes to endorse transactions. In this sense, the nodes in Hyperledger Fabric have different functions: the client nodes to propose, orchestrate, and broadcast transactions, the peer nodes to execute and validate transactions as well as to maintain the ledger and the smart contracts, and the ordering service nodes to mediate state updates and dependencies during transaction execution. To control the identity of these nodes, Hyperledger Fabric has a membership service provider to handle certificate authorities and public key infrastructure and, from them, issue credentials for authentication and authorization [41,62].

As already mentioned, we opt for the latter platform to implement our permissioned network. Most of the existing platforms, including Ethereum, implement a traditional active replication for the consensus protocol, which first orders and broadcasts transactions to all peers and second waits for each peer to perform such transactions sequentially (order-execute paradigm), limiting performance and requiring an additional mechanism to prevent denial-of-service attacks from untrusted codes [41]. Executing transactions only on a subset of peers, Hyperledger Fabric implements an execute-order-validate paradigm, which first performs and verifies the transactions, then orders through a consensus protocol, and finally validates such transactions by the application-specific trust assumptions [41]. Although there are scalability issues, Hyperledger Fabric has indeed exhibited better throughput and latency values than Ethereum and other blockchain platforms [42,43,62]. In addition to these characteristics, it provides an entire set of privacy-preserving mechanisms to create and submit private transactions [41,62]—a decisive quality that influenced our decision.

Blockchain-Based Architecture Design for PHR

Using Hyperledger Fabric release 2.2, our blockchain network is structured with N peer nodes (P1, P2, …, PN), with N greater than or equal to 3, and an ordering service node. The peer nodes are the basic elements of the network because they store ledgers (L) and smart contracts (S) [64]. Ideally, each peer infrastructure must be under the responsibility of a different corporation. In this sense, they can represent N interested parties—the government, health organizations, civil society institutions, hospitals, among others—acting for the maintenance and evolution of a PHR. Thus, the peer nodes provide network services such as the writing and reading of the ledgers for administrators and users relating to these parties. In theory, there is no upper bound for N other than that imposed by the hardware and software running the consensus protocol. In this

sense, we first investigate a 3-peer network because it is the smallest one in which the majority assumption is reasonable and, second, analyze the impact of increasing N.

The peers are associated with their respective client nodes (CL1, CL2, …, CLN)—the elements outside the network that allow an application to be connected to the blockchain, that is, an external application accesses ledgers and smart contracts via client-peer connection. By means of a software development kit [65], Hyperledger Fabric supplies an application programming interface with instructions to perform the aforementioned connection in order to submit transactions as well as to receive responses after these transactions are finished or interrupted earlier due to the lack of consensus. In addition, Hyperledger Fabric conceives of a channel (C) as a primary communication pathway by which peers and clients can establish a consortium with well-defined policies, thus providing a mechanism for isolating assets and transactions from the rest of the network. In this context, each smart contract and the respective ledger can be separately invoked on a specific channel only by users previously registered in the consortium, thereby ensuring interoperability and privacy [64].

The peers get assigned to the consortium—the government, health organizations, civil society institutions, and hospitals in our example—by their respective certificate authorities (CA1, CA2, …, CAN), the elements that generate public and private key infrastructure to issue identities via digital certificates [66]. Hyperledger Fabric has adopted the X.509 standard [67] as its primary certificate system. Whenever one of the consortium members establishes a client-peer connection to access the blockchain resources, these certificate authorities attest to the channel the digital identity of the applicant and her/his rights to use the required smart contract. As already mentioned, the Fabric component mapping identities with their own rights is the membership service provider, which inspects who participates in the network and their channels, identifying roles and limits of all administrators and users [64].

Lastly, the ordering service node mediates the interaction between peers during a transaction submission and ensures a consistent ledger after performing the consensus protocol. In Hyperledger Fabric, the endorsement policy occurs as a result of a 3-phase process: (1) proposal, (2) ordering and packing, and (3) validation and commit. Roughly speaking, in the first phase, a client node submits a transaction proposal, which is distributed to the endorsement peers and is independently executed by them, returning a set of endorsed responses—inconsistent responses can be already detected and discarded, finishing the workflow early. In the second phase, the ordering service node collects these responses and packages them into blocks, preparing for the next step. In the third phase, the ordering service node finally distributes the blocks to the peers, which in turn validate them to verify the endorsement phase and, only after that, commit to the ledger—failed transactions terminate the workflow without writing on the blockchain [64]. Figure 1 summarizes our architecture design, just omitting the ordering service node for a better visualization. The N peers in our network are configured to participate in the endorsement phase.

Turning the analysis to the ledgers and smart contracts, our approach considers 3 classes: (1) for personally identifiable information (PII), (2) for health record information (HRI), and (3) for record sharing information (RSI) (Figure 2). By opting for 2 or more ledgers (3 in our case), blockchains also evolve in an intricate and unpredictable way, which makes any attempt to tamper with health records even more difficult and unlikely as long as the system is in use. Besides the tamper resistance, such configuration permits the blockchain network to be structured in an oriented-ledger architecture design, making data organization aligned with the resource consumption.

PII is designed to store basic form data filled by the user at the moment of registration in the system. There are smart contracts to add, update, retrieve, and view history, respectively, to write a new record, rectify a registration error, perform a system login, and recover an updating log. To add a PII, the user needs to register with a password—converted into a hash value for security—and thus, receive a unique identifier (PII ID). Once registered, the PII ID is only recovered from a login, that is, identity number or email and the correct password hash. All other smart contracts, including those from HRI and RSI, are only able to write and read the ledger by means of a PII ID as the prefix of a composite key. In such a way, each user just accesses her/his data. HRI, in turn, is designed to store metadata from a health document, together with a hash value and a database ID, for reasons to be explained later in the text. Similar to the PII, there are smart contracts to add, update, retrieve—in this case, to recover a single record—and view history, and one further to list all records for a user. Finally, RSI is designed to store HIE logs in order to track every time a copy of a health document leaves the repository, either for downloading or sharing. There are smart contracts to add, retrieve, and list. To keep HIE logs unchanged, we opt for not creating a smart contract to update them; hence, neither one to view history.

Notwithstanding the necessity of smart contracts to list HRI and RSI, for the sake of security, PHR systems do not need one to list PII. One such smart contract would allow an administrator to list users and associate them with their respective HRI and RSI. To prevent such a situation and actually grant to a user the exclusive right of her/his health data ownership, the PII ID is only retrieved with the correct password hash. Because PII ID is a required index prefix to use HRI and RSI smart contracts, the absence of a PII listing function represents an additional security element directly configured in the operation rules of the system. Note that these settings are not just programming practices. Because smart contracts state the logic of the blockchain network, a set of security practices at the present time can evolve to rule status in the near future. Indeed, using smart contracts is a great opportunity to create a bylaw or business logic for PHR, defining which is and is not permitted regarding the access to patient information.

Although there are several smart contracts, they consist of 2 basic network operations: writing and reading. The former is used to invoke either the creation of a new state on the ledger or the modification of an existing one—without deleting past states, evidently. Smart contracts to add and update fall into this type. To perform writing, a client node needs to start an endorsement policy and reach consensus—a process that

involves all peers. The latter operation, in turn, is used to query the current state and history of a ledger. Smart contracts to retrieve, list, and view history fall into this another type. To perform reading, a client node just connects to its associated peer and thus queries the stored ledger, independently of the other peers. Similar to the client-peer connection resources, by means of another software development kit [68], Hyperledger Fabric supplies an application programming interface with instructions for the development of smart contracts and business logic. As already mentioned, Fabric provides support for Go, Java, and Node, but we adopt the latter as our primary programming language to build our architecture design.

Figure 1. Design of our blockchain network, considering N endorsement peers and their respective clients and certificate authorities. Each channel is associated with a specific set of ledgers and smart contracts, respectively named as personally identifiable information, health record information, and record sharing information. Ideally, each triple peer-client-certificate authority must be under the responsibility of a different organization or institution. HRI: health record information; PII: personally identifiable information; RSI: record sharing information; P: peer; S: smart contract; L: ledger; CL: client; CA: certificate authority; C: channel.

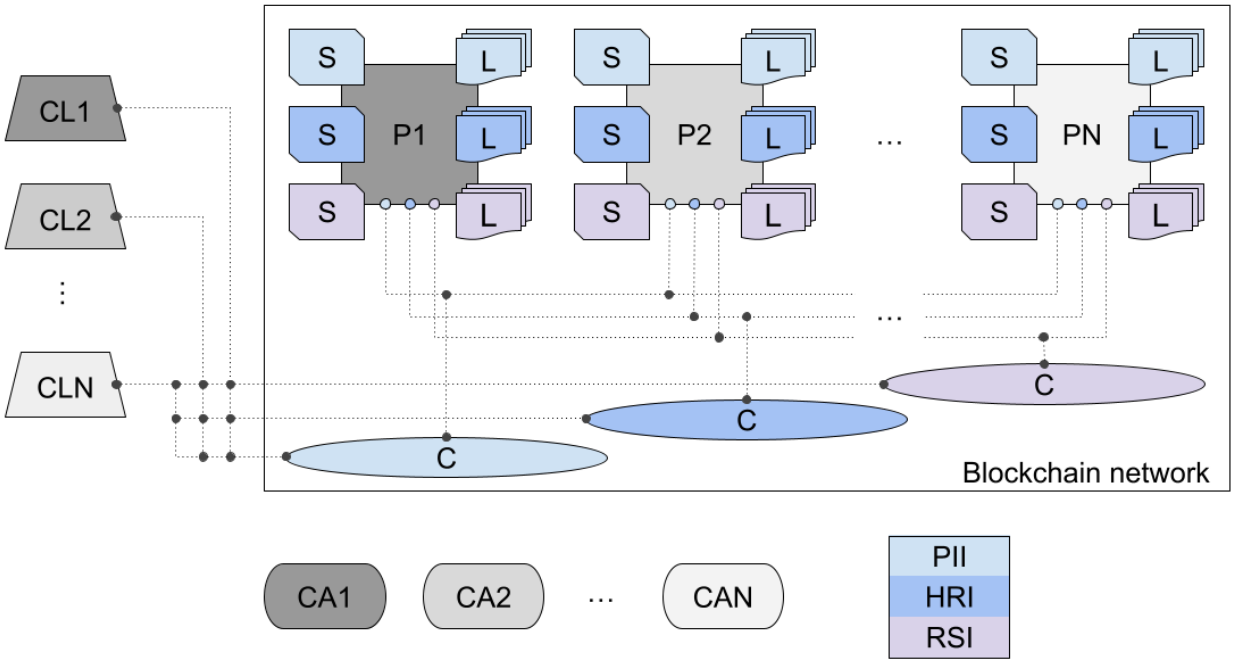
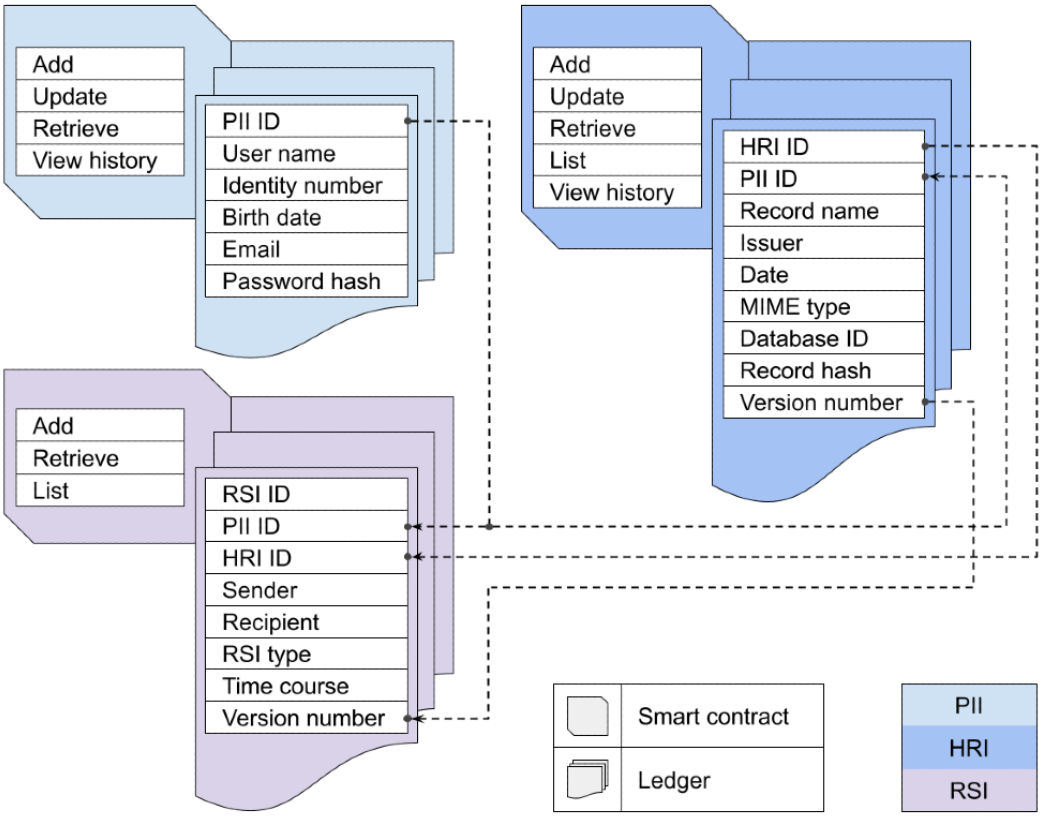


Figure 2. Design of the ledgers and their respective smart contracts. They fall into 3 classes: personally identifiable information, health record information, and record sharing information. HRI: health record information; MIME: Multipurpose Internet Mail Extensions; PII: personally identifiable information; RSI: record sharing information.



Health Database, Cryptographic Unit, and Server

Although blockchain technology provides security tools against record tampering, it is still not suitable for storing a large volume of data, despite the efforts made to meet this requirement [69]. Nowadays, only metadata such as PII, HRI, and RSI can be recorded and maintained in a blockchain network. Therefore, our system also includes a NoSQL database to permit the scaling of all sorts of health data (text, signals, and images) in clusters of machines. To implement our NoSQL health database, we adopted MongoDB, a document-oriented database, which indeed supports methods to distribute and replicate data across multiple machines and provides lower execution times than a relational one, making the scaling out easier for applications demanding both a large volume of data and a large number of queries [70]. In summary, while metadata (PII, HRI, and RSI) are stored on the blockchain network, data, that is, digital health documents, are stored on a distributed health database as soon as the network achieves consensus. In these circumstances, the health documents are hashed and their hash values are included as metadata in HRI to shield them from breaches. Note that the blockchain network represents an audit system [71] and the health documents can be anonymized in the health database, apart from a database ID in the sole possession of the user.

As a further safeguard, the data and metadata are encrypted. When a user registers in our system, she/he automatically receives a key to encrypt information entering the system as well as to decrypt that leaving out by means of a cryptographic unit. Each user obtains her/his own key and is only capable of decrypting her/his own data evidently. Because our health database is configured to store documents smaller than or equal to 100 MB, we opt for using the advanced encryption standard (AES), a symmetric key block encryption algorithm recommended by the National Institute of Standards and Technology. The AES handles block sizes of at least 128 bits and key sizes of 128, 192, and 256 bits. The AES also accepts 5 modes of operation, that is, electronic codebook, cipher block chaining (CBC), cipher feedback, output feedback, and counter, for preventing identical ciphertexts to be generated from blocks containing the same data, a breach that facilitates a malicious opponent to accumulate enough plaintext-ciphertext pairs and thus find the key by exhaustion in a feasible time. In particular, CBC requires an initialization vector, which takes an exclusive-OR operation with the first plaintext block and, if randomly generated, provides different ciphertexts from the same data [72,73]. We adopt CBC as our mode of operation and 256 bits as our key and initialization vector sizes, resulting in the AES-256-CBC algorithm. The key and initialization vector of each user are allocated in a private wallet/folder, alongside her/his digital certificate.

As a final module, we build a server infrastructure out of a Node framework to host the blockchain clients and, thereby, provide blockchain resources for external applications. Through a control unit, and performing specific calls for each smart contract as well as for each database operation, this server supports the registration and access of users, the inclusion, updating and retrieval of health documents, and the creation of links to download and share these documents—only with the consent

and supervision of the respective user, evidently. Roughly speaking, this server executes 3 basic steps: (1) it receives requests from external applications, (2) according to each request, it accesses the corresponding network and database resources, and (3) it returns consistent responses to those applications. Because the server works as an intermediate system between blockchain network, health database, and external applications, it conveniently accommodates the cryptographic unit. In this way, sensitive information is encrypted as soon as it enters the system and only decrypted when leaving out. Figure 3 highlights all these interconnected modules and Figure 4 exemplifies the flow of information during the query or record request of a health document.

Figure 3. Sketch of the overall system, exhibiting the interconnections between server, health database, and blockchain network, in order to provide personal health record resources for external applications. HRI: health record information; PII: personally identifiable information; RSI: record sharing information.

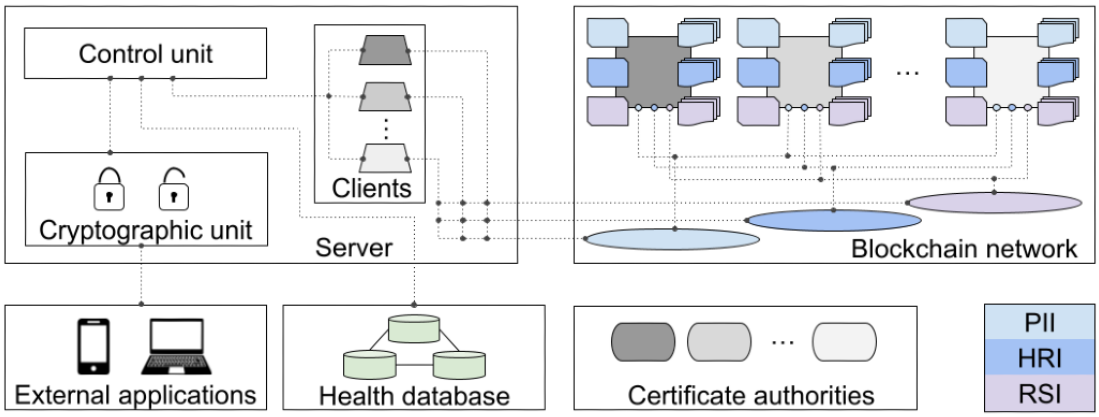
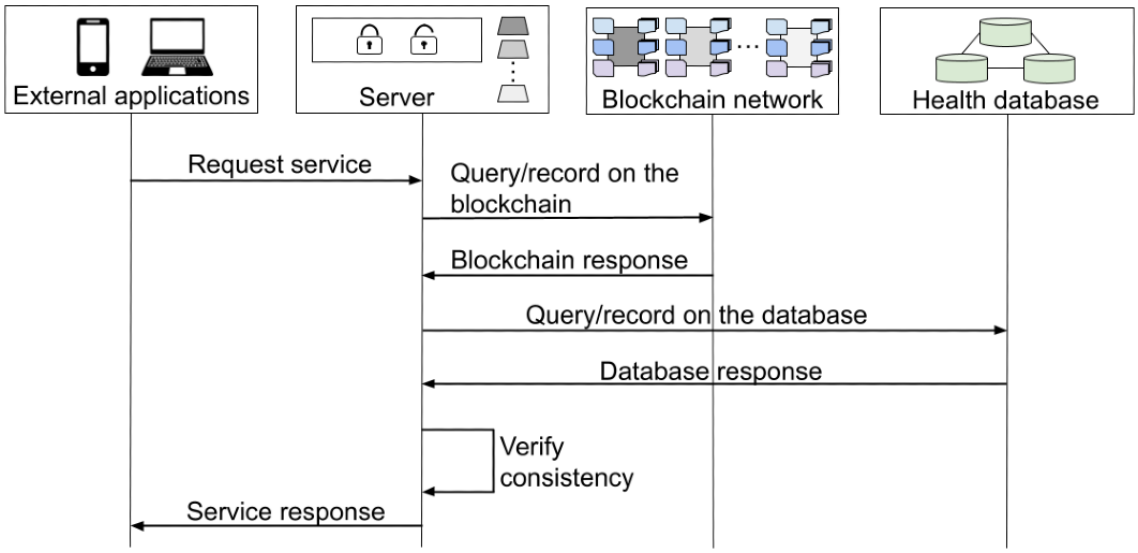


Figure 4. Flow of information during the query or record request of a health document. The server only returns a successful response if data and metadata are consistent. The flow can be interrupted earlier owing to lack of consensus.



Evaluation Benchmark

To evaluate our blockchain-based architecture design, we use Hyperledger Caliper—a benchmark tool released by the Hyperledger community for measuring the performance of blockchain systems and producing reports containing metrics commonly accepted, such as throughput and latency. Caliper supports Ethereum and Hyperledger Fabric, allowing computer scientists and engineers to compare EHR proposals developed from the 2 main platforms at present. It is capable of generating a workload for a system under test (SUT) and continuously monitoring responses from this SUT [44,74].

To run an experiment, Caliper requires a benchmark file, a network file, and workload modules. The first one presents custom configurations to run the benchmark, such as the number of workers to perform a workload, the round settings, the number of submissions, the round length in seconds, the rate at which transactions are sent to the blockchain, among others. The second one presents the layout of the SUT—basically, the addresses and identities of the nodes and the channels and smart contracts to be used during the test. Lastly, workload modules are Node functions exported to simulate client nodes sending requests to the SUT, that is, in each round, a different workload module can be used to generate and submit transactions to the SUT, according to the configurations in the benchmark and

network files. Therefore, Caliper can emulate many clients injecting workloads in a blockchain network [44,74].

As already mentioned, 2 basic metrics to assess blockchain performance are throughput and latency. The former, usually given in transactions per second (tps), represents the total number of valid transactions reached in a period of time [42]. In this sense, invalid transactions are subtracted from the total to yield the valid ones. Because transactions fall into reading and writing operations, throughput also falls into these types. On the one hand, reading throughput may be informative, but it only measures operations taken on a single client-peer connection, independently of the other peers, and therefore, is not a primary measure. On the other hand, writing throughput considers operations invoking the consensus protocol and thus committing transactions at all endorsement peers, making it much more informative than the preceding rate [75].

The latter, in turn, usually given in seconds, represents the time taken for a transaction to conclude and return a response [42]. Similar to the throughput, latency also falls into reading and writing types: the first one measures delays from a single client-peer connection, while the second one from all endorsement peers. In particular, writing latency includes the propagation and settling times due to the consensus protocol, considering delays measured over the entire network. Although this metric is generally calculated per transaction, the average latency is more suitable to assess blockchain performance [75].

Results

With a 3-peer network, our first benchmark is set to run a workload, from 100 to 2500 simultaneous submissions of health metadata, with steps of 100, on each smart contract of the PII, HRI, and RSI templates. We limit our test to 2500 loads because Hyperledger Fabric is standardly configured to perform a maximum of 2500 concurrent requests. Writing scenarios are configured to use 5 workers submitting at the same time 10,000 transactions, each one totalizing 50,000. Reading scenarios are configured to use the same 5 workers in parallel but to randomly request records during 600 seconds of continuous operation. The rate controller is kept in a fixed-load mode, starting at 50 tps and 500 tps, for writing and reading transactions, respectively, and growing to reach maximum rates. Because PII, HRI, and RSI are designed to store ciphertexts only, in our test, all simulated submissions of health metadata are randomly generated as strings of fixed length for each smart contract field. An empty blockchain network is raised in each load test to guarantee an equal condition. Our test environment consists of a machine having an Intel Xeon E-2246G processor (12 MB cache, 3.60 GHz, 6 cores, 12 threads), an NVIDIA Quadro P1000 graphic adapter, and a random access memory of 16 GB, running Ubuntu 18.04.5 LTS 64 bits operating system.

Figure 5 exhibits the throughputs and average latencies in relation to PII, HRI, and RSI smart contracts under workload. We do not report transaction errors because not one occurred. Disregarding the small variations inherent in each workload trial, and albeit with different baselines, the throughputs of all smart contracts remain fairly constant over the interval, a consistent behavior given that the system responses appear to be invariant to load. Smart contracts to add and update a record have rates with an order of magnitude close to 10^2 tps, while those to retrieve, list, and view history have rates close to 10^3 tps. As already suggested, this difference arises mainly because writing transactions trigger the consensus protocol, mobilize the network as a whole, and then need more time to process all submissions, whereas reading ones only involve a single client-peer connection. Although with different upward slopes, the average latencies of all smart contracts present a linear growth as workload range varies, a reasonable behavior inasmuch as an increase in submissions demands a proportional increase in processing. In this case, smart contracts to add and update a record have delays with an order of magnitude close to 10^1 seconds, while those to retrieve, list, and view history have delays close to 10^0 seconds. In analogy with the throughput, there is an obvious difference between writing and reading transactions, for the same reason as before.

Even though throughputs of reading transactions present a similar order of magnitude, they have significant differences between them. Smart contracts to retrieve, list, and view history have throughputs varying, respectively, from 1100 tps to 1300 tps, from 650 tps to 750 tps, and from 850 tps to 950 tps. Their latencies, in turn, grow at slightly different linear rates, albeit alike. These 2 pieces of evidence suggest that reading transactions can impact the overall system response if they are equally requested. An external application under a real situation has to consider the smallest of these values as the upper limit to avoid overload. With a fixed load at 2000 submissions, our second benchmark is set to increase the network size from 3 to 13 peers, with steps of 2, and perform, for each case, the writing and reading scenarios of the previous experimental protocol. We limit the largest network to 13 peers because by considering our test environment, Hyperledger Fabric has a very poor performance beyond this value, resulting in many transaction failures. Figure 6 displays the throughputs and average latencies when the size of the network increases. For reading smart contracts, they remain fairly constant over the interval, sustaining orders of magnitude close to 10^3 tps and 10^0 seconds, respectively, a consistent behavior given that such operations rely on a single client-peer connection. Writing smart contracts, in turn, start with throughputs close to 10^2 tps but end with rates close to 10^1 tps, exhibiting an exponential decay. They also start with latencies of 10^1 seconds but end with delays of 10^2 seconds, presenting a linear growth. Both pieces of evidence corroborate the well-known scalability issue of Hyperledger Fabric when the number of endorsement peers increases.

As a final comment when observing throughputs and average latencies in Figures 5 and 6, despite the obvious differences regarding each smart contract operation (to add, update, retrieve, list, and view history), the ongoing metrics of the 3 proposed templates (PII, HRI, RSI) do not reveal large deviations within a single operation, indicating a similar performance even with slightly different sizes of health metadata.

Figure 5. Throughput (measured in transactions per second) and average latency (measured in seconds) of all smart contracts under workload, ranging from 100 to 2500 concurrent submissions of health metadata, with steps of 100. HRI: health record information; PII: personally identifiable information; RSI: record sharing information; tps: transactions per second.

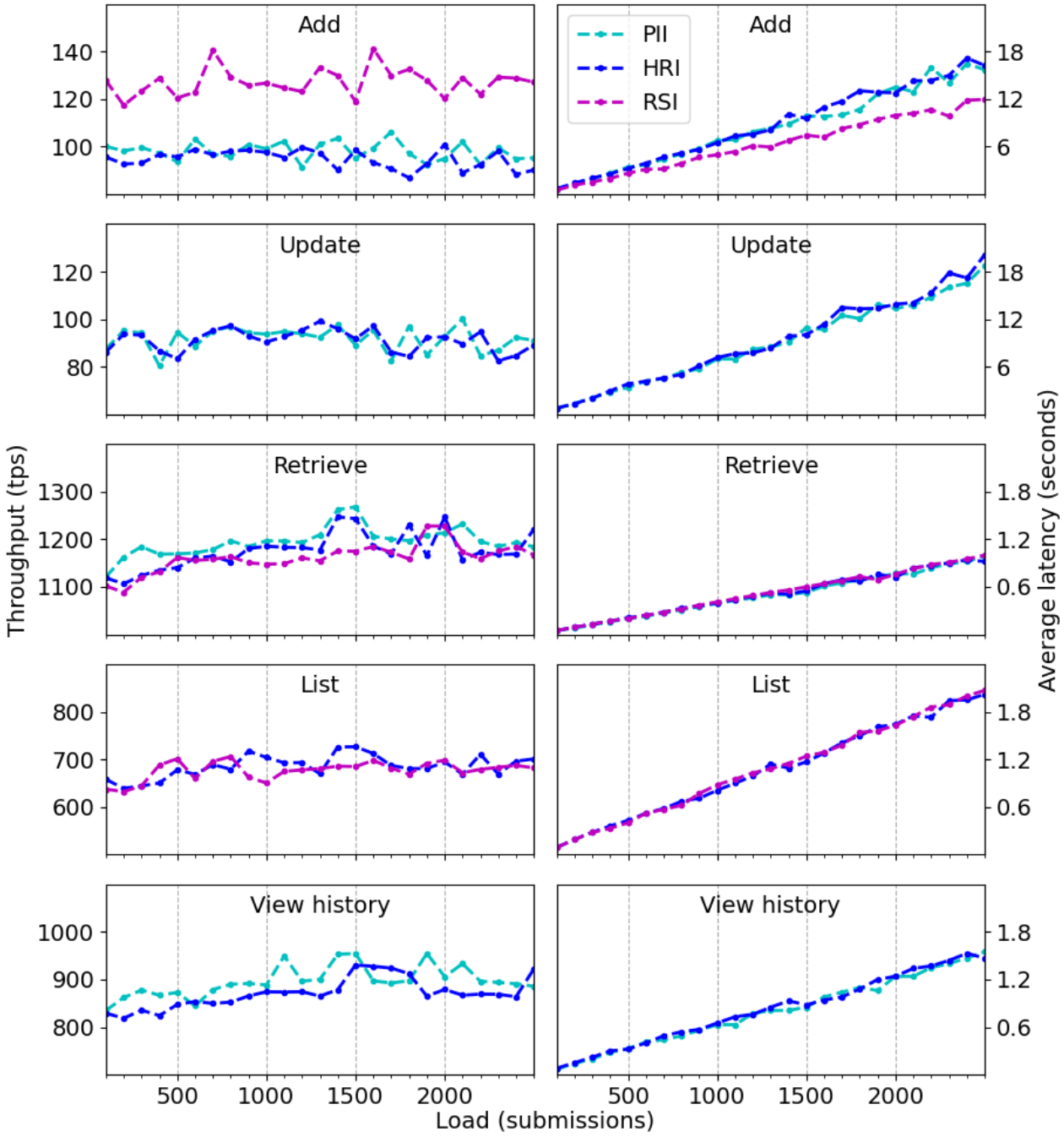
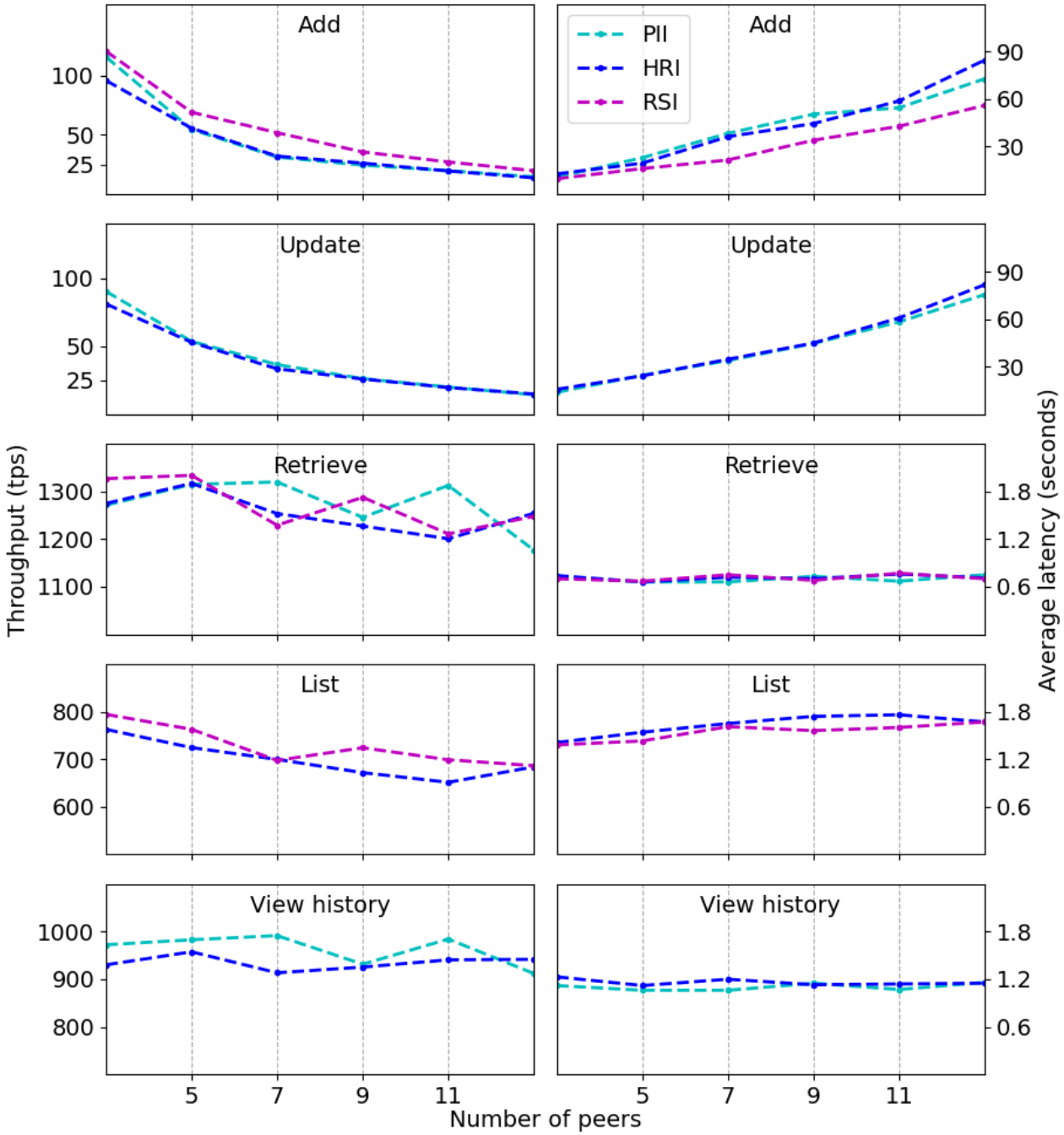


Figure 6. Throughput (measured in transactions per second) and average latency (measured in seconds) of all smart contracts, by considering a network increase from 3 to 13 endorsement peers, with steps of 2. HRI: health record information; PII: personally identifiable information; RSI: record sharing information; tps: transactions per second.



Discussion

The results of this study are comparable to those reported previously in the literature [42-44], indicating that blockchain systems achieve performances far below what the traditional distributed databases achieve [76,77]. Traditional databases make use of concurrency control, for example, 2-phase locking to ensure atomicity, consistency, isolation, and durability. By and large, they exhibit better performance because they consider simple failure models such as crash failure. Oppositely, blockchain systems consider Byzantine failure and, in the worst scenario, a hostile environment in which nodes can join and leave the network, which undeniably makes the overhead of concurrency control much more difficult to handle [42]. However, despite being widely recommended by the blockchain community [42-44], throughput and latency have not been commonly adopted metrics for evaluating PHR. Yue et al [28] did not even perform a system assessment; Roehrs et al [29] only simulated a peer-to-peer network and then, provided an inferred latency; Ichikawa et al [30] assessed the tamper resistance in a fault simulation context; Liang et al [45] and Liang et al [47] measured an average time cost to handle simultaneous records; Uddin et al [50] employed surviving generations value as well as central processing unit and memory monitoring; Omar et al [52] opted for Ethereum's crypto-fuel; and Lee et al [56] proposed a test scenario in which a person

and a doctor actually used the system [56]. Only Roehrs et al [53] observed how throughput and latency varied, under workload, from 50 to 500, from 1000 to 10,000, and from 13,000 to 40,000 concurrent requests as light, medium, and heavy scenarios, respectively. The authors achieved, in the heavy one, impressive values: 2298 tps and 0.404 seconds on average [53]. However, the authors arranged health data on single data blocks with writing and reading capabilities as a unified view of patients, thus not performing a bylaw or business logic for PHR and only assessed reading transactions considering these blocks. Furthermore, they did not develop their network from an open-source platform, hindering system reproducibility.

In practice, most of the works focus more on describing how blockchain models can be used in a PHR scenario than whether these models are in fact feasible to support a large number of users. Because the health industry can easily cover tens or even hundreds of millions of patients in a single country, we think the assessment of blockchain solutions for PHR is a major concern to be addressed before putting them into a real production. In view thereof, there is a latent necessity of standardizing evaluation metrics to facilitate the comparison between related works. We think that throughput and average latency are suitable metrics for this purpose as well as Hyperledger Caliper and BLOCKBENCH [42] adequate frameworks to perform this evaluation.

Toward a consistent, reproducible, and comparable PHR evaluation, and by regarding throughput and latency, we are the first to evaluate with Hyperledger Caliper the performance of a PHR blockchain architecture. Because Caliper is the official benchmark to access blockchain networks built out of Fabric, we believe that our results bring important insights to the limits and advantages of using Fabric to design PHR repositories. Moreover, Caliper can be adapted to access Ethereum-based systems, facilitating the comparison between architectures created with the 2 main open-source platforms at the present time. To the best of our knowledge, we are also the first to evaluate each smart contract separately. Previous works considered smart contracts as falling only into writing and reading transactions and have just identified dissimilarities between these 2 types. However, we reveal that, especially in relation to reading ones, throughput and latency can have significant differences, impacting the overall system response if these transactions are equally requested under the same workload.

Specifically in relation to our proposal, as a first implementation, the blockchain network, the health database, and the server are allocated through virtual machines on a single physical device, only simulating a decentralized system, which represents a limitation of our work. Furthermore, because we are primarily interested in the blockchain architecture, the health database and the server are incorporated in the model but they are not actually tested considering an external application under a real situation, which represents an additional limitation. We leave these improvements for future work because we believe that our current results already provide important advice to the biomedical and health informatics community.

In conclusion, the importance of blockchain-based architectures for PHR lies in the fact that they are thought and developed to allow a patient to control and at least partly collect health data, as well as to share health information on her/his own. Ideally, these systems should provide the full control of such data for the respective owner [78]; that is, each patient must authorize health care providers and stakeholders (s)he trusts before they can access her/his personal health data. Exactly because blockchain systems are tightly related to privacy and security concerns, several works are proposing blockchain-based solutions to the health care industry. In line with these efforts, we build a novel ledger-oriented architecture out of a permissioned distributed network in order to support a PHR system for patients to securely collect, store, share, and manage their health data. We emphasize the importance of suitable ledgers and smart contracts to operate the overall blockchain network and provide a detailed assessment of this network under workload, ranging from 100 to 2500 concurrent submissions, and increasing the network size from 3 to 13 peers. To the best of our knowledge, we are the first to evaluate with Hyperledger Caliper the performance of a PHR blockchain architecture and the first to evaluate each smart contract separately. However, our system elements are allocated through virtual machines on a single physical device, only simulating a decentralized system. Besides this limitation, our health database and server are incorporated in the model but they are not actually tested considering an external application under a real situation. We intend to perform these enhancements in future works.

Acknowledgments

We thank Foxconn for the financial support and Instituto do Coração, Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo for the research infrastructure.

Conflicts of Interest

None declared.

References

1. Häyrinen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. Int J Med Inform 2008 May;77(5):291-304. [doi: 10.1016/j.ijmedinf.2007.09.001] [Medline: 17951106]

2. Hripcsak G, Albers DJ. Next-generation phenotyping of electronic health records. J Am Med Inform Assoc 2013 Jan 01;20(1):117-121 [FREE Full text] [doi: 10.1136/amiajnl-2012-001145] [Medline: 22955496]
3. Ludwick DA, Doucette J. Adopting electronic medical records in primary care: lessons learned from health information systems implementation experience in seven countries. Int J Med Inform 2009 Jan;78(1):22-31. [doi: 10.1016/j.ijmedinf.2008.06.005] [Medline: 18644745]
4. Zahabi M, Kaber DB, Swangnetr M. Usability and Safety in Electronic Medical Records Interface Design: A Review of Recent Literature and Guideline Formulation. Hum Factors 2015 Aug;57(5):805-834. [doi: 10.1177/0018720815576827] [Medline: 25850118]
5. Mikkelsen G, Aasly J. Concordance of information in parallel electronic and paper based patient records. International Journal of Medical Informatics 2001 Oct;63(3):123-131. [doi: 10.1016/s1386-5056(01)00152-6]
6. Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. BMJ 2003 May 17;326(7398):1070 [FREE Full text] [doi: 10.1136/bmj.326.7398.1070] [Medline: 12750210]
7. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.M2025] [Medline: 16357345]
8. Archer N, Fevrier-Thomas U, Lokker C, McKibbon KA, Straus SE. Personal health records: a scoping review. J Am Med Inform Assoc 2011;18(4):515-522 [FREE Full text] [doi: 10.1136/amiajnl-2011-000105] [Medline: 21672914]
9. Roehrs A, da Costa CA, Righi RDR, de Oliveira KSF. Personal Health Records: A Systematic Literature Review. J Med Internet Res 2017 Jan 06;19(1):e13 [FREE Full text] [doi: 10.2196/jmir.5876] [Medline: 28062391]
10. Rudin RS, Motala A, Goldzweig CL, Shekelle PG. Usage and Effect of Health Information Exchange. Ann Intern Med 2014 Dec 02;161(11):803. [doi: 10.7326/m14-0877]
11. Williams C, Mostashari F, Mertz K, Hogin E, Atwal P. From the Office of the National Coordinator: the strategy for advancing the exchange of health information. Health Aff (Millwood) 2012 Mar;31(3):527-536. [doi: 10.1377/hlthaff.2011.1314] [Medline: 22392663]
12. Cimino JJ, Frisse ME, Halamka J, Sweeney L, Yasnoff W. Consumer-mediated health information exchanges: the 2012 ACMI debate. J Biomed Inform 2014 Apr;48:5-15 [FREE Full text] [doi: 10.1016/j.jbi.2014.02.009] [Medline: 24561078]
13. 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]
14. Gordon WJ, Catalini C. Blockchain Technology for Healthcare: Facilitating the Transition to Patient-Driven Interoperability. Comput Struct Biotechnol J 2018;16:224-230 [FREE Full text] [doi: 10.1016/j.csbj.2018.06.003] [Medline: 30069284]
15. Zhang P, White J, Schmidt DC, Lenz G, Rosenbloom ST. FHIRChain: Applying Blockchain to Securely and Scalably Share Clinical Data. Comput Struct Biotechnol J 2018;16:267-278 [FREE Full text] [doi: 10.1016/j.csbj.2018.07.004] [Medline: 30108685]
16. Murphy DR, Satterly T, Rogith D, Sittig DF, Singh H. Barriers and facilitators impacting reliability of the electronic health record-facilitated total testing process. Int J Med Inform 2019 Jul;127:102-108. [doi: 10.1016/j.ijmedinf.2019.04.004] [Medline: 31128821]
17. Tanwar S, Parekh K, Evans R. Blockchain-based electronic healthcare record system for healthcare 4.0 applications. Journal of Information Security and Applications 2020 Feb;50:102407. [doi: 10.1016/j.jisa.2019.102407]
18. Dagher GG, Mohler J, Milojkovic M, Marella PB. Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology. Sustainable Cities and Society 2018 May;39:283-297. [doi: 10.1016/j.scs.2018.02.014]
19. Zhang A, Lin X. Towards Secure and Privacy-Preserving Data Sharing in e-Health Systems via Consortium Blockchain. J Med Syst 2018 Jun 28;42(8):140. [doi: 10.1007/s10916-018-0995-5] [Medline: 29956061]
20. Cao S, Zhang G, Liu P, Zhang X, Neri F. Cloud-assisted secure eHealth systems for tamper-proofing EHR via blockchain. Information Sciences 2019 Jun;485:427-440. [doi: 10.1016/j.ins.2019.02.038]
21. Nakamoto S. Bitcoin: A peer-to-peer electronic cash system. Bitcoin: Open Source P2P Money. 2008. URL: https://bitcoin.org/bitcoin.pdf [accessed 2021-04-23]
22. Ferdous MS, Chowdhury MJM, Hoque MA. A survey of consensus algorithms in public blockchain systems for crypto-currencies. Journal of Network and Computer Applications 2021 May;182:103035. [doi: 10.1016/j.jnca.2021.103035]
23. 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]
24. McGhin T, Choo KR, Liu CZ, He D. Blockchain in healthcare applications: Research challenges and opportunities. Journal of Network and Computer Applications 2019 Jun;135:62-75. [doi: 10.1016/j.jnca.2019.02.027]
25. Vazirani AA, O'Donoghue O, Brindley D, Meinert E. Implementing Blockchains for Efficient Health Care: Systematic Review. J Med Internet Res 2019 Feb 12;21(2):e12439 [FREE Full text] [doi: 10.2196/12439] [Medline: 30747714]
26. Hussien HM, Yasin SM, Udzir SNI, Zaidan AA, Zaidan BB. A Systematic Review for Enabling of Develop a Blockchain Technology in Healthcare Application: Taxonomy, Substantially Analysis, Motivations, Challenges, Recommendations and Future Direction. J Med Syst 2019 Sep 14;43(10):320. [doi: 10.1007/s10916-019-1445-8] [Medline: 31522262]

27. Azaria A, Ekblaw A, Vieira T, Lippman A. MedRec: Using blockchain for medical data access and permission management. 2016 Presented at: 2016 2nd International Conference on Open and Big Data (OBD); August 22-24; Vienna, Austria p. 25-30. [doi: 10.1109/obd.2016.11]
28. Yue X, Wang H, Jin D, Li M, Jiang W. Healthcare Data Gateways: Found Healthcare Intelligence on Blockchain with Novel Privacy Risk Control. J Med Syst 2016 Oct;40(10):218. [doi: 10.1007/s10916-016-0574-6] [Medline: 27565509]
29. Roehrs A, da Costa CA, da Rosa Righi R. OmniPHR: A distributed architecture model to integrate personal health records. J Biomed Inform 2017 Jul;71:70-81 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.012] [Medline: 28545835]
30. Ichikawa D, Kashiyama M, Ueno T. Tamper-Resistant Mobile Health Using Blockchain Technology. JMIR Mhealth Uhealth 2017 Jul 26;5(7):e111 [FREE Full text] [doi: 10.2196/mhealth.7938] [Medline: 28747296]
31. Mannaro K, Baralla G, Pinna A, Ibba S. A Blockchain Approach Applied to a Teledermatology Platform in the Sardinian Region (Italy). Information 2018 Feb 23;9(2):44. [doi: 10.3390/info9020044]
32. Ji Y, Zhang J, Ma J, Yang C, Yao X. BMPLS: Blockchain-Based Multi-level Privacy-Preserving Location Sharing Scheme for Telecare Medical Information Systems. J Med Syst 2018 Jun 30;42(8):147. [doi: 10.1007/s10916-018-0998-2] [Medline: 29961160]
33. Kleinaki A, Mytis-Gkometh P, Drosatos G, Efraimidis PS, Kaldoudi E. A Blockchain-Based Notarization Service for Biomedical Knowledge Retrieval. Comput Struct Biotechnol J 2018;16:288-297 [FREE Full text] [doi: 10.1016/j.csbj.2018.08.002] [Medline: 30181840]
34. Jamil F, Hang L, Kim K, Kim D. A Novel Medical Blockchain Model for Drug Supply Chain Integrity Management in a Smart Hospital. Electronics 2019 May 07;8(5):505. [doi: 10.3390/electronics8050505]
35. Patel V. A framework for secure and decentralized sharing of medical imaging data via blockchain consensus. Health Informatics J 2019 Dec;25(4):1398-1411 [FREE Full text] [doi: 10.1177/1460458218769699] [Medline: 29692204]
36. Jamil F, Ahmad S, Iqbal N, Kim D. Towards a Remote Monitoring of Patient Vital Signs Based on IoT-Based Blockchain Integrity Management Platforms in Smart Hospitals. Sensors (Basel) 2020 Apr 13;20(8):2195 [FREE Full text] [doi: 10.3390/s20082195] [Medline: 32294989]
37. Dubovitskaya A, Baig F, Xu Z, Shukla R, Zambani PS, Swaminathan A, et al. ACTION-EHR: Patient-Centric Blockchain-Based Electronic Health Record Data Management for Cancer Care. J Med Internet Res 2020 Aug 21;22(8):e13598 [FREE Full text] [doi: 10.2196/13598] [Medline: 32821064]
38. Hasselgren A, Kralevska K, Gligoroski D, Pedersen SA, Faxvaag A. Blockchain in healthcare and health sciences-A scoping review. Int J Med Inform 2020 Feb;134:104040 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.104040] [Medline: 31865055]
39. Fang HSA, Tan TH, Tan YFC, Tan CJM. Blockchain Personal Health Records: Systematic Review. J Med Internet Res 2021 Apr 13;23(4):e25094 [FREE Full text] [doi: 10.2196/25094] [Medline: 33847591]
40. Baumgart I, Heep B, Krause S. OverSim: A flexible overlay network simulation framework. 2007 Presented at: 2007 IEEE Global Internet Symposium; May 11; Anchorage, AK, USA p. 79-84. [doi: 10.1109/gi.2007.4301435]
41. Androulaki E, Barger A, Bortnikov V, Cachin C, Christidis K, De CA, et al. Hyperledger fabric: a distributed operating system for permissioned blockchains. 2018 Presented at: Thirteenth EuroSys Conference; April 23-26; Porto, Portugal p. 1-15. [doi: 10.1145/3190508.3190538]
42. Dinh T, Wang J, Chen G, Liu R, Ooi B, Tan K. BLOCKBENCH: A framework for analyzing private blockchains. 2017 Presented at: 2017 ACM International Conference on Management of Data; May 14-19; Chicago, Illinois, USA p. 1085-1100. [doi: 10.1145/3035918.3064033]
43. Nasir Q, Qasse IA, Abu Talib M, Nassif AB. Performance Analysis of Hyperledger Fabric Platforms. Security and Communication Networks 2018 Sep 09;2018:1-14. [doi: 10.1155/2018/3976093]
44. Kuzlu M, Pipattanasomporn M, Gurses L, Rahman S. Performance analysis of a hyperledger fabric blockchain framework: throughput, latency and scalability. 2019 Presented at: 2019 IEEE International Conference on Blockchain (Blockchain); July 14-17; Atlanta, GA, USA p. 536-540. [doi: 10.1109/blockchain.2019.00003]
45. Liang X, Zhao J, Shetty S, Liu J, Li D. Integrating blockchain for data sharingcollaboration in mobile healthcare applications. 2017 Presented at: IEEE 28th Annual International Symposium on Personal, Indoor, Mobile Radio Communications (PIMRC); Oct 8-13; Montreal, QC, Canada p. 8-13. [doi: 10.1109/pimrc.2017.8292361]
46. Chainpoint: Blockchain Proof and Anchoring Standard. URL: https://chainpoint.org/ [accessed 2020-04-20]
47. Liang X, Shetty S, Zhao J, Bowden D, Li D, Liu J. Towards decentralized accountability and self-sovereignty in healthcare systems. 2017 Presented at: Information and Communications Security; Apr 4-6; Irbid, Jordan p. 387-398. [doi: 10.1007/978-3-319-89500-0_34]
48. Pires R, Pasin M, Felber P, Fetzer C. Secure content-based routing using intel software guard extensions. 2016 Presented at: 17th International Middleware Conference; Dec 12-16; Trento, Italy p. 1-10. [doi: 10.1145/2988336.2988346]
49. Tierion: Blockchain Proof Engine. URL: https://tierion.com/ [accessed 2020-04-20]
50. Uddin MA, Stranieri A, Gondal I, Balasubramanian V. Continuous Patient Monitoring With a Patient Centric Agent: A Block Architecture. IEEE Access 2018;6:32700-32726. [doi: 10.1109/access.2018.2846779]
51. Buterin V. Ethereum White Paper. 2014. URL: https://blockchainlab.com/pdf/Ethereum_white_paper-a_next_generation_smart_contract_and_decentralized_application_platform-vitalik-buterin.pdf [accessed 2021-05-11]

52. Omar AA, Bhuiyan MZA, Basu A, Kiyomoto S, Rahman MS. Privacy-friendly platform for healthcare data in cloud based on blockchain environment. Future Generation Computer Systems 2019 Jun;95:511-521. [doi: 10.1016/j.future.2018.12.044]
53. Roehrs A, da Costa CA, da Rosa Righi R, da Silva VF, Goldim JR, Schmidt DC. Analyzing the performance of a blockchain-based personal health record implementation. J Biomed Inform 2019 Apr;92:103140 [FREE Full text] [doi: 10.1016/j.jbi.2019.103140] [Medline: 30844481]
54. Apache Kafka. URL: https://kafka.apache.org/ [accessed 2021-05-11]
55. Apache ZooKeeper. URL: https://zookeeper.apache.org/ [accessed 2021-05-11]
56. Lee H, Kung H, Udayasankaran JG, Kijsanayotin B, B Marcelo A, Chao LR, et al. An Architecture and Management Platform for Blockchain-Based Personal Health Record Exchange: Development and Usability Study. J Med Internet Res 2020 Jun 09;22(6):e16748 [FREE Full text] [doi: 10.2196/16748] [Medline: 32515743]
57. Zheng Z, Xie S, Dai H, Chen X, Wang H. An overview of blockchain technology: Architecture, consensus, and future trends. 2017 Presented at: 2017 IEEE International Congress on Big Data (BigData Congress); Jun 25-30; Honolulu, HI, USA p. 557-564. [doi: 10.1109/bigdatacongress.2017.85]
58. Khan MA, Salah K. IoT security: Review, blockchain solutions, and open challenges. Future Generation Computer Systems 2018 May;82:395-411. [doi: 10.1016/j.future.2017.11.022]
59. Miller A. Permissioned and permissionless blockchains. In: Shetty S, Kamhoua C, Njilla L, editors. Blockchain for Distributed Systems Security. Hoboken: John Wiley & Sons, Inc; 2019:193-204.
60. Helliar CV, Crawford L, Rocca L, Teodori C, Veneziani M. Permissionless and permissioned blockchain diffusion. International Journal of Information Management 2020 Oct;54:102136. [doi: 10.1016/j.ijinfomgt.2020.102136]
61. Wüst K, Gervais A. Do you need a blockchain? 2018 Presented at: Crypto Valley Conference on Blockchain Technology (CVCBT); Jun 20-22; Zug, Switzerland p. 45-54. [doi: 10.1109/cvcbt.2018.00011]
62. Polge J, Robert J, Le Traon Y. Permissioned blockchain frameworks in the industry: A comparison. ICT Express 2021 Jun;7(2):229-233. [doi: 10.1016/j.icte.2020.09.002]
63. Wood G. Ethereum: a secure decentralised generalised transaction ledger. EIP-150 Revision. 2014. URL: https://gavwood.com/paper.pdf [accessed 2021-06-23]
64. A blockchain platform for the enterprise. Hyperledger-Fabricdocs Master Documentation. URL: https://hyperledger-fabric.readthedocs.io/en/release-2.2/ [accessed 2020-08-06]
65. Hyperledger Fabric SDK for Node.js Index. URL: https://hyperledger.github.io/fabric-sdk-node/release-2.2/index.html [accessed 2020-08-06]
66. Welcome to hyperledger fabric CA (certificate authority). Hyperledger-Fabric-Cadocs Master Documentation. URL: https://hyperledger-fabric-ca.readthedocs.io/en/release-1.4/ [accessed 2020-08-06]
67. Kwon T. Privacy preservation with X.509 standard certificates. Information Sciences 2011 Jul;181(13):2906-2921. [doi: 10.1016/j.ins.2011.02.016]
68. Hyperledger Fabric SDK for Node.js Index. URL: https://hyperledger.github.io/fabric-chaincode-node/release-2.2/api/index.html [accessed 2020-08-06]
69. BigchainDB 2.0 The Blockchain Database. BigchainDB - Whitepaper. 2018. URL: https://www.bigchaindb.com/whitepaper/bigchaindb-whitepaper.pdf [accessed 2021-06-23]
70. Gyorodi C, Gyorodi R, Pecherle G, Olah A. A comparative study: MongoDB vs MySQL. 2015 Presented at: 13th International Conference on Engineering of Modern Electric Systems; Jul 11-12; Oradea, Romania p. 1-6. [doi: 10.1109/emes.2015.7158433]
71. Zyskind G, Nathan O, Pentland AS. Decentralizing privacy: Using blockchain to protect personal data. 2015 Presented at: 2015 IEEE Security Privacy Workshops; May 21-22; San Jose, CA, USA p. 180-184. [doi: 10.1109/spw.2015.27]
72. Burr W. Selecting the Advanced Encryption Standard. IEEE Secur. Privacy 2003 Mar;1(2):43-52. [doi: 10.1109/msecp.2003.1193210]
73. Frankel S, Glenn R, Kelly S. The AES-CBC cipher algorithm and its use with IPsec. hjp: doc: RFC 3602. 2003. URL: https://www.hjp.at/doc/rfc/rfc3602.html [accessed 2021-08-23]
74. Hyperledger Caliper. URL: https://hyperledger.github.io/caliper/ [accessed 2020-08-06]
75. Hyperledger blockchain performance metrics white paper. Hyperledger Foundation. URL: https://www.hyperledger.org/learn/publications/blockchain-performance-metrics [accessed 2020-08-06]
76. Corbett JC, Hochschild P, Hsieh W, Kanthak S, Kogan E, Li H, et al. Spanner. ACM Trans Comput Syst 2013 Aug 01;31(3):1-22. [doi: 10.1145/2518037.2491245]
77. Stonebraker M, Madden S, Abadi D, Harizopoulos S, Hachem N, Helland P. The end of an architectural era: it's time for a complete rewrite. In: Making Databases Work: the Pragmatic Wisdom of Michael Stonebraker. New York: ACM; 2018:463-489.
78. Kish LJ, Topol EJ. Unpatients-why patients should own their medical data. Nat Biotechnol 2015 Sep;33(9):921-924. [doi: 10.1038/nbt.3340] [Medline: 26348958]

Abbreviations

AES: advanced encryption standard
CBC: cipher block chaining
EHR: electronic health record
HIE: health information exchange
HRI: health record information
PHR: personal health record
PII: personally identifiable information
RSI: record sharing information
SUT: system under test
tps: transactions per second

Edited by G Eysenbach; submitted 16.11.21; peer-reviewed by A Uddin, TT Kuo; comments to author 22.12.21; revised version received 14.02.22; accepted 07.03.22; published 13.04.22.

Please cite as:
Costa TBDS, Shinoda L, Moreno RA, Krieger JE, Gutierrez M
Blockchain-Based Architecture Design for Personal Health Record: Development and Usability Study
J Med Internet Res 2022;24(4):e35013
URL: https://www.jmir.org/2022/4/e35013
doi:10.2196/35013
PMID:35416782

Original Paper

Differences in Learning and Persistency Characterizing Behavior in Chronic Pain for the Iowa Gambling Task: Web-Based Laboratory-in-the-Field Study

Lili Zhang[1,2], MSc; Himanshu Vashisht[1], MSc; Alekhya Nethra[1], MSc; Brian Slattery[3], BA, PhD; Tomas Ward[1,2], PhD

[1]School of Computing, Dublin City University, Dublin, Ireland

[2]Insight Science Foundation Ireland Research Centre for Data Analytics, Dublin City University, Dublin, Ireland

[3]School of Psychology, Dublin City University, Dublin, Ireland

Corresponding Author:
Lili Zhang, MSc
School of Computing
Dublin City University
McNulty Building
Glasnevin Campus
Dublin 9
Ireland
Phone: 353 0873747816
Email: lili.zhang27@mail.dcu.ie

Abstract

Background: Chronic pain is a significant worldwide health problem. It has been reported that people with chronic pain experience decision-making impairments, but these findings have been based on conventional laboratory experiments to date. In such experiments, researchers have extensive control of conditions and can more precisely eliminate potential confounds. In contrast, there is much less known regarding how chronic pain affects decision-making captured via laboratory-in-the-field experiments. Although such settings can introduce more experimental uncertainty, collecting data in more ecologically valid contexts can better characterize the real-world impact of chronic pain.

Objective: We aim to quantify decision-making differences between individuals with chronic pain and healthy controls in a laboratory-in-the-field environment by taking advantage of internet technologies and social media.

Methods: A cross-sectional design with independent groups was used. A convenience sample of 45 participants was recruited through social media: 20 (44%) participants who self-reported living with chronic pain, and 25 (56%) people with no pain or who were living with pain for <6 months acting as controls. All participants completed a self-report questionnaire assessing their pain experiences and a neuropsychological task measuring their decision-making (ie, the Iowa Gambling Task) in their web browser at a time and location of their choice without supervision.

Results: Standard behavioral analysis revealed no differences in learning strategies between the 2 groups, although qualitative differences could be observed in the learning curves. However, computational modeling revealed that individuals with chronic pain were quicker to update their behavior than healthy controls, which reflected their increased learning rate (95% highest–posterior-density interval [HDI] 0.66-0.99) when fitted to the Values-Plus-Perseverance model. This result was further validated and extended on the Outcome-Representation Learning model as higher differences (95% HDI 0.16-0.47) between the reward and punishment learning rates were observed when fitted to this model, indicating that individuals with chronic pain were more sensitive to rewards. It was also found that they were less persistent in their choices during the Iowa Gambling Task compared with controls, a fact reflected by their decreased outcome perseverance (95% HDI −4.38 to −0.21) when fitted using the Outcome-Representation Learning model. Moreover, correlation analysis revealed that the estimated parameters had predictive value for the self-reported pain experiences, suggesting that the altered cognitive parameters could be potential candidates for inclusion in chronic pain assessments.

Conclusions: We found that individuals with chronic pain were more driven by rewards and less consistent when making decisions in our laboratory-in-the-field experiment. In this case study, it was demonstrated that, compared with standard statistical

summaries of behavioral performance, computational approaches offered superior ability to resolve, understand, and explain the differences in decision-making behavior in the context of chronic pain outside the laboratory.

(J Med Internet Res 2022;24(4):e26307) doi:10.2196/26307

KEYWORDS

chronic pain; decision-making; computational modeling; Iowa Gambling Task; lab-in-the-field experiment

Introduction

Background

Chronic pain is defined as pain persisting or reoccurring for more than 3 to 6 months [1] and has been recognized as one of the most significant health issues of the 21st century [2,3]. Approximately 100 million adults in the United States experience chronic pain, resulting in an annual cost of US $560-635 billion in medical treatment and lost productivity [4]. Worldwide, it is estimated that approximately 20% of the population lives with chronic pain [5,6]. As a result, there is significant ongoing research into understanding chronic pain and supporting people who live with this condition. A key area of research is the impact of chronic pain on cognitive or neuropsychological abilities. It has been reported that at least 20% of clinical patients with chronic pain, including those without a history of mental disorders, complain of cognitive impairments that cause significant difficulties in their social life and daily functioning [7]. In other studies, researchers have found that cognitive deficits occur across a range of pain conditions, including fibromyalgia [8], migraine [9], chronic back pain [10], and chronic neuropathic pain [11].

Although pain is an attention-demanding sensory process, cognitive alterations cannot be simply attributed to the extra attentional demand from ongoing pain [10]. Functional magnetic resonance imaging has revealed decreased gray matter density in the medial prefrontal cortex (mPFC) area [12-14] and less brain activation in cortical structures during response inhibition in patients with chronic back pain [15]. These findings are important as the mPFC also plays a critical role in other cognitive functions such as decision-making [16], executive control [17], learning [18], and memory [19]. The latest research findings have confirmed that reduced glutamate in the mPFC significantly impairs emotional and cognitive processing in people with chronic pain [20]. This suggests that chronic pain may have a negative impact on the mPFC and related neural structures and could be considered a cognitive state that may be competing with other cognitive abilities, especially those involving the mPFC such as decision-making, which is one of the cognitive domains in which individuals with chronic pain are commonly impaired.

The Iowa Gambling Task (IGT) developed by Bechara et al [21] is one of the most widely used neuropsychological paradigms for simulating complex and experience-based decision-making. Participants in this task are required to choose cards from one of 4 decks, two of which (decks A and B) are good decks and the remaining 2 (decks C and D) being bad decks. The bad decks yield negative long-term outcomes, whereas the good decks yield positive long-term outcomes. It has been successfully used to distinguish various clinical populations from healthy populations, such as patients with lesions in the ventromedial prefrontal cortex [22,23], obsessive-compulsive disorder [24], and even chronic cannabis use [25-27]. These earlier applications of the IGT found that healthy controls could learn to choose more frequently from good decks than from bad ones, whereas clinical populations tended to more regularly choose from bad decks throughout the task. With relevance to this discussion, the IGT has also been applied to investigate abnormalities in decision-making among people living with various chronic pain conditions, yielding significant findings. After extracting behavioral responses to the IGT, Apkarian et al [12] found that patients with chronic pain more frequently chose cards from bad decks, were less persistent, and exhibited a negative correlation between gambling performance outcome and reported intensity of chronic pain. The participants in the study by Verdejo-García et al [28] were required to complete both the original IGT, where reward was immediate and punishment was delayed, and an IGT variant where punishment was immediate and reward was delayed. The authors summarized their behavioral choices and found that women with fibromyalgia had significantly lower scores on the third block, which was referred to as the hunch period of the task, on the original IGT, whereas intact performance on the IGT variant suggested that these patients were hypersensitive to rewards. Similar results were obtained in the study by Tamburin et al [10], where people with chronic back pain won significantly less money relative to healthy controls, and their IGT scores did not change significantly throughout the task.

Objective

It is worth noting that all the relevant studies to date have been conducted in a laboratory setting. In these settings, the participants were under tight experimental control. No study to date has investigated decision-making tasks such as the IGT in the context of chronic pain in more natural environments where the experimenter has much less control. A laboratory-in-the-field approach is adopted in this study and, although such a setting can introduce experimental *noise* and potential confounds that may bias the results, it is closer to observing more representative behavior for this population [29]. In addition, carrying out web-based behavioral experiments is the only entirely risk-free method currently available under the typical movement restrictions imposed by the threat of COVID-19. Thus, in this study, we are interested in investigating the differences in characterizations of decision-making between individuals with chronic pain and normal controls in their everyday living environment. We are also interested in the analysis approaches that are best able to extract behavioral signals in *noisy* experimental environments. In terms of the experimental task, the participants were required to complete the pain assessments and the IGT on their web browser in an environment where they

carried out the task at a location of their choice, at a time of their choice, and without supervision. To the best of our knowledge, this is among the first research studies investigating chronic pain through internet-based technology combining decision-making tasks and self-reports.

Given the higher variability of data when collection takes place outside a laboratory setting, conventional behavioral data statistics may not be sufficient to reveal signals in the noisier data collected. In previous studies, this conventional analysis has been based on the measurement of the proportion of choices from the good decks relative to the bad decks. Furthermore, these behavioral summaries are agnostic to the underlying cognitive mechanisms that drive the behavioral performance on the IGT, thus limiting their interpretability. Therefore, in this study, we apply computational modeling analysis as a complementary approach to explicitly decompose behavioral performance on the IGT into cognitive parameters. It has been documented that estimated parameters from such models are able to reveal group differences in cognitive processes despite the absence of group differences in conventional IGT measurements [30]. The extracted parameters can then be used to understand the source of the decision-making alterations. We hypothesize that the noisier experimental environment might produce data that reveal little difference between the experimental groups when considered through conventional analysis, whereas the *filtering* enabled by the computational modeling might reveal significant differences in some cognitively interpretable parameters. Although computational modeling analyses have been successfully applied to capture the complex interplay of cognitive processes for people with a variety of other health issues [31,32], we have only found 1 reference in which a simple heuristic model was designed to differentiate the behavioral performance of individuals with chronic pain and healthy controls on the IGT [33]; thus, the changes in latent cognitive parameters that drive the impaired performance of people living with chronic pain remain unexplored until now. Given that more competitive cognitive models and more advanced parameter estimation methods have been developed to more precisely characterize the underlying cognitive mechanisms, we hypothesize that computational modeling analysis is more effective in capturing the differences in decision-making from the data set collected in a laboratory-in-the-field environment.

Methods

Recruitment

The participants in this study were recruited through social media and local pain advocacy groups. A total of 64 people, including 28 (44%) symptomatic participants who had lived with chronic pain for months and 36 (56%) healthy controls who had never lived with chronic pain or had experienced pain for <6 months, were interested in participating in the experiments. They were directed to a webpage containing the plain language statement of the experiment. After providing informed consent, they were linked to a questionnaire and the IGT through their computer or mobile phone. A total of 8 symptomatic participants and 11 healthy controls were excluded from the study as they failed to complete the IGT. Thus, a convenience sample of 45 participants (45/64, 70% of the total) was recruited finally, consisting of 20 (44%) symptomatic participants (14/20, 70% women; mean age 40, SD 12 years) and 25 (56%) healthy controls (12/25, 48% women; mean age 38, SD 12 years). The groups did not significantly differ in age (P=.43) or proportion of women (P=.07).

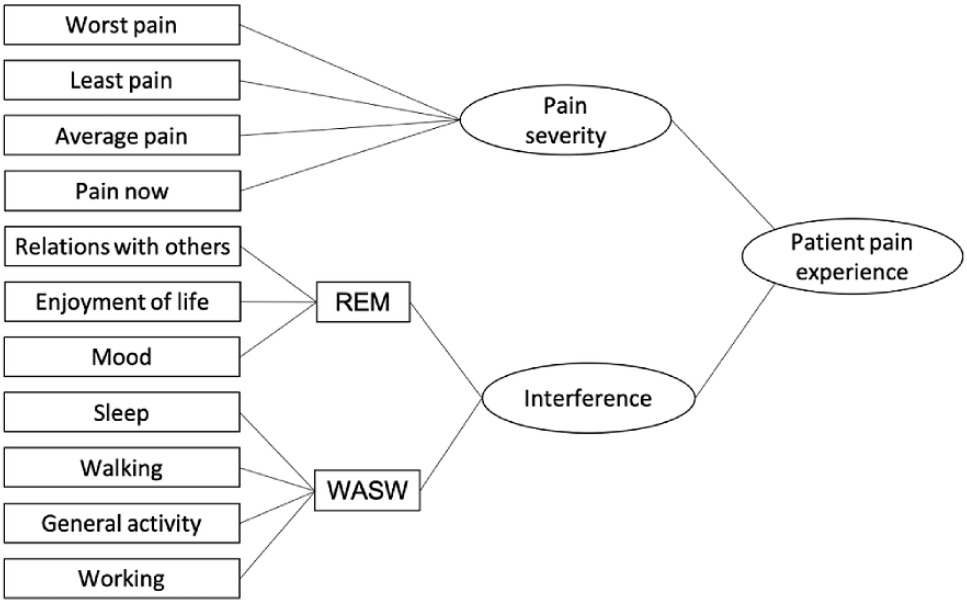
Ethics Approval

This study was approved by the local ethics committee of the School of Computing, Dublin City University (DCUREC/CA/2019/1).

Assessment of Pain Experience

The Brief Pain Inventory–Short Form (BPI-SF) is a validated, 9-item self-administered questionnaire used to evaluate the severity of the patient's pain and its impact on the patient's daily functioning. The patient is asked to rate their worst, least, average, and current pain intensity; list current treatments and their perceived effectiveness; and rate the degree to which pain interferes with general activity, mood, walking ability, normal work, relations with other persons, sleep, and enjoyment of life on a 10-point scale. The pain interference is then divided into affective subdimensions (ie, relations with others, enjoyment of life, and mood [REM]) and activity subdimensions (ie, walking, general activity, sleep, and work [WASW]). The BPI-SF has been used with a variety of populations and has been shown to be a valid and reliable measure with adequate internal reliability across these studies (eg, α=.86-.96) [34]. A graphical representation of the conceptual framework of the measurement is shown in Figure 1.

Figure 1. A graphical representation of the conceptual framework of the measurement. REM: relations with others, enjoyment of life, and mood subdimension of the Brief Pain Inventory–Short Form; WASW: walking, general activity, sleep, and work subdimension of the Brief Pain Inventory–Short Form.



The IGT Paradigm

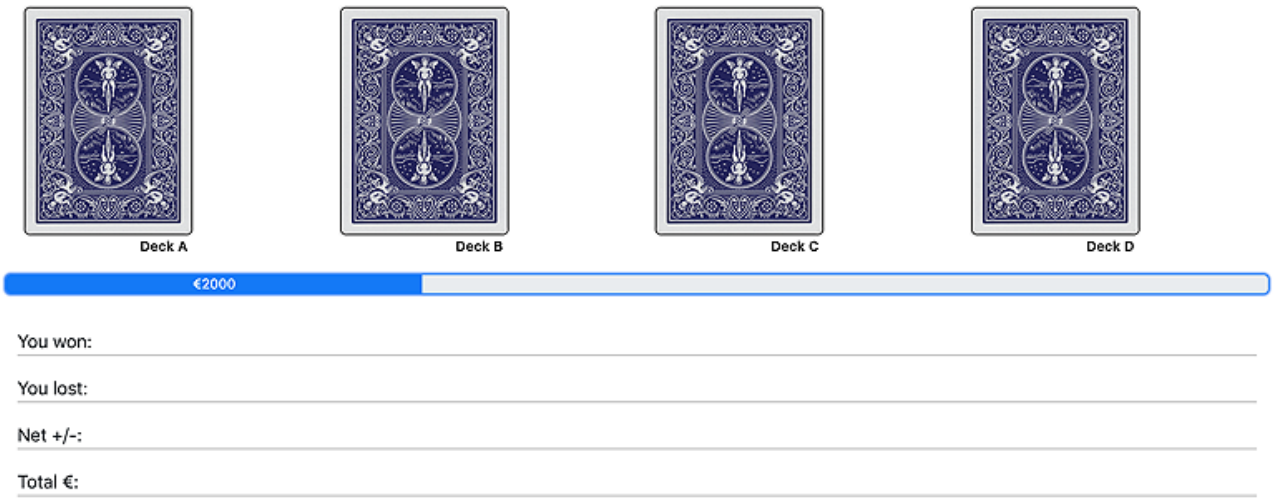
Participants in the IGT are initially given €2000 (US $2216.10) in virtual money and presented with 4 decks of cards labeled A, B, C, and D. Each card in these decks can generate gains and sometimes cause losses. Participants have to choose 1 card from these 4 decks consecutively until the task shuts off automatically after 100 trials. In each trial, feedback on the rewards and losses of their choice and the running tally over all trials so far are given to the participants, but no information is given regarding how many trials they will play and how many trials they have completed during the task. Participants are instructed that they can choose cards from any deck and switch decks at any time. They are also told to make as much money as possible, thus minimizing losses.

Table 1 shows the payoffs of the 4 decks. As can be seen in the table, decks A and B are 2 bad decks that generate high immediate, constant rewards but even higher unpredictable, occasional losses. Thus, the long-term net outcome associated with decks A and B is negative. In contrast, decks C and D are 2 good decks that generate low immediate, constant rewards but even lower unpredictable, occasional losses. Thus, the long-term net outcome associated with decks C and D is positive. In addition to the payoff magnitudes, the 4 decks also differ in the frequency of losses (ie, decks A and C are associated with a higher frequency of losses, whereas decks B and D are associated with a lower frequency of losses). The key to obtaining a higher long-term net outcome in this task is to explore all the decks in the initial stage and then exploit the 2 good decks (see Figure 2 for the screenshot of the web-based IGT that we implemented).

Table 1. Summary of the payoff of the Iowa Gambling Task.

	Deck A (bad deck with frequent losses)	Deck B (bad deck with infrequent losses)	Deck C (good deck with frequent losses)	Deck D (good deck with infrequent losses)
Reward/trial (€)	100	100	50	50
Number of loss trials/10 trials	5	1	5	1
Loss/10 trials (€)	−1250	−1250	−250	−250
Net outcome/10 trials (€)	−250	−250	250	250

Figure 2. Screenshot of the web-based Iowa Gambling Task.



Standard Behavioral Data Analysis

To compare the behavioral differences between the 2 groups in the IGT, 3 parameters were measured. First, the total amount of gain at the end of the task was calculated for each participant to measure their overall performance on the task. An unpaired 2-tailed *t* test was used to determine if the difference between the 2 groups was significant on this measure. Second, to obtain a visual exploration of the group-level deck preferences across trials, we calculated the proportions of good deck selections (decks C and D) and the learning IGT scores (ie, the difference between the number of good deck selections [decks C and D] and the number of bad deck selections [decks A and B]) across the task. Specifically, 5 new variables were created through the division of the 100 trials into 5 blocks of 20 trials each without overlap. The proportion of good deck selections and the difference between the number of good deck selections and the learning score in each block were calculated. In this way, 5 proportions and 5 learning scores, 1 for each block, were obtained for each participant. The comparison between the 5 learning scores is regarded as an index of learning. A learning score increasing from the first block to the last block suggests that a participant is developing a preference for good decks and an effective selection strategy. Given the repeated learning scores of the 2 groups over the 5 blocks of trials, a block-by-group Bayesian repeated-measure analysis of variance was performed (within-participant factor: block 1-5; between-participant factor: group healthy vs chronic pain group) to reveal whether the 2 groups differed in learning curves.

Computational Modeling Analysis

Overview

Poor performance on the IGT can be due to a variety of altered underlying neurocognitive processes such as poor learning, memory, hypersensitivity to rewards and losses, or less response consistency. To more precisely identify the psychological processes that drive participants' behavioral performances on the IGT, multiple cognitive models have been proposed, such as the Expectancy–Valence Learning model (which is also the first proposed cognitive model for the IGT) [35], the Prospect Valence Learning model with Delta (PVL-Delta) [36], the PVL model with decay (PVL-Decay) [26], and the Values-Plus-Perseverance (VPP) model [26] (these 3 models are derived from the original Expectancy–Valence Learning model but show better performance), as well as the recently proposed Outcome-Representation Learning (ORL) model [37] (see Textbox 1 for the parameter specifications of the 4 IGT models). We fitted the 4 models using the hBayesDM package in R (R Foundation for Statistical Computing) [38].

Textbox 1. Parameter specifications of the 4 Iowa Gambling Task models.

Models and their parameters

- Prospect Valence Learning (PVL)-Delta (four parameters): outcome sensitivity (α), loss aversion (λ), learning rate (A), and response consistency (c)
- PVL-Decay (four parameters): outcome sensitivity (α), loss aversion (λ), decay parameter (A), and response consistency (c)
- Values-Plus-Perseverance (eight parameters): outcome sensitivity (α), loss aversion (λ), learning rate (A), decay parameter (K), gain impact parameter (EP_p), loss impact parameter ($EP_{N\_}$), weight parameter (w), and response consistency (c)
- Outcome-Representation Learning (five parameters): reward learning rate (A_+), punishment learning rate ($A\_$), decay parameter (K), outcome frequency weight (β_F), and perseverance weight (β_p)

The PVL Models

Both the PVL-Delta and PVL-Decay models consist of a utility function, a learning rule, and an action selection rule. They are identical except that they use different learning rules. The utility function determines the weight given to gains relative to losses. Both PVL variants applied the Prospect utility function [39] that featured diminishing sensitivity to increases in magnitude and different sensitivity to losses versus gains. The utility $u(t)$ of each net outcome $x(t)$—that is, the difference between the amount of rewards and losses—in trial t is calculated as follows:

Where $u(t)$ is the subjective utility of the experienced net outcome $x(t)$, α is the outcome sensitivity parameter ($0<\alpha<1$) that controls the shape of the utility function, and λ is the loss aversion parameter ($0<\lambda<10$) that governs the sensitivity to losses relative to gains. A higher value of α suggests that individuals have greater sensitivity to feedback outcomes. A value of $\lambda<1$ indicates that individuals are more sensitive to gains than to losses, and a value of $\lambda>1$ indicates that they are more sensitive to losses than to gains.

The learning rule in the PVL models is used to update the expectancies of the decks $E(t)$ based on the subjective utility value. In the delta rule, a simplified version of the Rascorla-Wagner rule is applied, in which only the expectancy of the chosen selection is updated, whereas the expectancies for other decks remain unchanged: $E_i(t + 1) = E_j(t) + A \cdot (u(t) - E_j(t))$ (equation 2), where A is the learning rate parameter ($0 \leq A \leq 1$) that determines how much weight the decision maker gives to the recent outcomes when updating expectancies. However, in the decay rule, the expectancies of all decks are discounted in each trial except for the chosen deck, which is updated by the current outcome utility: $E_i\ (t + 1) = A \cdot E_j(t) + \delta_j(t) \cdot u(t)$ (equation 3).

Here, A is the decay parameter ($0 \leq A \leq 1$) that determines how much the past expectancy is discounted, and $\delta_j(t)$ is a dummy variable that equals 1 when deck i is chosen and 0 otherwise.

The action selection rule generates the choice possibilities $\Pr(D(t + 1) = i)$ for each deck in the next trial using a softmax function:

Where $D(t)$ is the chosen deck on trial t, θ is assumed to be trial-independent and set to $3^c - 1$, and c ($0 \leq c \leq 5$) is a response consistency parameter. A higher value of c indicates that the decision maker has a higher tendency to select choices with higher expected values, which means that they are responding more deterministically.

VPP Model

The VPP model adds a perseveration term $P_i(t)$ for the chosen deck i on trial t based on the PVL-Delta model:

K is a decay parameter that determines how much the perseveration value of each deck is discounted in each trial. The tendency to perseverate or switch is incremented each time and updated by a gain impact parameter EP_p ($-Inf < EP_p < Inf$) and a loss impact parameter EP_N ($-Inf < EP_N < Inf$) based on whether the net outcome in the previous trial was a loss or a gain. Positive values for these parameters indicate stronger tendencies for decision makers to perseverate the deck chosen in the previous trial, whereas negative values indicate switching tendencies.

The expected value and perseveration term are then integrated into a single-value signal: $V_i(t + 1) = w \cdot E_i(t + 1) + (1 - w) \cdot P_i(t + 1)$ (equation 6), where w ($0<w<1$) is a weight parameter that controls the weight given to the expected value in each trial. A greater value of w represents a greater weight given to the expected value. The values of $V_i\ (t + 1)$ are then entered into the softmax function to calculate the probabilities of each option being chosen.

ORL Model

The recently proposed ORL model assumes that people track the expected value ($EV(t)$) and the win frequency ($EF(t)$) separately. In addition, for positive and negative net outcomes, the decision makers update the expectancy of the chosen deck i with different learning rates:

Here, A_+ ($0<A_+<1$) and $A\_$ ($0<A\_<1$) are the reward and punishment learning rates, respectively, used to update the expected value of the chosen deck after rewards and punishment. The updating process for the win frequency ($EF(t)$) of the chosen option is as follows:

Here, A_+ and $A\_$ are the same learning rates as those used to update the expected value, and $sgn(x(t))$ returns 1, 0, or −1 for positive, 0, or negative outcome values on trial t, respectively. The expected outcome frequencies for unchosen decks j' are also updated in each trial, in which the learning rates are also shared from the expected value learning rule:

Here, c is the number of alternative choices for the chosen deck j, which is 3 in the case of the IGT. The ORL model also assumes that the decision makers have tendencies to stay or switch their choices regardless of the outcome in the last trial, and this tendency can be captured by a perseverance weight ($PS_i(t)$):

Here, K is the decay parameter that controls how quickly the past deck selections are forgotten and is determined by $K = 3^{K'} - 1$ (equation 11).

$K' \in [0,5]$, so $K \in [0,242]$. A single-value signal for each deck is then produced by integrating the expected value, frequency, and perseverance into a linear function: $V_i(t + 1) = EV_i(t + 1) + EF_i(t + 1) \cdot \beta_F + PS_i(t + 1) \cdot \beta_p$ (equation 12).

βF ($-Inf < \beta F < Inf$) and βp ($-Inf < \beta p < Inf$) are 2-weight parameters that reflect the weight given to the outcome

frequency and perseverance, respectively, relative to the expected value of each deck. Finally, the probability of each choice is determined by passing the expected values through the softmax function:

Results

Overview

Individuals living with chronic pain (n=20) and healthy controls (n=25) completed the web-based BPI-SF and the original version of the IGT at a location of their choice and at a time of their choice without supervision. The full demographic information is presented in Table 2.

Table 2. Demographic information for the healthy controls and people living with chronic pain (N=45).

Characteristic	Healthy controls (n=25),[a] mean (SD; range)	Participants with chronic pain (n=20),[b] mean (SD; range)	*P* value[c]
Age (years)	37.2 (12.5; 24-63)	40.2 (11.9; 23-62)	.43
Pain duration (months)	1.2 (1.3; 0-5)	78.6 (76.3; 6-264)	<.001
Pain severity	2.07 (2.5; 0-10)	4.3 (2.3; 0-8)	.004
Interference (REM[d])	2.0 (2.9; 0-9.3)	4.6 (3.3; 0-9)	.007
Interference (WASW[e])	1.9 (2.6; 0-8.3)	4.2 (2.5; 0-8.5)	.004

[a]12 females.

[b]14 females.

[c]*P* value for females was .07.

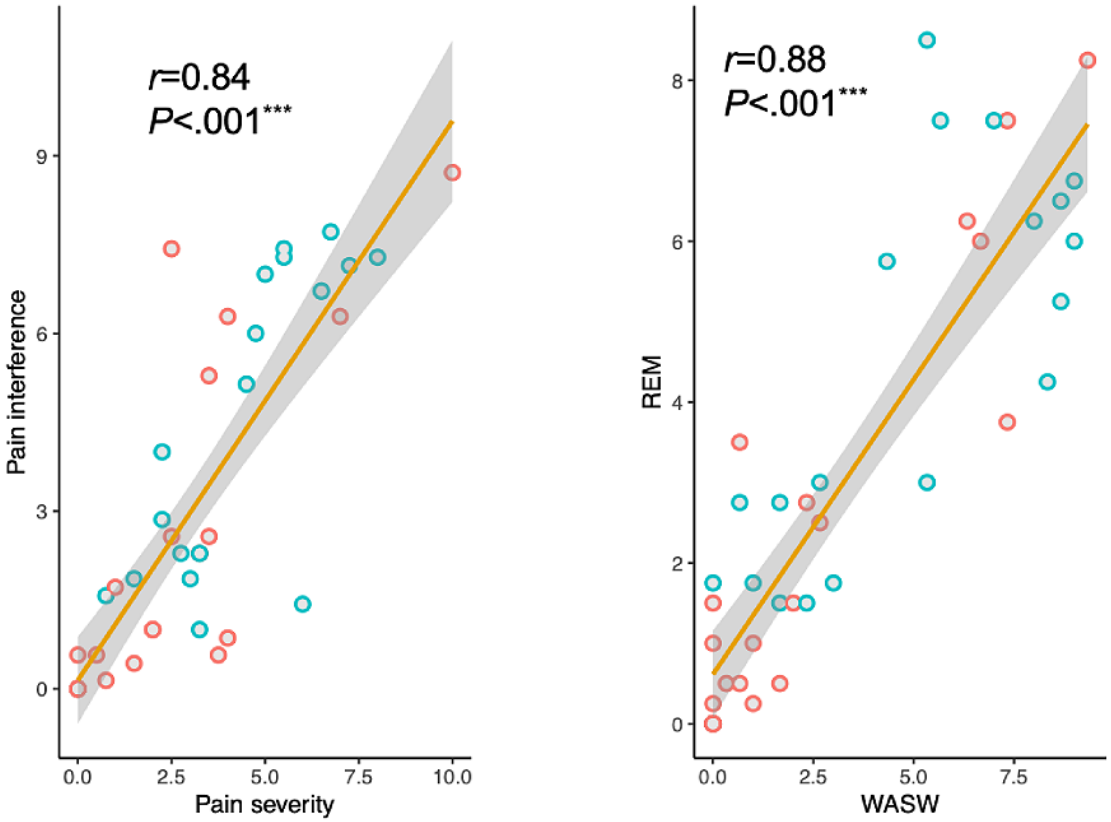
[d]REM: relations with others, enjoyment of life, and mood subdimension of the Brief Pain Inventory–Short Form.

[e]WASW: walking, general activity, sleep, and work subdimension of the Brief Pain Inventory–Short Form.

Self-report Analysis

As expected, individuals with chronic pain demonstrated higher levels of pain severity (t_{43}=–3.06; *P*=.004; Cohen *d*=–0.92, 95% CI −3.64 to −0.75), and their daily activities were more influenced by pain (t_{43}=–3.05; *P*=.004; *d*=–0.92, 95% CI −4.1 to −0.83) in comparison with healthy controls. Moreover, individuals living with chronic pain reported higher levels of subdimensional interference in REM (t_{43}=–2.86; *P*=.007; *d*=–0.86, 95% CI −4.50 to −0.78) and WASW (t_{43}=–3.02; *P*=.004; *d*=–0.91, 95% CI −3.84 to −0.77). As expected, pain severity was strongly correlated with pain interference (r_{40}=0.78, 95% CI 0.63-0.88; logBF_{10}=7.18; *P*<.001). The 2 subdimensional interferences were positively correlated (r_{40}=0.87, 95% CI 0.76-0.93; logBF_{10}=10.8; *P*<.001) as well (Figure 3).

Figure 3. Pain severity plotted against pain interference (left) and activity subdimension plotted against affective subdimension (right). Healthy controls (n=25) are plotted in red, and people living with chronic pain (n=20) are plotted in blue. The *r* values were calculated between the paired pain measures for the whole sample (*P<.5, **P<.01, ***P<.001). REM: relations with others, enjoyment of life, and mood subdimension of the Brief Pain Inventory–Short Form; WASW: walking, general activity, sleep, and work subdimension of the Brief Pain Inventory–Short Form.



Standard Behavioral Data Analysis

The total amount of gain at the end of the IGT did not significantly differ between individuals with chronic pain (€; mean 1997, SD 1187) and healthy controls (mean 1756, SD 645; t_{43}=0.81; P=.42). However, pain severity was significantly correlated with total gain (r_{43}=–0.39, 95% CI −0.62 to −0.11; P=.008; Figure 4). Figure 5 shows the proportion of choices from each deck as a function of the 5 blocks for the healthy and chronic pain groups separately and the proportion of choices from the good and bad decks. The choice pattern of the chronic pain group was qualitatively different (visual inspection of plots) from that of the healthy controls, although both groups demonstrated a clear avoidance of bad deck A. People with chronic pain showed an obvious preference for disadvantageous deck B. In contrast, healthy controls consistently favored deck D as the task progressed. Both decks B and D, which featured low-frequency losses, were generally chosen more often than decks A and C, which featured high-frequency losses. Decision makers both healthy and with chronic pain selected more cards from good decks than from bad decks at the beginning of the task. After learning whether each deck was good or bad in the second block, the healthy controls continued to select more from good decks than from bad decks. However, the choices of decision makers with chronic pain seemed to fluctuate more across the advantageous and disadvantageous decks throughout the task. The final proportion of good deck selection of healthy decision makers was higher than that of decision makers with chronic pain.

Figure 4. Pain severity plotted against total amount of gain by the end of the task. Healthy controls (n=25) are plotted in red, and people living with chronic pain (n=20) are plotted in blue. The *r* value was calculated between the pain measure and the task performance measure for the whole sample (**P*<.5, ***P*<.01, ****P*<.001).

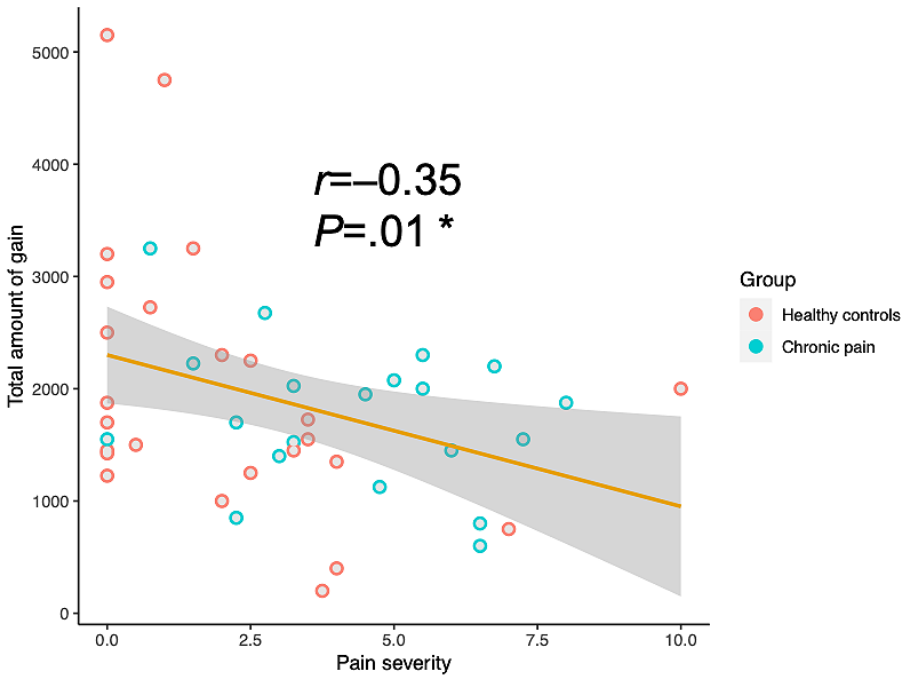


Figure 5. Mean proportion of choices from each deck within 5 blocks of both groups of decision makers (top 2 graphs) and mean proportion of choices from good decks and bad decks of both groups of decision makers (last 2 graphs). Each block contains 20 trials.

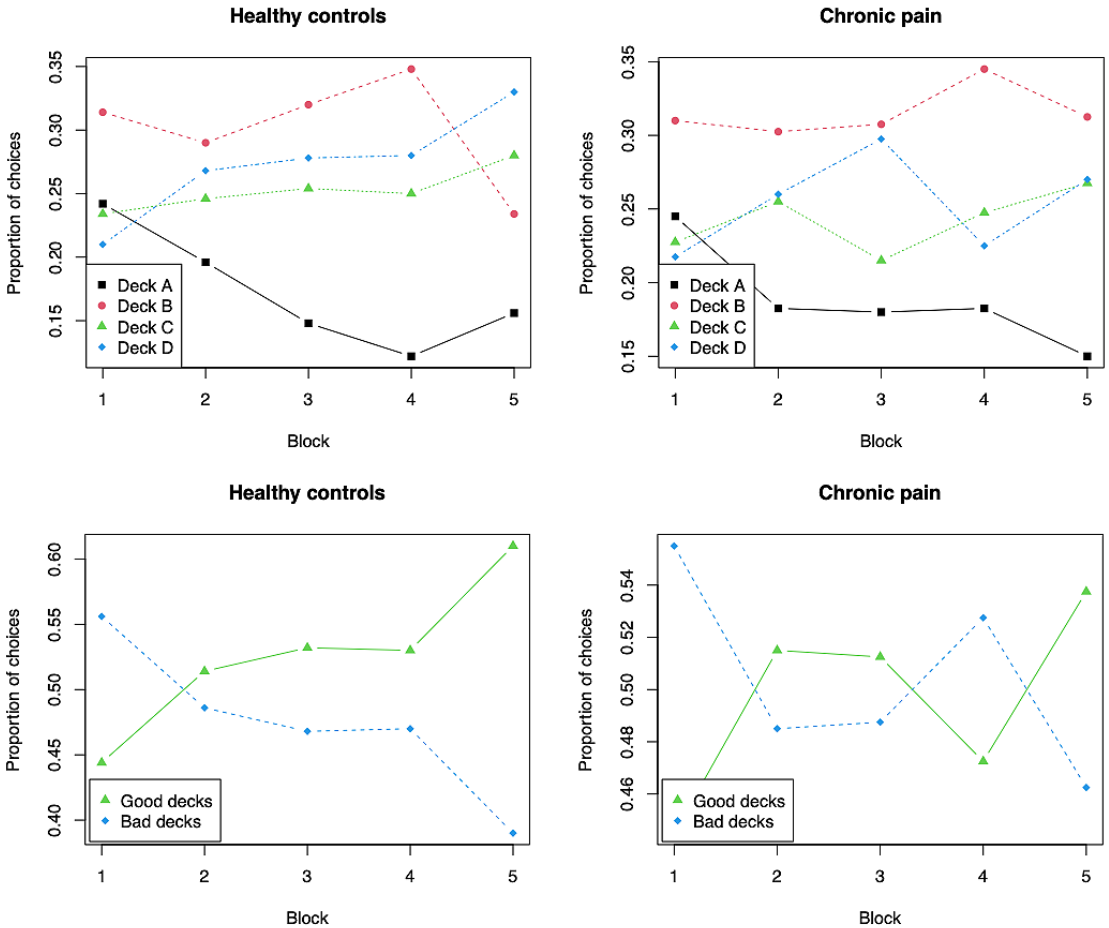


Figure 6 shows the learning scores across the 5 blocks of the IGT. A learning process was apparent in the healthy control group, in which the learning score progressively improved across the 5 blocks. Although the learning scores of individuals with chronic pain also showed an increasing trend, there was a clear dip in block 4. It is worth pointing out that the choice variances of the 2 groups in our study, as shown in Figure 6, were both higher than the variances reported in previous studies that administered the IGT in a laboratory setting [10,40]. This is evidence perhaps of the additional noise introduced by the unsupervised experimental setting. To quantify the group differences, we applied the Bayesian repeated-measure analysis of variance test in the form of a 5 (block) × 2 (health status) to the learning scores (Table 3). To our surprise, the results showed that neither the block nor the group factor had a significant impact on deck selection because the evidence in favor of the null hypothesis was 3.33:1 in favor of the alternative hypothesis that assumes an effect of group and 1.45:1 in favor of the alternative hypothesis that assumes an effect of block. This suggests that people living with chronic pain and healthy controls did not show significant deck preferences in the IGT, and neither group developed a strong learning curve during the task. We will discuss the possible implications of this observation in the *Discussion* section.

Figure 6. The learning Iowa Gambling Task (IGT) scores across the 5 different blocks of the IGT in healthy controls and people living with chronic pain.

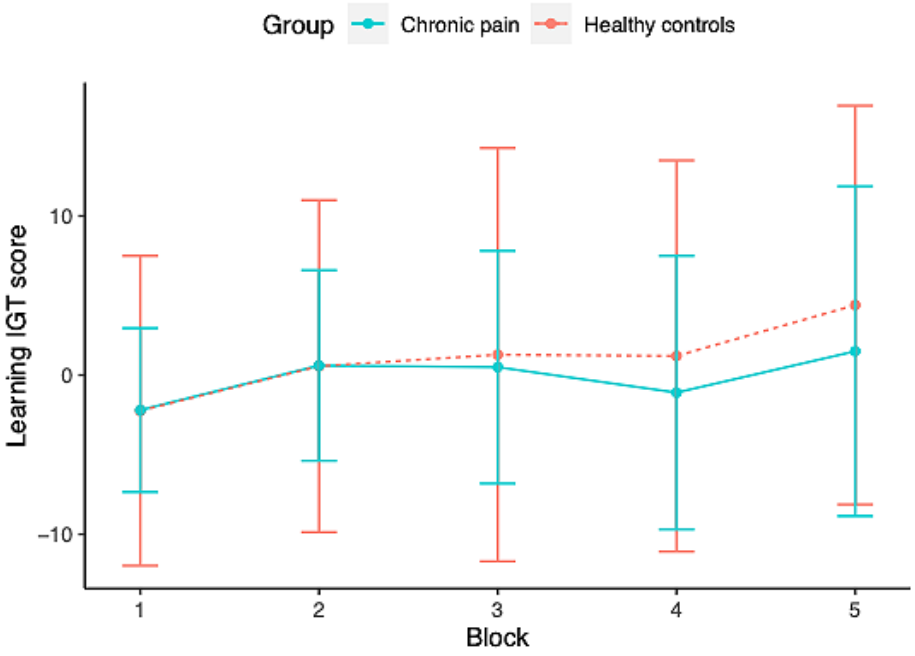


Table 3. Output of the Bayesian repeated-measure analysis of variance conducted in JASP.

Model	BF_{01}
Null model	*1.000* [a]
Group	3.33
Block	1.45
Block + group	4.65
Block + group + block × group	90.91

[a]1: no evidence; 1-3: anecdotal evidence for H0; 3-10: moderate evidence for H1; 10-30: strong evidence for H1; 30-100: very strong evidence for H1.

Computational Modeling Analysis for the IGT

Although the behavioral data statistics suggest that decision makers both healthy and with chronic pain did not show significantly different deck preferences in the IGT, there might still be group differences in the cognitive processes underlying the choices. To investigate this possibility, we decomposed the IGT performance of the 2 groups using the cognitive modeling analysis introduced earlier.

We first checked which model provided the best short-term prediction performance as measured by the one-step-ahead leave-one-out information criterion (LOOIC). The smaller a model's LOOIC score is, the better its model fit is. As shown in Table 4, the VPP model demonstrated the best overall model fit relative to other models followed by the ORL model. The ORL model ranked best in the chronic pain group.

Table 4. Models and prior fits.

Model	Pain LOOIC[a]	Healthy LOOIC	Sum LOOIC
IGT[b] ORL[c]	*4593*[d]	4600	9194
IGT VPP[e]	4623	*4544*	*9168*
IGT PVL[f]-Decay	4756	4867	9624
IGT PVL-Delta	5054	5622	10,676

[a]LOOIC: leave-one-out information criterion.

[b]IGT: Iowa Gambling Task.

[c]ORL: Outcome-Representation Learning.

[d]The smaller the value, the better the model fits the data.

[e]VPP: Values-Plus-Perseverance.

[f]PVL: Prospect Valence Learning.

Next, we used the best-fitting models (VPP and ORL) to compare the 2 groups. Figure 7 shows the posterior distributions of the group-level mean parameters of the VPP and ORL models fitted with two priors (one for each group) separately for healthy decision makers and decision makers with chronic pain. The extracted parameters for the VPP model demonstrated significantly elevated learning rates in individuals with chronic pain relative to normal controls (95% highest–posterior-density interval [HDI] 0.66-0.99). The chronic pain group also showed strong evidence of increased reward learning rate (95% HDI 0.22-0.55), punishment learning rate (95% HDI 0.03-0.11) and difference between the reward and punishment learning rate (95% HDI 0.16-0.47), and decreased decay rate (95% HDI −0.76 to −0.25) and outcome perseverance (95% HDI −4.38 to −0.21) than healthy controls when fitting to the ORL model (the 95% HDI for the comparison across groups did not overlap zero; Figure 8).

Figure 7. Group-level Values-Plus-Perseverance (top panel) and Outcome-Representation Learning (bottom panel) parameters across healthy controls and people living with chronic pain.

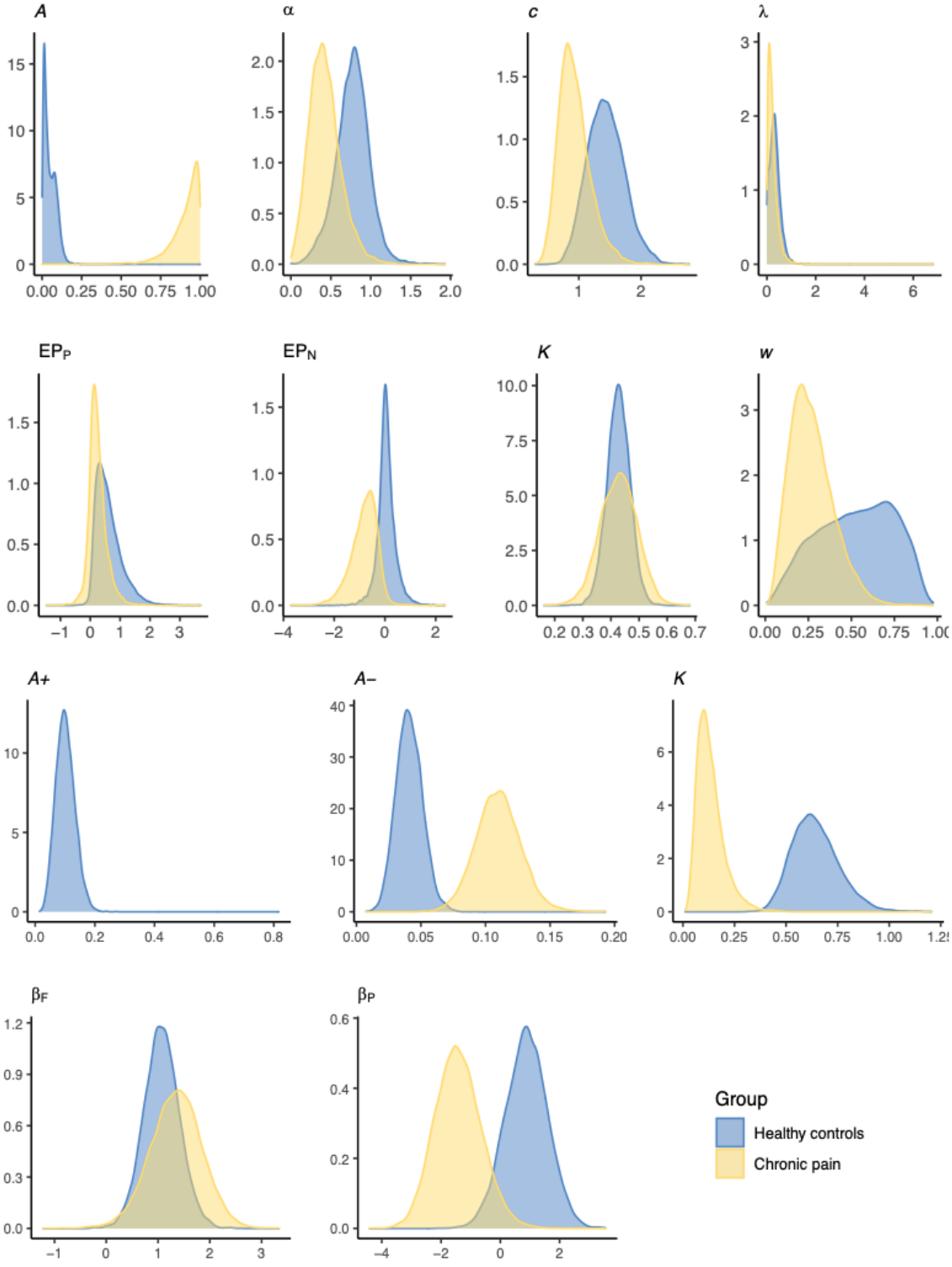
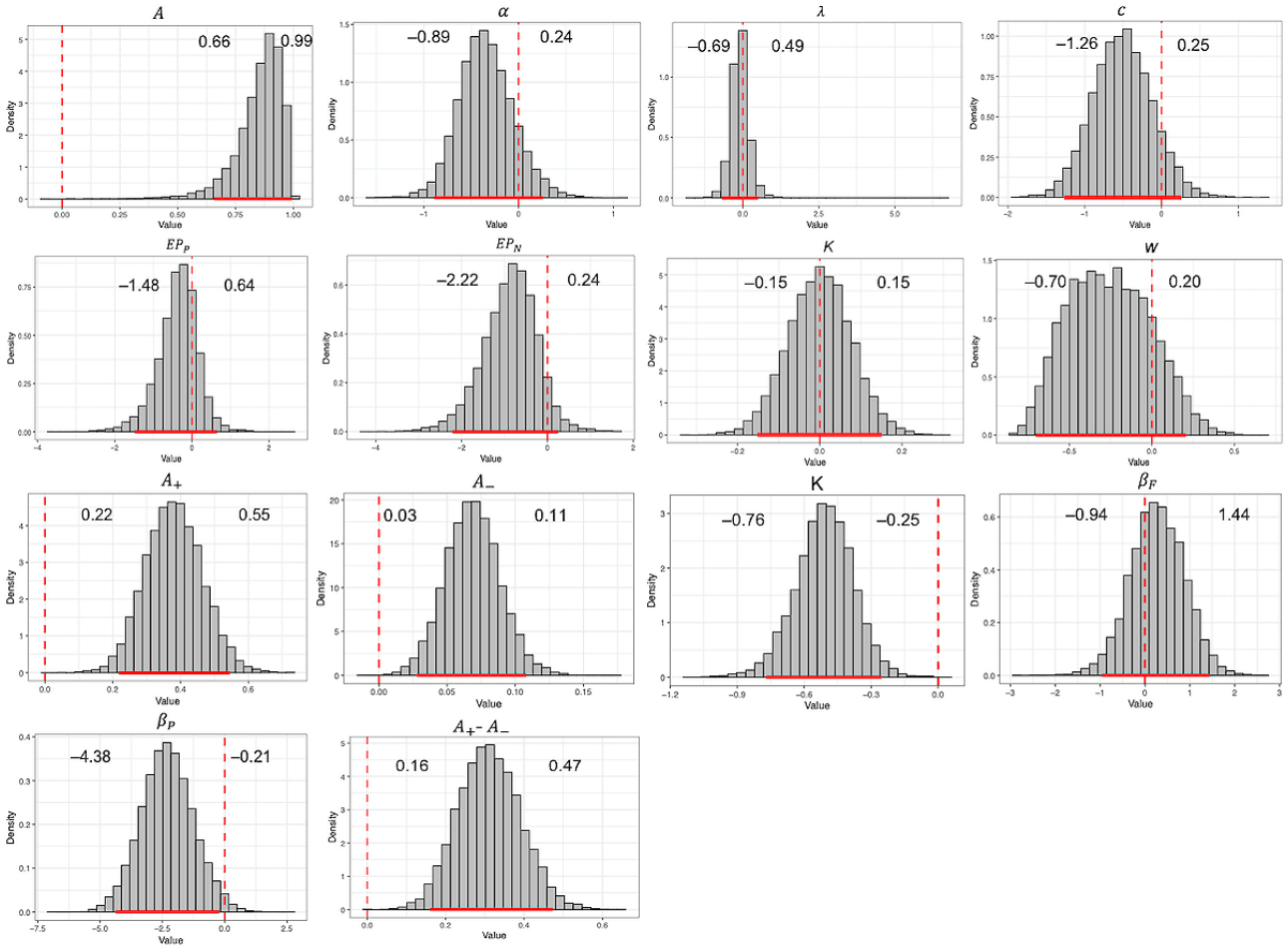


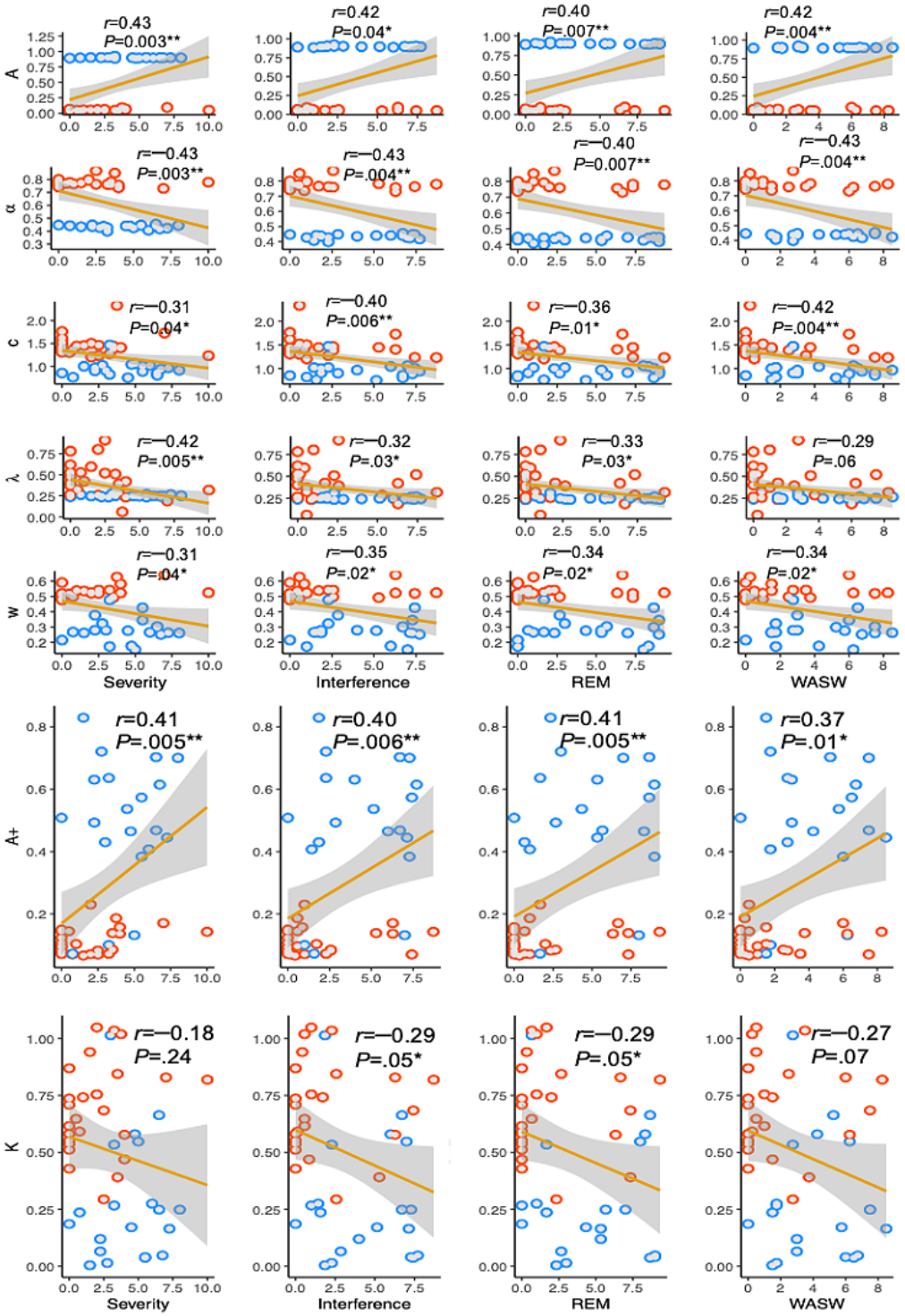
Figure 8. Differences in group-level Values-Plus-Perseverance (VPP) and Outcome-Representation Learning (ORL) parameter distribution between healthy and symptomatic groups. Solid red lines covered the 95% highest–posterior-density interval (HDI), and dashed red lines marked the 0 point. Values on the left and right sides of each graph are the lower and upper bounds of the 95% HDI of the comparison between the symptomatic and healthy control groups. If 0 point is included in the HDI, we consider there to be a nonsignificant difference between the groups.



Extracting each individual's posterior mean estimated parameters for the VPP model, shrinkage effects [41] were observed in the individual estimations of the learning rate and outcome sensitivity parameter in Figure 9, in which the individual estimations of these 2 parameters were shrunk toward the population mean. Significant evidence of positive correlations was identified between learning rate (A) and pain severity (r_{43}=0.43, 95% CI 0.15-0.64; logBF_{10}=2.48; P=.003), average pain interference (r_{43}=0.42, 95% CI 0.15-0.64; logBF_{10}=2.37; P=.04), and the 2 subdimensional interferences. However, a negative correlation was observed between the 4 pain measures and the outcome sensitivity (α), response consistency (c), loss aversion (λ), and weight parameter (w) except for WASW and loss aversion. We did not find significant correlations between the other 3 parameters in the VPP model and the pain measures.

Extracting individual estimations for the ORL model provided evidence for the existence of positive correlations between pain severity and reward learning rate (r_{43}=0.41, 95% CI 0.12-0.62; logBF_{10}=2.18; P=.005). Similar correlations were obtained between pain interference (including 2 subdimensional interferences) and this model parameter. However, a negative correlation was observed between pain interference and the decay rate parameter (r_{43}=–0.29, 95% CI −0.54 to −0.00; logBF_{10}=0.18; P=.05), but this correlation with the decay rate parameter did not apply to pain severity. A similar correlation was only identified between REM and the decay rate parameter (r_{43}=–0.29, 95% CI −0.54 to −0.00; logBF_{10}=0.18; P=.05), but no supported correlation for WASW (r_{43}=–0.27, 95% CI −0.53 to −0.02; logBF_{10}=–0.04; P=.07) was identified.

Figure 9. Estimated parameters of the Values-Plus-Perseverance model and Outcome-Representation Learning model plotted against 4 pain measures. Healthy controls are plotted in red, and patients with chronic pain are plotted in blue. The *r* values were calculated between the model parameters and pain measures for the whole sample (*P<.05, **P<.01, ***P<.001). REM: relations with others, enjoyment of life, and mood subdimension of the Brief Pain Inventory–Short Form; WASW: walking, general activity, sleep, and work subdimension of the Brief Pain Inventory–Short Form.



Discussion

Principal Findings

In this study, we explored the differences in decision-making between individuals living with chronic pain and healthy controls in a laboratory-in-the-field environment by collecting their behavioral responses to the web-based IGT. The main finding of our study is that people with chronic pain did not show significant differences in decision-making in the IGT based on standard behavioral statistics. However, further computational modeling analysis revealed that people with chronic pain had an elevated learning rate for rewards and punishments when fitting to the VPP model. Further results

were obtained when fitting the data to the ORL model, where individuals with chronic pain were more dominated by rewards. Meanwhile, the symptomatic group demonstrated a decreased decay rate and perseverance weight when fitting to the ORL model. We also explored the connection between the self-reported pain experiences, standard inferential statistics, and cognitive parameters. The main finding is that the total amount of gain at the end of the task was negatively correlated with the degree of pain severity. Moreover, several cognitive parameters could also predict pain severity and pain interference assessed by the self-reported pain experiences.

Using standard inferential statistics, the total amount of gain obtained by the end of the task was not significantly lower in people living with chronic pain than in healthy controls, indicating that there was no difference between the 2 groups in terms of the overall performance on the IGT. However, the total amount of gain was negatively correlated with the pain severity measure, suggesting that higher pain severity could be a factor that impairs performance on the IGT. Although people with chronic pain tended to select more bad decks relative to healthy controls, as seen from the learning curves, the statistical analysis identified neither group nor block as a significant factor that affected deck preference. This indicates no significant differences between the 2 groups, and both lacked evidence of significant learning across the 10 blocks during the task. These results are inconsistent with previous studies that identified significant group effects, in which patients won significantly less money and failed to adopt the advantageous decision-making strategy quickly learned by healthy controls [28,42,43]. A possible interpretation of this analysis is that the laboratory-in-the-field approach, which sacrifices experimenter control of the participants' environment, causes greater variances in participants' behavior, as revealed through their choices.

As expected, even with the relatively noisier data set, significant differences were identified in several cognitive components when comparing groups using the best-fitting models (VPP and ORL) measured by LOOIC in computational modeling analysis. The learning rate parameter in the VPP model determines how much weight is placed on past experiences of the chosen deck versus the most recent outcome from the deck. The chronic pain group demonstrated a much higher learning rate than the healthy controls, indicating that the recent outcome had a larger influence on the expectancy of the chosen deck and that forgetting was more rapid for individuals with chronic pain. The reward and punishment learning rates in the ORL model were used to update expectations after positive and negative outcomes, respectively. These 2 parameters account for the degree of the participants' sensitivity to losses and gains.

Individuals with chronic pain demonstrated both elevated reward and punishment learning rates when fitting to the ORL model, suggesting that they gave more weight to recent outcomes, which is consistent with the results obtained from the VPP model. It was suggested in the study by Haines et al [37] that comparing the differences between the reward and punishment learning rates for the 2 groups was more useful, although they were defined separately. The larger the differences between the 2 learning rates, the more the learning is dominated by either rewards or punishments. The significantly higher reward learning rates that caused larger differences between the 2 learning rates in individuals with chronic pain suggested that they were more sensitive to gains over losses relative to healthy controls. In other words, individuals with chronic pain appeared to be more driven by rewards, which could be a possible reason that made them choose more cards from deck B (as seen in Figure 5), the bad deck with a higher reward magnitude but also a higher punishment magnitude. This finding differs from the result reported in the study by Elvemo et al [44], where people with chronic pain only demonstrated significantly reduced scores on reward responsiveness but not on the self-reported tendency to pursue rewards. It can be seen from Equation (12), that the outcome frequency weight and perseverance weight parameter in the ORL model collectively influence the total value of the expected value of each deck. Values for the outcome frequency <0 or >0 indicate that decision makers prefer decks with low or high win frequency, respectively. This value was >0 for the 2 groups and did not show significant differences, suggesting that both groups preferred decks with high win frequency (ie, decks B and D), which can be reflected in the learning curves plotted in Figure 5, where healthy individuals ended up selecting deck D the most and patients with chronic pain ended up selecting deck B the most. Values for the perseverance weight <0 or >0 indicate that decision makers prefer to switch or stay with their recently chosen decks. The mean value of this parameter for the healthy controls was >0, whereas this value for the patients with chronic pain was <0. Meanwhile, there was a significant difference between the 2 groups. This means that people with chronic pain were less persistent in their previous choices during the task. In this sense, our findings are in agreement with a previous study [33] where a simple heuristic model was proposed to discriminate between patients and healthy controls on IGT performance by tuning the degree of randomness and importance given to losses and gains. Patients with chronic pain in this study demonstrated significantly less persistent behavior, which was characterized by giving more emphasis to gains than to losses and increasing decision randomness. However, contrary to our expectation, decision makers with chronic pain presented lower decay rates, suggesting that they base decisions on longer histories than healthy controls. However, substantial existing studies have consistently reported impaired memory functions in patients with chronic pain [45,46], and memory complaints are one of the most common complaints in patients with chronic pain and cognitive deficits [47]. In addition, this result is inconsistent with the findings fitting to the VPP model, where decision makers with chronic pain had higher learning rates and, therefore, relatively worse memory. It was argued in the study by Ahn et al [26] that this situation where the parameters of a model with good model fit might not correctly reflect the underlying cognitive constructs could be caused by the insufficient number of trials such that insufficient information was extracted to reliably estimate the free parameters in the model; thus, it might be helpful to perform external tests in future research [48].

When analyzing the correlations between the cognitive parameters and self-reported pain experiences, we found that the learning rate in the VPP model was positively associated with the 4 pain measures, whereas the outcome sensitivity,

response consistency, loss aversion, and weight parameter were negatively associated with the 4 pain measures. Higher reward learning rates in the ORL model could significantly predict higher self-reported pain severity and pain interference. Lower decay rates in the ORL model were only associated with higher self-reported pain interference, especially with the affective subdimension of interference. Given the correlations observed between the cognitive parameters and pain experiences, cognitive tasks might be an important tool to consider when evaluating individuals at risk of developing chronic pain conditions.

In summary, by recruiting participants on the web, administering the pain inventory and the IGT in natural environments over the web, and breaking down their performance into distinct psychological processes using computational modeling analysis, we revealed that participants with chronic pain displayed increased reward sensitivity and reduced choice persistency to their previous choices relative to healthy controls, and some of the cognitive parameters could predict the participants' pain severity and pain interference. Compared with conventional statistical analysis of behavioral performance, computational modeling analysis revealed much more evidence of distinct differences in decision-making between individuals with chronic pain and healthy controls in the noisier laboratory-in-the-field environment.

Limitations

First, only sex, age, and pain duration were collected in the demographic data, but no other measures of education or psychological status, such as anxiety and depression, were considered apart from pain severity and pain interference. It is known that experiencing chronic pain puts a person at increased risk of developing anxiety and depression disorders, and these 2 factors have been documented to have a significant influence on decision-making [49,50]. We did not consider the impact of medications on decision-making either (ie, we did have the participants' medication information but did not exclude participants who were receiving medical treatments). Another important factor that could influence decision-making is environmental distractions, such as family members, distractions in the environment, and time of the day. It is also worth noting that virtual money may not be as effective as real financial incentives for the decision-making task. However, we did achieve good adherence to and compliance with the task. Furthermore, we recruited a mixed sample of participants who might have had various kinds of pain conditions that might cause various cognitive abnormalities and, thus, have different influences on decision-making. As a result, caution must be exercised when interpreting the results obtained and generalizing them to pain populations in specific environments and particular chronic pain conditions. Future research should conduct a more comprehensive assessment of the participants and analyze the potential impacts of the aforementioned variables. Finally, a sample size of 45 could be considered relatively small, especially given that our recruitment was not a laboratory-based paradigm. This may be the cause of the analyses that revealed results that were not consistent with the existing literature. We take this smaller scale of study as pilot research to validate our research methodology. The computational methods demonstrated a great advantage in distinguishing between the underlying cognitive processes of participants with chronic pain and healthy controls in that regard.

Acknowledgments

This publication has emanated from research supported by the Science Foundation Ireland under grant SFI/12/RC/2289_P2, cofunded by the European Regional Development Fund. TW and LZ are also funded by Allied Irish Banks for this research.

Conflicts of Interest

None declared.

References

1. Merskey H. Classification of chronic pain: descriptions of chronic pain syndromes and definitions of pain terms. Pain Suppl 1986;3:1-226. [Medline: 3461421]
2. Banks S, Kerns R. Explaining high rates of depression in chronic pain: a diathesis-stress framework. Psychol Bull 1996;119(1):95-110. [doi: 10.1037/0033-2909.119.1.95]
3. Kerns RD, Jacob MC. Psychological aspects of back pain. Baillière's Clin Rheumatol 1993 Jun;7(2):337-356. [doi: 10.1016/s0950-3579(05)80093-7]
4. Simon LS. Relieving pain in America: a blueprint for transforming prevention, care, education, and research. J Pain Palliat Care Pharmacother 2012 Jul 05;26(2):197-198 [FREE Full text] [doi: 10.3109/15360288.2012.678473]
5. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. Eur J Pain 2006 May;10(4):287-333. [doi: 10.1016/j.ejpain.2005.06.009] [Medline: 16095934]
6. Goldberg DS, McGee SJ. Pain as a global public health priority. BMC Public Health 2011 Oct 06;11(1):770 [FREE Full text] [doi: 10.1186/1471-2458-11-770] [Medline: 21978149]
7. McGuire BE. Chronic pain and cognitive function. Pain 2013 Jul;154(7):964-965. [doi: 10.1016/j.pain.2013.04.008] [Medline: 23659915]

8. Luerding R, Weigand T, Bogdahn U, Schmidt-Wilcke T. Working memory performance is correlated with local brain morphology in the medial frontal and anterior cingulate cortex in fibromyalgia patients: structural correlates of pain-cognition interaction. Brain 2008 Dec;131(Pt 12):3222-3231. [doi: 10.1093/brain/awn229] [Medline: 18819988]
9. Mongini F, Keller R, Deregibus A, Barbalonga E, Mongini T. Frontal lobe dysfunction in patients with chronic migraine: a clinical-neuropsychological study. Psychiatry Res 2005 Jan 30;133(1):101-106. [doi: 10.1016/j.psychres.2003.12.028] [Medline: 15698682]
10. Tamburin S, Maier A, Schiff S, Lauriola MF, Di Rosa E, Zanette G, et al. Cognition and emotional decision-making in chronic low back pain: an ERPs study during Iowa gambling task. Front Psychol 2014;5:1350 [FREE Full text] [doi: 10.3389/fpsyg.2014.01350] [Medline: 25505440]
11. Moriarty O, Ruane N, O'Gorman D, Maharaj C, Mitchell C, Sarma K, et al. Cognitive impairment in patients with chronic neuropathic or radicular pain: an interaction of pain and age. Front Behav Neurosci 2017;11:100 [FREE Full text] [doi: 10.3389/fnbeh.2017.00100] [Medline: 28659771]
12. Apkarian AV, Sosa Y, Sonty S, Levy RM, Harden RN, Parrish TB, et al. Chronic back pain is associated with decreased prefrontal and thalamic gray matter density. J Neurosci 2004 Nov 17;24(46):10410-10415 [FREE Full text] [doi: 10.1523/JNEUROSCI.2541-04.2004] [Medline: 15548656]
13. Cauda F, Palermo S, Costa T, Torta R, Duca S, Vercelli U, et al. Gray matter alterations in chronic pain: a network-oriented meta-analytic approach. Neuroimage Clin 2014;4:676-686 [FREE Full text] [doi: 10.1016/j.nicl.2014.04.007] [Medline: 24936419]
14. Yuan C, Shi H, Pan P, Dai Z, Zhong J, Ma H, et al. Gray matter abnormalities associated with chronic back pain. Clin J Pain 2017;33(11):983-990. [doi: 10.1097/ajp.0000000000000489]
15. Glass JM, Williams DA, Fernandez-Sanchez M, Kairys A, Barjola P, Heitzeg MM, et al. Executive function in chronic pain patients and healthy controls: different cortical activation during response inhibition in fibromyalgia. J Pain 2011 Dec;12(12):1219-1229 [FREE Full text] [doi: 10.1016/j.jpain.2011.06.007] [Medline: 21945593]
16. Bechara A, Damasio AR. The somatic marker hypothesis: a neural theory of economic decision. Games Econo Behav 2005 Aug;52(2):336-372. [doi: 10.1016/j.geb.2004.06.010]
17. Posner MI, Rothbart MK, Sheese BE, Tang Y. The anterior cingulate gyrus and the mechanism of self-regulation. Cogn Affect Behav Neurosci 2007 Dec 01;7(4):391-395. [doi: 10.3758/cabn.7.4.391] [Medline: 18189012]
18. Rushworth M, Noonan M, Boorman E, Walton M, Behrens T. Frontal cortex and reward-guided learning and decision-making. Neuron 2011 Jun 23;70(6):1054-1069 [FREE Full text] [doi: 10.1016/j.neuron.2011.05.014] [Medline: 21689594]
19. Frankland PW, Bontempi B, Talton LE, Kaczmarek L, Silva AJ. The involvement of the anterior cingulate cortex in remote contextual fear memory. Science 2004 May 07;304(5672):881-883 [FREE Full text] [doi: 10.1126/science.1094804] [Medline: 15131309]
20. Naylor B, Hesam-Shariati N, McAuley JH, Boag S, Newton-John T, Rae CD, et al. Reduced glutamate in the medial prefrontal cortex is associated with emotional and cognitive dysregulation in people with chronic pain. Front Neurol 2019 Dec 3;10:1110 [FREE Full text] [doi: 10.3389/fneur.2019.01110] [Medline: 31849800]
21. Bechara A, Damasio AR, Damasio H, Anderson SW. Insensitivity to future consequences following damage to human prefrontal cortex. Cognition 1994 Apr;50(1-3):7-15. [doi: 10.1016/0010-0277(94)90018-3]
22. Bechara A, Damasio H, Damasio AR, Lee GP. Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. J Neurosci 1999 Jul 01;19(13):5473-5481. [doi: 10.1523/jneurosci.19-13-05473.1999]
23. Bechara A, Tranel D, Damasio H. Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. Brain 2000 Nov;123(11):2189-2202. [doi: 10.1093/brain/123.11.2189] [Medline: 11050020]
24. Cavedini P, Riboldi G, D'Annucci A, Belotti P, Cisima M, Bellodi L. Decision-making heterogeneity in obsessive-compulsive disorder: ventromedial prefrontal cortex function predicts different treatment outcomes. Neuropsychologia 2002 Jan;40(2):205-211. [doi: 10.1016/s0028-3932(01)00077-x]
25. Fridberg DJ, Queller S, Ahn W, Kim W, Bishara AJ, Busemeyer JR, et al. Cognitive mechanisms underlying risky decision-making in chronic cannabis users. J Math Psychol 2010 Feb 01;54(1):28-38 [FREE Full text] [doi: 10.1016/j.jmp.2009.10.002] [Medline: 20419064]
26. Ahn W, Vasilev G, Lee S, Busemeyer JR, Kruschke JK, Bechara A, et al. Decision-making in stimulant and opiate addicts in protracted abstinence: evidence from computational modeling with pure users. Front Psychol 2014 Aug 12;5:849 [FREE Full text] [doi: 10.3389/fpsyg.2014.00849] [Medline: 25161631]
27. Bechara A, Damasio H. Decision-making and addiction (part I): impaired activation of somatic states in substance dependent individuals when pondering decisions with negative future consequences. Neuropsychologia 2002 Jan;40(10):1675-1689. [doi: 10.1016/s0028-3932(02)00015-5]
28. Verdejo-García A, López-Torrecillas F, Calandre EP, Delgado-Rodríguez A, Bechara A. Executive function and decision-making in women with fibromyalgia. Arch Clin Neuropsychol 2009 Feb 11;24(1):113-122. [doi: 10.1093/arclin/acp014] [Medline: 19395361]
29. Gneezy U, Imas A. Chapter 10 - Lab in the field: measuring preferences in the wild. In: Banerjee AV, editor. Handbook of Economic Field Experiments. Amsterdam: Elsevier; Jan 2017:439-464.

30. Yechiam E, Kanz JE, Bechara A, Stout JC, Busemeyer JR, Altmaier EM, et al. Neurocognitive deficits related to poor decision making in people behind bars. Psychon Bull Rev 2008 Feb 01;15(1):44-51 [FREE Full text] [doi: 10.3758/pbr.15.1.44] [Medline: 18605478]
31. Aylward J, Valton V, Ahn W, Bond RL, Dayan P, Roiser JP, et al. Altered learning under uncertainty in unmedicated mood and anxiety disorders. Nat Hum Behav 2019 Oct 17;3(10):1116-1123 [FREE Full text] [doi: 10.1038/s41562-019-0628-0] [Medline: 31209369]
32. Huys QJ, Maia TV, Frank MJ. Computational psychiatry as a bridge from neuroscience to clinical applications. Nat Neurosci 2016 Mar 23;19(3):404-413 [FREE Full text] [doi: 10.1038/nn.4238] [Medline: 26906507]
33. Hess L, Haimovici A, Muñoz MA, Montoya P. Beyond pain: modeling decision-making deficits in chronic pain. Front Behav Neurosci 2014;8:263 [FREE Full text] [doi: 10.3389/fnbeh.2014.00263] [Medline: 25136301]
34. Mendoza T, Mayne T, Rublee D, Cleeland C. Reliability and validity of a modified Brief Pain Inventory short form in patients with osteoarthritis. Eur J Pain 2006 May;10(4):353-361. [doi: 10.1016/j.ejpain.2005.06.002] [Medline: 16051509]
35. Busemeyer J, Stout J. A contribution of cognitive decision models to clinical assessment: decomposing performance on the Bechara gambling task. Psychol Assess 2002;14(3):253-262. [doi: 10.1037/1040-3590.14.3.253]
36. Ahn W, Busemeyer J, Wagenmakers E, Stout J. Comparison of decision learning models using the generalization criterion method. Cogn Sci 2008 Dec;32(8):1376-1402 [FREE Full text] [doi: 10.1080/03640210802352992] [Medline: 21585458]
37. Haines N, Vassileva J, Ahn WY. The outcome-representation learning model: a novel reinforcement learning model of the Iowa gambling task. Cogn Sci 2018 Nov;42(8):2534-2561 [FREE Full text] [doi: 10.1111/cogs.12688] [Medline: 30289167]
38. Ahn W, Haines N, Zhang L. Revealing neurocomputational mechanisms of reinforcement learning and decision-making with the hBayesDM package. Computat Psychiatry 2017 Dec;1:24-57. [doi: 10.1162/cpsy_a_00002]
39. Kahneman D, Tversky A. Handbook of the Fundamentals of Financial Decision Making. Singapore: World Scientific Publishing Co Pte Ltd; Jun 2012:99-127.
40. Biars J, Johnson N, Nespeca M, Busch R, Kubu C, Floden D. Iowa gambling task performance in Parkinson disease patients with impulse control disorders. Arch Clin Neuropsychol Oxford Academic 2019;34(3):310-318. [doi: 10.1093/arclin/acy036]
41. Gelman A, Carlin J, Stern H, Dunson D, Vehtari A, Rubin D. Bayesian Data Analysis, Third Edition. Boca Raton, Florida: CRC Press; 2013.
42. Apkarian VA, Sosa Y, Krauss BR, Thomas SP, Fredrickson BE, Levy RE, et al. Chronic pain patients are impaired on an emotional decision-making task. Pain 2004 Mar;108(1-2):129-136. [doi: 10.1016/j.pain.2003.12.015] [Medline: 15109516]
43. Walteros C, Sánchez-Navarro JP, Muñoz MA, Martínez-Selva JM, Chialvo D, Montoya P. Altered associative learning and emotional decision making in fibromyalgia. J Psychosom Res 2011 Mar;70(3):294-301. [doi: 10.1016/j.jpsychores.2010.07.013] [Medline: 21334501]
44. Elvemo N, Landrø NI, Borchgrevink P, Håberg AK. Reward responsiveness in patients with chronic pain. Eur J Pain 2015 Nov;19(10):1537-1543 [FREE Full text] [doi: 10.1002/ejp.687] [Medline: 25766961]
45. Dick BD, Rashiq S. Disruption of attention and working memory traces in individuals with chronic pain. Anesth Analg 2007 May;104(5):1223-1229. [doi: 10.1213/01.ane.0000263280.49786.f5] [Medline: 17456678]
46. Jorge L, Gerard C, Revel M. Evidences of memory dysfunction and maladaptive coping in chronic low back pain and rheumatoid arthritis patients: challenges for rehabilitation. Eur J Phys Rehabil Med 2009 Dec;45(4):469-477 [FREE Full text] [Medline: 20032904]
47. Liu X, Li L, Tang F, Wu S, Hu Y. Memory impairment in chronic pain patients and the related neuropsychological mechanisms: a review. Acta Neuropsychiatr 2014 Aug;26(4):195-201. [doi: 10.1017/neu.2013.47] [Medline: 25279415]
48. Wallsten T, Pleskac T, Lejuez C. Modeling behavior in a clinically diagnostic sequential risk-taking task. Psychol Rev 2005;112(4):862-880. [doi: 10.1037/0033-295x.112.4.862]
49. de Siqueira AS, Flaks MK, Biella MM, Mauer S, Borges MK, Aprahamian I. Decision making assessed by the Iowa gambling task and major depressive disorder a systematic review. Dement Neuropsychol 2018;12(3):250-255 [FREE Full text] [doi: 10.1590/1980-57642018dn12-030005] [Medline: 30425788]
50. Soshi T, Nagamine M, Fukuda E, Takeuchi A. Pre-specified anxiety predicts future decision-making performances under different temporally constrained conditions. Front Psychol 2019 Jul 9;10:1544 [FREE Full text] [doi: 10.3389/fpsyg.2019.01544] [Medline: 31354572]

Abbreviations

BPI-SF: Brief Pain Inventory–Short Form
HDI: highest–posterior-density interval
IGT: Iowa Gambling Task
LOOIC: leave-one-out information criterion
mPFC: medial prefrontal cortex
ORL: Outcome-Representation Learning
PVL: Prospect Valence Learning
REM: relations with others, enjoyment of life, and mood subdimension of the Brief Pain Inventory–Short Form

VPP: Values-Plus-Perseverance
WASW: walking, general activity, sleep, and work subdimension of the Brief Pain Inventory–Short Form

Edited by G Eysenbach; submitted 08.12.20; peer-reviewed by B Thompson, JM de la Serna; comments to author 06.04.21; revised version received 14.04.21; accepted 27.12.21; published 06.04.22.

Please cite as:
Zhang L, Vashisht H, Nethra A, Slattery B, Ward T
Differences in Learning and Persistency Characterizing Behavior in Chronic Pain for the Iowa Gambling Task: Web-Based Laboratory-in-the-Field Study
J Med Internet Res 2022;24(4):e26307
URL: https://www.jmir.org/2022/4/e26307
doi:10.2196/26307
PMID:35384855

Original Paper

The Effect of a Sepsis Interprofessional Education Using Virtual Patient Telesimulation on Sepsis Team Care in Clinical Practice: Mixed Methods Study

Wei Ling Chua[1], BSc (Hons), PhD; Sim Leng Ooi[1], BSc (Hons); Gene Wai Han Chan[2,3], MBBS, MMed; Tang Ching Lau[3,4], MBBS, MMed, MMedSc; Sok Ying Liaw[1], BHSN, MHS, PhD

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

[2]Emergency Medicine Department, National University Hospital, National University Health System, Singapore, Singapore, Singapore

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

[4]Division of Rheumatology, Department of Medicine, National University Health System, Singapore, Singapore, Singapore

Corresponding Author:
Wei Ling Chua, BSc (Hons), PhD
Alice Lee Centre for Nursing Studies
National University of Singapore
Level 3, Clinical Research Centre, Block MD11
10 Medical Drive
Singapore, 117597
Singapore
Phone: 65 6516 7449
Email: chuaweiling@nus.edu.sg

Abstract

Background: Improving interprofessional communication and collaboration is necessary to facilitate the early identification and treatment of patients with sepsis. Preparing undergraduate medical and nursing students for the knowledge and skills required to assess, escalate, and manage patients with sepsis is crucial for their entry into clinical practice. However, the COVID-19 pandemic and social distancing measures have created the need for interactive distance learning to support collaborative learning.

Objective: This study aimed to evaluate the effect of sepsis interprofessional education on medical and nursing students' sepsis knowledge, team communication skills, and skill use in clinical practice.

Methods: A mixed methods design using a 1-group pretest-posttest design and focus group discussions was used. This study involved 415 undergraduate medical and nursing students from a university in Singapore. After a baseline evaluation of the participants' sepsis knowledge and team communication skills, they underwent didactic e-learning followed by virtual telesimulation on early recognition and management of sepsis and team communication strategies. The participants' sepsis knowledge and team communication skills were evaluated immediately and 2 months after the telesimulation. In total, 4 focus group discussions were conducted using a purposive sample of 18 medical and nursing students to explore their transfer of learning to clinical practice.

Results: Compared with the baseline scores, both the medical and nursing students demonstrated a significant improvement in sepsis knowledge (P<.001) and team communication skills (P<.001) in immediate posttest scores. At the 2-month follow-up, the nursing students continued to have statistically significantly higher sepsis knowledge (P<.001) and communication scores (P<.001) than the pretest scores, whereas the medical students had no significant changes in test scores between the 2-month follow-up and pretest time points (P=.99). A total of three themes emerged from the qualitative findings: greater understanding of each other's roles, application of mental models in clinical practice, and theory-practice gaps. The sepsis interprofessional education—particularly the use of virtual telesimulation—fostered participants' understanding and appreciation of each other's interprofessional roles when caring for patients with sepsis. Despite noting some incongruities with the real-world clinical practice and not encountering many sepsis scenarios in clinical settings, participants shared the application of mental models using interprofessional communication strategies and the patient assessment framework in their daily clinical practice.

Conclusions: Although the study did not show long-term knowledge retention, the use of virtual telesimulation played a critical role in facilitating the application of mental models for learning transfer and therefore could serve as a promising education modality for sepsis training. For a greater clinical effect, future studies could complement virtual telesimulation with a

mannequin-based simulation and provide more evidence on the long-term retention of sepsis knowledge and clinical skills performance.

(J Med Internet Res 2022;24(4):e35058) doi:10.2196/35058

KEYWORDS

sepsis; interprofessional education; team training; nurse-physician communication; simulation; telesimulation

Introduction

Background

Sepsis is defined as a "life-threatening organ dysfunction caused by a dysregulated host response to infection" [1]. Delays in sepsis recognition and slow initiation of treatment have been associated with poor patient outcomes [2,3]. A recent global burden of disease study on sepsis estimated approximately 48.9 million cases of sepsis with 11.0 million sepsis-related deaths worldwide in 2017 [3]. One key effort stipulated by the World Health Organization, which aims to reduce the global burden of sepsis, is to educate health care professionals on the early identification and management of sepsis [4].

As a time-critical illness, early identification, prompt escalation of care, and immediate treatment initiation for sepsis are critical to minimize patient deterioration and improve patient outcomes [3]. Although sepsis is usually managed in intensive care units, there has been an observed increase in the prevalence of sepsis in general wards [5,6]. Thus, nursing and medical graduates must have adequate knowledge and skills to recognize and initiate appropriate management of sepsis because they are typically the first contact point with patients in general wards. Hence, for undergraduate medical and nursing students' entry into clinical practice, it is important to equip them with the adequate knowledge and skillset required to assess, recognize, escalate care of, and manage patients with sepsis. However, existing literature has revealed that medical and nursing students are underprepared in relation to sepsis recognition and management, suggesting inadequate and ineffective coverage of sepsis education in undergraduate medical and nursing curricula [7-11]. Thus, there is a need to design sepsis education programs that adequately prepare medical and nursing students for entry-level practice.

Successful management of sepsis hinges not only on prompt recognition and immediate response with appropriate escalation of care to critical care if required but also on effective physician-nurse collaboration [12]. Nurses at the bedside play a key role in assessing and recognizing early signs and symptoms of sepsis and then escalating promptly for medical review, whereas junior physicians play a crucial role in considering the possibility of sepsis, initiating prompt initial management, and escalating in a timely manner for intensive care unit care. During this process, interprofessional teamwork and communication are integral to coordinating patient care and delivering timely medical treatment. Therefore, an educational approach that integrates sepsis education and interprofessional team training would further enhance health care professionals' knowledge and practice of sepsis care. Moreover, the call for interprofessional team training at the prelicensure level makes a strong case for incorporating team training elements, such as interprofessional communication and teamwork, into undergraduate interprofessional education (IPE) programs [13].

Simulation is a popular teaching method to deliver health care team–based training because it provides learners with experiential opportunities to work together as a multidisciplinary team to manage patient care and develop a shared understanding of each other's roles within a team in regard to patient care [14]. However, logistical challenges, such as conflicting schedules among learners from different professional groups, the availability of facilities and facilitators, and the high cost involved, impede the implementation of in-person simulation–based team training [15]. As such, health care educators have turned to computer learning technology to enhance cost-effectiveness and overcome the barriers to traditional in-person simulation training [16]. Furthermore, restrictions of the COVID-19 pandemic and safe distancing measures have disrupted conventional in-person simulation training [17,18]. This has resulted in an unprecedented push for the adoption of interactive distance learning techniques, such as telesimulation and virtual simulation, to remotely provide simulation-based education [18].

Retaining the experiential strengths of simulation-based learning, telesimulation utilizes both internet-based communication technology and simulation resources to provide simulation-based education when learners and facilitators are at off-site locations [19]. Currently, most telesimulations are hosted on 2D video conferencing software with webcams and screen sharing functions [18]. Facilitators simulate a patient encounter by projecting prerecorded videos of physical examination findings and patient monitors that the facilitators can control based on the clinical scenario flow and participants' actions [18]. Conversely, virtual reality simulation uses immersive technology to create a 3D computer-based simulation that mimics real-life clinical situations in a virtual environment and can also provide real-time responses according to participants' actions and decision-making [20,21]. Research has shown that virtual simulation–based team training can improve knowledge retention, clinical reasoning, teamwork attitudes, and communication skills performance, as well as learners' satisfaction with learning [22-26]. This suggests that virtual simulation is a promising learning strategy for interprofessional team–based learning.

Objectives

Given that sepsis care is intrinsically interprofessional, requiring input from various health care professionals [12], the combination of IPE and sepsis care provides dual benefits and joint synergy in teaching two important aspects of contemporary health care practice: interprofessional collaborative practice and sepsis care principles. Using a blended learning approach that

incorporates virtual telesimulation, we designed and implemented a sepsis IPE program that fulfills the needs of both practical sepsis education and IPE for undergraduate medical and nursing students. An earlier study described the integration of an IPE program into undergraduate medical and nursing curricula using the implementation science method [27]. This study aimed to (1) examine the effects of sepsis IPE using virtual telesimulation on sepsis knowledge and team communication skills of medical and nursing students and (2) describe students' perceived impact of sepsis IPE on their clinical practice.

Methods

Study Design

A mixed methods study design was used, combining a 1-group pretest-posttest design and focus group discussions (FGDs) to evaluate students' perceived effects of sepsis IPE using virtual telesimulation.

Study Setting and Participants

To substitute for in-person interprofessional learning during the COVID-19 pandemic, a virtual sepsis IPE program was integrated into undergraduate third-year nursing and fourth-year medical curricula at the National University of Singapore. A total of 96% (288/300) of medical students and 98% (293/299) of nursing students attended this compulsory program. As part of the students' learning process, all students were required to complete presimulation and postsimulation quizzes.

It was made known to the students that evaluation research would be conducted to evaluate the sepsis IPE program, and participation in the research was completely voluntary. Before starting the sepsis IPE program, a participant information sheet explaining the purpose of the research and outlining the entire research process was sent via email to all students. They were asked to provide consent for the use of their presimulation and postsimulation quiz results for the evaluation research, as well as consent to be contacted to complete a 2-month follow-up test and participate in a one-time FGD after completion of the program.

Sepsis IPE

Presimulation Activities

As part of the presimulation learning activities, the participants attended didactic e-learning on team communication skills strategies and sepsis education on their own time. The learning involved team communication skills strategies adapted from the Team Strategies and Tools to Enhance Performance and Patient Safety curriculum [28], which included the Identity, Situation, Background, Assessment and Recommendation (ISBAR) communication tool; the Concerned, Uncomfortable, and Safety (CUS) tool; and feedback to acknowledge, call out, and check back. The sepsis education adopted a case study teaching method that focused on adult sepsis pathophysiology and clinical manifestations, risk factors of sepsis, assessment of sepsis using the Airway, Breathing, Circulation, Disability, Exposure (ABCDE) approach, diagnostic and laboratory investigations for sepsis, and general management of sepsis and septic shock based on the surviving sepsis campaign guidelines [29]. References to further essential web-based readings were also provided.

Virtual Telesimulation

Table 1 summarizes the virtual telesimulation implementation process. Participants were scheduled to participate remotely in a 3-hour virtual telesimulation that we developed using the Unity 5 game engine (Unity Technologies) [30]. The design and learning activities in the sepsis virtual telesimulation were underpinned by the experiential learning theory by Kolb [31] and the theory of social constructivism by Vygotsky and Cole [32]. The experiential learning theory by Kolb [31], which constitutes the learning process of gaining knowledge from experimentation followed by reflection, informed the learning mechanisms of role-playing and debriefing in the virtual telesimulation. The approach of medical-nursing student collaborative learning through role-playing and debriefing supports the theory of social constructivism, which emphasizes learning through interpersonal interaction and discussion [32].

For each session, 2 medical students and 2 nursing students were assigned to a group, and each group included a nursing facilitator or debriefer and a simulated patient. The virtual telesimulation was implemented over a period of 4 months, from August to December 2020, with approximately 150 sessions conducted. In total, 28 clinicians who were nursing alumni of the university and had at least three years of clinical practice experience were recruited as facilitators for the simulation. Every facilitator had to undergo a one-time training session that covered the program's learning objectives, lesson plans, simulation scenarios, and facilitation and debriefing pointers and instructions on how to navigate in the virtual environment.

The virtual telesimulation required only standard computer equipment, and instructions were given to all students to install the virtual simulation software before the telesimulation. At the start of each session, the students were oriented by the facilitator on the Zoom videoconferencing software. During the orientation, the students learned how to navigate in the 3D virtual environment using their human-controlled avatar roles and perform assessments and management on the patient avatar. Both the players (ie, medical and nursing students) and the facilitator can use the computer's keyboard or mouse to freely navigate inside the virtual hospital and verbally communicate with one another and the simulated patient in real time using headsets or earphones with a microphone. Figure 1 illustrates the views presented to the different avatar roles.

Table 1. Technical and educational components of Sepsis IPE[a] virtual telesimulation.

Task and personnel	Technology	Process
Simulation orientation		
Facilitator and technical support staff	Zoom (Zoom Video Communications, Inc)	• Welcome students • Introduce team: facilitator and students • Reinforce confidentiality and ground rules • Learn how to navigate in the virtual environment using avatar roles and testing of audio system for communication • Assign students into medical-nursing pair
Simulation		
Facilitator	Unity 5 games engine	• Introduce scenario • Allow student players to read case scenario
Simulated patient	Zoom	• Responding or answering to questions
Students (in team 1 medical-nursing pair)	Unity 5 games engine	• Perform patient assessment of the patient avatar • Initiate interventions and treatments by clicking on the treatment trolley or equipment • Communicate with each other and patient avatar using headsets and clickable gestures
Debriefing with students		
Facilitators	Zoom	• Announce end of simulation scenario and instruct students to return to Zoom for debriefing • Instruct everyone to turn on video function • Engage students in reflection during scenario and from the sepsis IPE

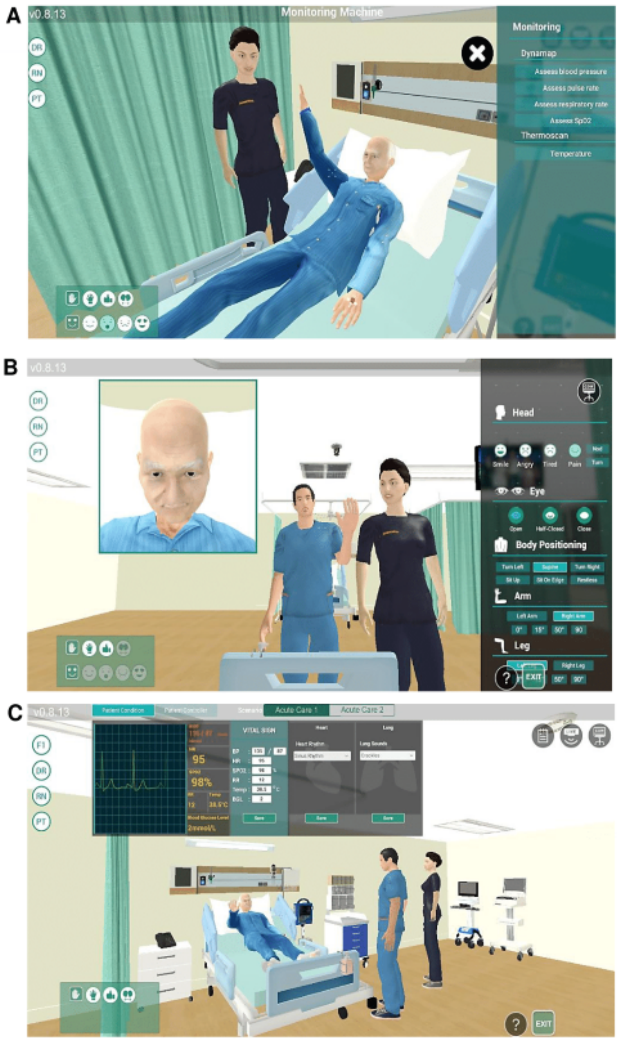
[a]IPE: interprofessional education.

The 4 students in each group were randomly paired to form 2 medical-nursing student pairs. Each medical-nursing student pair took turns as role-players and observers in the 2 simulation scenarios. The first scenario simulated a postoperative patient with early manifestations of sepsis, which required the team to perform a clinical assessment of patients with suspected sepsis and early goal-directed therapy for sepsis, including oxygen therapy, septic workup, and intravenous antibiotic therapy. The second scenario simulated the same patient who had deteriorated and required airway management and fluid resuscitation.

Before the start of the simulation scenario, each medical-nursing student pair was given 15 minutes to read the case history. Thereafter, each scenario began with the nursing student assessing the patient using the ABCDE approach and initiating immediate nursing management before escalating to their medical teammate. The medical student was expected to perform a patient assessment and collaborate with the nursing student on the treatment plans. They were also expected to use the communication strategies (eg, ISBAR, CUS, and the feedback to acknowledge, check back, and call out strategy) to communicate with each other.

Each simulation scenario lasted approximately 15 to 20 minutes and was followed by a 30-minute semistructured reflective debriefing conducted by a facilitator on the Zoom videoconferencing software. The students were asked to reflect on their performance and discuss the assessment and management of sepsis and septic shock, as well as the process of nurse-physician teamwork, communication, and collaboration. Upon the conclusion of the 2 scenarios, the participants returned to the Zoom videoconferencing software for a team debriefing. The facilitators first asked the participants about their overall thoughts and impressions of the virtual telesimulation. They then reviewed the main clinical knowledge and teamwork learning points and ended with each learner citing their major take-home points from the sepsis IPE.

Figure 1. Viewpoints of different avatar roles: (A) nursing or medical student's view, (B) simulated patient's view, and (C) facilitator's view during role-playing.



Data Collection

Both quantitative and qualitative data were obtained in this study. Whereas the quantitative data focused on knowledge evaluation and retention of the sepsis IPE, the qualitative data aimed to capture the participants' perspectives on the transferability of learning to clinical practice.

Quantitative outcomes were collected at three time points using the same set of quizzes: before (pretest), immediately after (posttest), and 2 months after (follow-up) the virtual simulation. The quiz consisting of 4 short case studies on sepsis with 25 multiple-choice questions was designed by the study team to assess sepsis knowledge (18 questions) and team communication skills (7 questions; Multimedia Appendix 1). The quiz topics were covered in the e-learning modules and virtual simulation. The quiz questions were developed with reference to the Surviving Sepsis Campaign Guidelines for Management of Sepsis and Septic Shock 2016 [33] and Team Strategies and Tools to Enhance Performance and Patient Safety curriculum [28]. The quiz was content validated by a multidisciplinary team consisting of 1 medical doctor, 2 nursing academics, and 1 advanced practice nurse, who were involved in undergraduate medical and nursing teaching programs.

Qualitative data were collected through FGDs. A total of 4 focus groups were conducted approximately 5 months after the completion of the sepsis IPE. They were conducted by a moderator on Zoom using a semistructured topic guide (Textbox 1), and field notes were taken by a research assistant. The moderator was a research fellow who did not have any previous interaction with the participants. This helped minimize any influence of the moderator on the participants during the FGDs. Each focus group consisted of 4 to 6 participants from the same profession and lasted 50 to 75 minutes. All FGDs were audio recorded.

Textbox 1. Focus group discussion topic guide.

Recall of sepsis interprofessional education (IPE) learning

- What have you learned from the sepsis IPE program?
- What are your thoughts about this interprofessional sepsis care training? In what ways were it effective?
- Which aspect of your knowledge and action skills on sepsis care do you think has improved?

Application of learning in clinical practice

- Since the end of the sepsis IPE program, have you had any opportunity to apply any of the knowledge or skills learned in clinical practice? Please share.
- Tell me about your experience and level of comfort with using the communication strategies?
- Has the program changed your way of communication with other health care professionals in clinical setting?
- What are some barriers to the application or transfer of learning to clinical practice?

Areas of improvement for future sepsis IPE program

- How can we improve future sepsis IPE program?
- Do you think it is necessary to have interdisciplinary facilitators for the virtual simulation?

Data Analysis

All quantitative data were analyzed using IBM SPSS Statistics for Windows (version 26.0; IBM Corp) [34]. Descriptive statistics were used to summarize the participants' demographics and sepsis knowledge and team communication scores at each time point. A 2-tailed, paired sample *t* test was used to examine changes between the baseline and immediate posttest scores, whereas repeated-measures ANOVA with Bonferroni correction was used to determine the effect of the intervention over the 3 time points, as well as on the respective profession group. For all analyses, the level of statistical significance was set at $P<.05$.

The audio-recorded FGDs were transcribed verbatim and subjected to the thematic analysis of Braun and Clarke [35]. Field notes were inserted into the transcripts. In addition, 2 investigators independently coded the transcripts and generated emerging themes before comparing and discussing the final set of themes and subthemes. Thematic saturation was deemed to have been achieved through the 4 FGDs.

Ethical Considerations

This study was approved by the National University of Singapore Institutional Review Board (Reference Code S-17-107). Informed consent for voluntary participation was obtained from all participants. Furthermore, participants were assured of confidentiality, anonymity, and their right to withdraw from the study at any time without any repercussions. However, they were informed that the data that had been collected until the time of their withdrawal would be retained and analyzed to enable a complete and comprehensive evaluation of the study.

Results

Effect of Sepsis IPE on Sepsis Knowledge and Team Communication

Of the 581 students who attended the program, 415 students consented to participate in the sepsis IPE research study. A total of 214 (73%) out of 293 nursing students and 201 (69.8%) out of 288 medical students who attended the program consented to participate in the research study. All 415 participants completed the presimulation and postsimulation tests. The medical students had statistically significant higher pretest scores for both sepsis and communication than nursing students (Table 2). After attending the sepsis IPE, the medical students continued to have significantly higher posttest scores than the nursing students, except for communication scores ($P=.32$). Nonetheless, as shown in Table 3, both the medical and nursing students showed significant improvements in their posttest scores for sepsis knowledge and communication after attending the sepsis IPE.

Table 2. Summary of mean scores before and after virtual simulation.

Test	Medicine (n=201), mean (SD)	Nursing (n=214), mean (SD)	*P* value
Pretest			
Total score (0-25)	16.69 (2.77)	12.67 (2.99)	<.001
Sepsis knowledge (0-18)	12.33 (2.14)	8.82 (2.22)	<.001
Communication (0-7)	4.34 (1.59)	3.85 (1.59)	.002
Posttest			
Total score (0-25)	19.09 (2.41)	16.30 (2.90)	<.001
Sepsis knowledge (0-18)	13.56 (1.68)	10.90 (2.20)	<.001
Communication (0-7)	5.53 (1.46)	5.40 (1.40)	.32

Table 3. Comparison of the pretest and posttest mean scores within the medical and nursing students.

Test	Total score (0-25)	Sepsis knowledge (0-18)	Communication (0-7)
All participants (N=415)			
Pretest, mean (SD)	14.62 (3.51)	10.52 (2.80)	4.09 (1.60)
Posttest, mean (SD)	17.66 (3.01)	12.19 (2.37)	5.47 (1.43)
P value	<.001	<.001	<.001
Nursing (n=214)			
Pretest, mean (SD)	12.67 (2.99)	8.82 (2.22)	3.85 (1.59)
Posttest, mean (SD)	16.30 (2.90)	10.90 (2.20)	5.40 (1.40)
P value	<.001	<.001	<.001
Medicine (n=201)			
Pretest, mean (SD)	16.69 (2.77)	12.33 (2.14)	4.34 (1.59)
Posttest, mean (SD)	19.09 (2.41)	13.56 (1.68)	5.53 (1.46)
P value	<.001	<.001	<.001

2-Month Follow-up Results

A total of 35% (75/214) of nursing students and 24.9% (50/201) of medical students completed the 2-month follow-up test. The Bonferroni test indicated that there were significant differences in the students' knowledge levels across time. The post hoc comparisons of the presimulation and 2-month follow-up test scores and postsimulation and 2-month follow-up test scores within each group are provided in Table 4.

At 2-month follow-up, the nursing students continued to have statistically significantly higher sepsis knowledge (mean 10.16, SD 2.42) and communication scores (mean 5.0, SD 1.55) than the pretest scores (sepsis knowledge: mean 9.21, SD 1.93; communication: mean 3.91, SD 1.56). For the medical students, there was a significant increase in sepsis knowledge test scores and communication scores from the presimulation test to the postsimulation test (sepsis knowledge: mean 12.26 SD, 1.94 vs mean 13.56, SD 1.95, *P*<.001; communication: mean 4.66 SD, 1.53 vs mean 5.60, SD 1.68, *P*<.001), but no significant changes in test scores between the 2-month follow-up and presimulation tests. Nevertheless, the medical students continued to have significantly higher sepsis knowledge scores at 2 months' follow-up than the nursing students (mean 12.38, SD 1.97 v mean 10.16, SD 2.42, *P*<.001). Although there was no significant difference in the communication scores between the medical and nursing students (*P*=.37), the change between the pretest scores and 2-month follow-up test scores was higher among the nursing students (nursing: mean 1.09, SD 1.89, *P*<.001 vs medicine: mean 0.060, SD 2.316, *P*=.86).

Table 4. Comparison of pretest-posttest scores and 2-month follow-up test scores.

Test	Nursing (n=75)			Medicine (n=50)		
	Value, mean (SD)	Follow-up, mean (SD)	*P* value	Value, mean (SD)	Follow-up, mean (SD)	*P* value
Total score						
Pretest	13.12 (2.66)	15.07 (3.05)	<.001[a]	17.18 (3.31)	17.10 (2.38)	.99[a]
Posttest	16.89 (3.13)	15.07 (3.05)	<.001[b]	19.16 (2.87)	17.10 (2.38)	.001[b]
P value	<.001[c]	N/A[d]	N/A	<.001[c]	N/A	N/A
Sepsis knowledge						
Pretest	9.21 (1.93)	10.16 (2.42)	.008[a]	12.26 (1.94)	12.38 (1.97)	.99[a]
Posttest	11.39 (2.32)	10.16 (2.42)	.001[b]	13.56 (1.95)	12.38 (1.97)	.001[b]
P value	<.001[c]	N/A	N/A	.002[c]	N/A	N/A
Communication						
Pretest	3.91 (1.56)	5.00 (1.55)	<.001[a]	4.66 (1.53)	4.72 (1.90)	.99[a]
Posttest	5.51 (1.52)	5.00 (1.55)	.04[b]	5.60 (1.68)	4.72 (1.90)	.01[b]
P value	<.001[c]	N/A	N/A	.02[c]	N/A	N/A

[a]Comparison of 2-month follow-up test scores with pretest scores.

[b]Comparison of 2-month follow-up test scores with posttest scores.

[c]Comparison of pretest scores with posttest scores.

[d]N/A: not applicable.

Effect on Clinical Practice

Overview

A total of 10 medical and 8 nursing students participated in 4 FGDs. Among the participants, there were 4 male medical students and 1 male nursing student. The thematic analysis yielded three main themes: (1) greater understanding of each other's roles, (2) application of learning in clinical settings, and (3) theory-practice gaps.

Theme 1: Greater Understanding of Each Other's Role

Generally, both the medical and nursing students agreed that the use of sepsis case scenarios was suitable for interprofessional team–based training because it provided them with a greater understanding and appreciation of each other's interprofessional roles when caring for patients with sepsis:

> *Sepsis scenario allows medical and nursing students to work together because it is a condition that requires team effort to treat the patient.* [Nursing FGD1, P2]

> *I think it [sepsis IPE learning] was a good experience to learn from our nursing colleagues. It helped me see what they [nurses] focus on and their thought process before deciding whether to escalate to the doctors and what information that they chose to pass down that they felt was important...also more of understanding the roles of both the doctors and nurses.* [Medical FGD2, P2]

Participants also stated that the use of virtual telesimulation and gamification was not only refreshing in fostering the process of interprofessional teamwork and communication in the management of septic patients but also assisted with the application of knowledge, as well as knowledge retention:

> *I thought it [sepsis virtual telesimulation] was quite fun...I learnt how to manage sepsis and how to communicate with the nurses, and then use it [acquired knowledge] in the game itself. I also got to understand what the nurses could do in their capacity, [and have a] better understanding of their roles and the team.* [Medical FGD1, P4]

> *I think having a virtual telesimulation with regards to sepsis would definitely help us to retain the information better. If without the telesimulation, we will just be reading off slides and textbooks, so the retention is definitely not as good as when we have experienced it in a virtual setting or in a real-life setting.* [Nursing FGD2, P5]

Generally, participants were satisfied with having just a nursing facilitator to facilitate the sepsis IPE. They felt that the nursing facilitators could provide insights into the interprofessional teamwork approach to the medical and nursing management of sepsis that catered to both professions:

> *My nursing facilitator is very experienced. She knows a lot as to bringing in her nursing experiences as well as some clinical value. I don't think that a medical facilitator will be essential. Because if there are many people, everyone wants to say something. Then, it becomes a bit more artificial...if we want to learn more about the sepsis and protocol, we can learn from the doctors in the wards.* [Medical FGD2, P1]

Theme 2: Application of Mental Models in Practice

During the FGDs, all participants shared their main takeaways from the program. The nursing participants gained knowledge predominantly on using the ABCDE approach to assess and manage patients with sepsis and effective nurse-physician communication strategies. Following the sepsis IPE program, the nursing students attended their transition-to-practice clinical practicum, which required them to take on the responsibilities of a full-fledged registered nurse under supervision. Although most participants did not encounter sepsis during their clinical practicum, they shared the application of the mental models, including the ABCDE assessment framework and communication strategies, in their daily clinical practice:

> *For me, learning about the ABCDE assessment part and what to do when patient deterioration happens, was main takeaway.* [Nursing FGD1, P1]
>
> *I don't think it's about [knowledge] retention, but application. I'm sure that there are some points that we can apply in other scenarios such as ABCDE assessments and communication strategies.* [Nursing FGD2, P3]

In contrast, the medical students benefited more from interprofessional teamwork and communication skills. They reported having greater awareness of the use of these communication strategies, although they cited a lack of opportunities to apply the interprofessional communication strategies in clinical settings because the nature of their clinical practicum was to *shadow* the medical team and attend case discussions:

> *The main takeaway would be the communication strategies because the callout and checkback, they are all new to me until the IPE session. Even though I haven't got to use it directly as a medical student in the ward, it made me more aware of such communication techniques.* [Medical FGD1, P4]
>
> *Experiencing on a virtual platform that there are these communication strategies, you start to observe this in the clinical settings when you see your seniors speak to nurses or other healthcare professionals. Then, you start to realise that they do use it on day-to-day basis.* [Medical FGD2, P1]

Theme 3: Theory-Practice Gaps

Despite observing the value of the ABCDE framework and communication strategies, the nursing participants highlighted a disconnect between what they were taught and what they practiced in the clinical setting. The nursing participants reported having to adapt the ABCDE and ISBAR frameworks to suit clinical workflows and dynamics. Such instances include nonadherence to the ABCDE framework of patient assessment and modification of ISBAR when escalating to the medical team:

> *They [ward nurses] know what to do first then they will continue assessing patient. The ABCDE is not always [done] in systematic order...For us [nursing students], I think the ABCDE is quite a good framework. At least you have it at the back of your mind and you know what to do next when your patient deteriorates.* [Nursing FGD1, P3]
>
> *I didn't apply the full ISBAR. I never give the background of this patient because the doctor always seems to be in a rush, and they don't really want to hear you talk so much about patient's background because they can find out themselves afterwards. So, I just give them assessment and what is happening to the patient.* [Nursing FGD1, P2]

Although both the medical and nursing students welcomed the virtual telesimulation, they also acknowledged the value of physical simulation in terms of its realism and ability for psychomotor skills practice. They suggested that virtual telesimulation should complement and be integrated into their existing curriculum for continuity:

> *I really appreciate that it was a game because it was more engaging. But I would prefer if we can also do it in real life. I think, in terms of assessing the patients, it is a bit superficial to just click on the buttons. If you do it in real life, there is more practical [hands-on] aspect.* [Medicine FGD1, P1]
>
> *Having it [virtual telesimulation] like a more continuous or regular thing rather than just a one-off thing instead.* [Nursing FGD2, P6]

Discussion

Principal Findings

To the best of our knowledge, this is the first study to evaluate the effect of a sepsis IPE program using virtual telesimulation on the sepsis knowledge and team communication skills of undergraduate medical and nursing students. Our mixed methods approach enabled the evaluation of the program at the second and third levels of Kirkpatrick's model of training evaluation [36]. Level 2, *learning*, was assessed through the quantitative sepsis knowledge and communication quiz, whereas level 3, *behavior changes*, was evaluated through FGDs on the effect of the sepsis IPE on students' clinical practice [36].

Comparison With Previous Work

In this study, we built the students' knowledge by scaffolding the knowledge base, starting with didactic e-learning, followed by a skills practice session through virtual simulated cases with debriefing sessions. Both the medical and nursing students had improved knowledge acquisition as measured by their sepsis knowledge and communication quiz scores immediately after training. This finding is analogous to those of previous studies [21,37], which observed an immediate measurable improvement in theoretical knowledge with learning through virtual simulation. Through role-playing exercises and engaging in reflective debriefing, the virtual telesimulation provided opportunities for the medical and nursing students to work together in a realistic environment and practice their skills in problem solving, decision-making, and team communication, which are essential for the management of patients with sepsis [21]. This application of experiential learning through virtual telesimulation helped deepen students' learning from

self-directed e-learning by building connections between theory and clinical practices [38].

At 2-month follow-up, both groups had lost knowledge over time. Although this is as expected and aligns well with other findings of retention effects of virtual simulation training in health education [39-41], the benefit of virtual simulation—namely, unlimited opportunities to engage in repetitive practice within a safe and realistic clinical environment [41]—was not maximized in this study. Repetition of virtual simulation sessions is key to the long-term retention of learning [42]. However, the virtual telesimulation was a one-off session in this study, which could have limited the true retention effects of virtual simulation training. Similar to an earlier study by Liaw et al [22], to some extent, the use of multiuser human-controlled avatars posed a challenge to bringing the medical and nursing students together at the same time to form interprofessional teams for regular virtual simulation sessions. The development of embodied virtual agents (EVAs; ie, physicians, nurses, simulated patients, or even facilitators) controlled by computer algorithms to allow for a single-player mode could potentially be a solution to achieve better scalability and sustainability of team-based training; however, further evaluation is needed to compare the effectiveness of EVAs (single-player mode) with human-controlled avatars (multiplayer mode) in virtual team-based simulations [22].

When compared with the preintervention test, the nursing students demonstrated significantly higher sepsis knowledge and communication scores on the 2-month follow-up test. This finding suggests that there is little knowledge decay among the nursing students. However, the medical students did not show significant differences in either sepsis or communication scores when compared with the preintervention test scores, suggesting no knowledge retention. There are 3 plausible explanations for the more positive results among the nursing students. First, medical students had higher mean pretest scores than nursing students, and it was predicted that learners with higher pretest scores would have lower learning gains than learners with lower pretest scores [43]. Second, one could theorize that opportunities for repetitive practice are crucial in enhancing knowledge retention [41]. Shortly after the sepsis IPE, the nursing students completed their final clinical practicum, in which they were expected to function as registered nurses. The requirements of their clinical practicum would have provided them with opportunities to deliberately practice the acquired knowledge, thus aiding knowledge retention. This was corroborated in both the quantitative and qualitative data, whereby the nursing participants continued to have significantly higher communication scores at 2-months follow-up from the preintervention test and reported the application of team communication strategies in their daily clinical practice. Congruently, the significant drop in communication scores at 2 months' follow-up to a mean score on par with the preintervention test among the medical students could be attributed to a lack of opportunities to practice interprofessional communication. The third reason could be related to the lack of medicine facilitators in the interprofessional sepsis team training. Although the medical students did not feel a compelling need to have a medicine facilitator, including facilitators of the respective professions (medicine or nursing) could enhance students' learning through the sharing of their respective professional perspectives and practices in real clinical practice [44,45].

Overall, although long-term sepsis knowledge retention may not be apparent in our quantitative data, our qualitative data showed that the sepsis IPE had a positive effect on students' awareness of sepsis and fostered a better understanding and appreciation of each other's interprofessional roles. This finding is consistent with those of previous studies [22,23], where significant improvements in attitudes toward health care teams and interprofessional collaboration were observed following interprofessional team–based training using virtual simulation. Similar to conventional in-person interprofessional simulation, virtual telesimulation is an experience-based learning strategy that gives learners an opportunity to experience interprofessional collaborative practice in a realistic and risk-free environment [17,20]. Although this study demonstrated that virtual telesimulation is not inferior to in-person simulations in interprofessional team training, further research is needed to evaluate this educational modality in the development of students' clinical and team communication skills.

Our method of incorporating mnemonic tools (ie, the ABCDE patient assessment framework and ISBAR communication tool) as mental models into the simulation learning was observed to facilitate the transfer of learning from the sepsis IPE to the clinical setting. Despite the lack of opportunities to practice communication skills in their clinical practice, the medical students were more cognizant of the communication strategies for effective communication within interprofessional health care teams. For the nursing students, the teaching of the systematic ABCDE assessment tool and team communication skills was applied to their everyday practice in the clinical setting. Although the nursing students noted some variability in real-world clinical practice, especially in relation to patient assessment, this inconsistency is not surprising. Instead of taking a step-by-step ABCDE approach to patient assessment, more experienced nurses tend to collect a range of focused and relevant cues to obtain a complete picture of the patient's situation because they are better able to anticipate the patient's problems [46]. Conversely, for students who lack clinical exposure, the use of mnemonic tools provides mental models to enhance learning and boost the recall of learned information, which in turn aids in the application of learning to practice [47].

Interestingly, although the value of using virtual telesimulation to deliver the sepsis IPE was well received by the students, both the medical and nursing students were in favor of an additional in-person simulation after the virtual simulation to consolidate and reinforce their learning. This is unsurprising; several studies have shown a strong preference for the kinesthetic learning style among medical and nursing students [48-52]. Despite the aforementioned merits of virtual telesimulation, it is lacking in terms of procedural skill enhancement. For a topic such as sepsis, which involves clinical procedural skills, in-person simulations provide kinesthetic learning in a realistic situation that could better develop procedural skills [53]. Thus, a blended simulation approach (ie, virtual telesimulation followed by

high-fidelity mannequin-based simulation) could optimize sepsis care learning because of its ability to provide both formative and summative assessments of knowledge and skills acquisition [38,54].

Limitations

This study had a few limitations that warrant attention. First, the effectiveness of our sepsis IPE using virtual simulation was limited by the absence of a control group and an evaluation of long-term knowledge retention that was confined to a relatively small sample size (125/415, 30.1%) of students' knowledge 2 months after the program. The high rate of loss to follow-up was likely due to the students' heavy workload, as the data collection period coincided with their clinical practicum and midterm tests. Notwithstanding, the FGDs provided some insights into the students' transfer of learning to the clinical setting. Second, the program was evaluated based on a multiple-choice quiz that assessed the lower levels of clinical competence—that is, the cognition domain of fact gathering and application of knowledge—rather than clinical performance and actual practice, which are typically evaluated using a simulation test with an assessor checklist or workplace-based assessment [55,56]. Third, our evaluation may be subject to recall or practice effect bias because we used the same set of quizzes to evaluate the postsimulation test and follow-up test. It might have been better if we had modified the set of quizzes, but to a similar difficulty level as the presimulation quiz. Furthermore, the FGDs were conducted 5 months after the simulation learning, which might have resulted in erroneous recall of responses from the participants. Finally, the true effect of the program may have been constrained by the one-time exposure to the virtual simulation experience when the intent was to provide greater access to enable students to deliberate practice opportunities.

Future Directions

Given that the virtual telesimulation was a one-off session in this study, further research with a control group is needed to determine whether exposure to repetitive virtual telesimulation training can mitigate knowledge decay over time and contribute to the long-term retention of clinical competency through high-fidelity mannequin-based simulation assessments. Further development and evaluation of EVAs controlled by computer algorithms to allow for the single-player mode as opposed to the multiplayer mode could address the desire for time flexibility, accessibility to repetitive training, and scalability to train a large number of learners. Future studies could also evaluate the effect of virtual telesimulation followed by high-fidelity mannequin-based simulation on the long-term retention of team performance on sepsis management, as well as evaluate the transfer of students' learning to clinical practice as junior house officers and new graduate nurses. On a broader context, more study is warranted to evaluate virtual telesimulation as an educational modality on clinical skills development.

Conclusions

A sepsis IPE program using a virtual simulation was developed to enhance sepsis knowledge and team communication skills among medical and nursing students. Although long-term sepsis knowledge retention was not demonstrated in this study, virtual telesimulation played a critical role in facilitating the application of knowledge and skill utilization in the clinical setting. The study also achieved one of its objectives, namely, strengthening interprofessional collaboration, whereby students fostered a better understanding and appreciation of each other's interprofessional roles in sepsis care. Future studies could complement the virtual telesimulation with a mannequin-based simulation and provide more evidence on the long-term retention of sepsis knowledge and clinical skills performance.

Acknowledgments

The authors thank all participants and nursing facilitators who facilitated the sepsis interprofessional education. We also thank Elite Editing for editing the manuscript. This study was internally funded by a Research Fellow Start Up Grant awarded by the Alice Lee Centre for Nursing Studies, Yong Loo Lin School of Medicine, National University of Singapore, Singapore.

Authors' Contributions

WLC contributed to conceptualization, methodology, investigation, formal analysis, results interpretation, and drafting and revision of the manuscript. SLO contributed to formal analysis and drafting and revision of the manuscript. GWHC contributed to results interpretation and revision of the manuscript. TCL contributed to results interpretation and revision of the manuscript. SYL contributed to conceptualization, methodology, investigation, results interpretation, and revision of the manuscript. All the authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sepsis IPE Quiz.
[PDF File (Adobe PDF File), 165 KB - jmir_v24i4e35058_app1.pdf]

References

1. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). JAMA 2016 Feb 23;315(8):801-810 [FREE Full text] [doi: 10.1001/jama.2016.0287] [Medline: 26903338]
2. Levy MM, Rhodes A, Phillips GS, Townsend SR, Schorr CA, Beale R, et al. Surviving sepsis campaign: association between performance metrics and outcomes in a 7.5-year study. Crit Care Med 2015 Jan;43(1):3-12. [doi: 10.1097/CCM.0000000000000723] [Medline: 25275252]
3. Rhee C, Jones TM, Hamad Y, Pande A, Varon J, O'Brien C, Centers for Disease Control and Prevention (CDC) Prevention Epicenters Program. Prevalence, underlying causes, and preventability of sepsis-associated mortality in US acute care hospitals. JAMA Netw Open 2019 Feb 01;2(2):e187571 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.7571] [Medline: 30768188]
4. Schlapbach LJ, Kissoon N, Alhawsawi A, Aljuaid MH, Daniels R, Gorordo-Delsol LA, et al. World Sepsis Day: a global agenda to target a leading cause of morbidity and mortality. Am J Physiol Lung Cell Mol Physiol 2020 Sep 01;319(3):L518-L522 [FREE Full text] [doi: 10.1152/ajplung.00369.2020] [Medline: 32812788]
5. Zaccone V, Tosoni A, Passaro G, Vallone CV, Impagnatiello M, Li Puma DD, Internal Medicine Sepsis Study Group. Sepsis in internal medicine wards: current knowledge, uncertainties and new approaches for management optimization. Ann Med 2017 Nov;49(7):582-592. [doi: 10.1080/07853890.2017.1332776] [Medline: 28521523]
6. Szakmany T, Lundin RM, Sharif B, Ellis G, Morgan P, Kopczynska M, Welsh Digital Data Collection Platform Collaborators. Sepsis prevalence and outcome on the general wards and emergency departments in wales: results of a multi-centre, observational, point prevalence study. PLoS One 2016;11(12):e0167230 [FREE Full text] [doi: 10.1371/journal.pone.0167230] [Medline: 27907062]
7. Harley A, Massey D, Ullman AJ, Reid-Searl K, Schlapbach LJ, Takashima M, et al. Final year nursing student's exposure to education and knowledge about sepsis: a multi-university study. Nurse Educ Today 2021 Feb;97:104703. [doi: 10.1016/j.nedt.2020.104703] [Medline: 33360011]
8. Datta R, Di Tanna GL, Harley A, Schlapbach LJ, Nunnink L, Venkatesh B. An assessment of knowledge and education about sepsis among medical students: a multi-university survey. Crit Care Resusc 2021 Mar 1;23(1):117-118. [doi: 10.51893/2021.1.rl2]
9. Valičević G, Friganović A, Kurtović B, Rotim C, Ledinski Fičko S, Krupa S. Knowledge of sepsis in nursing students-a cross-sectional study. Int J Environ Res Public Health 2021 Nov 26;18(23):12443 [FREE Full text] [doi: 10.3390/ijerph182312443] [Medline: 34886169]
10. Stefani S, Trisyani Y, Setyawati A. The knowledge of nursing internship program students about early detection of sepsis. Open Access Maced J Med Sci 2021 Nov 17;9(T6):116-121. [doi: 10.3889/oamjms.2021.7602]
11. Marshall-Brown P, Namboya F, Pollach G. Evaluating sepsis training for medical students and nonphysicians in Malawi. J Clin Anesth 2016 Nov;34:352-357. [doi: 10.1016/j.jclinane.2016.05.013] [Medline: 27687409]
12. Chung H, Medina D, Fox-Robichaud A. Interprofessional sepsis education module: a pilot study. CJEM 2016 Mar;18(2):143-145. [doi: 10.1017/cem.2015.42] [Medline: 26063312]
13. Nelson S, White CF, Hodges BD, Tassone M. Interprofessional team training at the prelicensure level: a review of the literature. Acad Med 2017 May;92(5):709-716. [doi: 10.1097/ACM.0000000000001435] [Medline: 27782914]
14. Fernandez R, Shah S, Rosenman ED, Kozlowski SW, Parker SH, Grand JA. Developing team cognition: a role for simulation. Simul Healthc 2017 Apr;12(2):96-103 [FREE Full text] [doi: 10.1097/SIH.0000000000000200] [Medline: 28704287]
15. Liaw SY, Zhou WT, Lau TC, Siau C, Chan SW. An interprofessional communication training using simulation to enhance safe care for a deteriorating patient. Nurse Educ Today 2014 Feb;34(2):259-264. [doi: 10.1016/j.nedt.2013.02.019] [Medline: 23518067]
16. Sarker MN, Wu M, Cao Q, Alam GM, Li D. Leveraging digital technology for better learning and education: a systematic literature review. Int J Inf Educ Technol 2019;9(7):453-461. [doi: 10.18178/ijiet.2019.9.7.1246]
17. Ray JM, Wong AH, Yang TJ, Buck S, Joseph M, Bonz JW, et al. Virtual telesimulation for medical students during the COVID-19 pandemic. Acad Med 2021 Oct 01;96(10):1431-1435 [FREE Full text] [doi: 10.1097/ACM.0000000000004129] [Medline: 33883398]
18. Thomas A, Burns R, Sanseau E, Auerbach M. Tips for conducting telesimulation-based medical education. Cureus 2021 Jan 04;13(1):e12479 [FREE Full text] [doi: 10.7759/cureus.12479] [Medline: 33552792]
19. McCoy CE, Sayegh J, Alrabah R, Yarris LM. Telesimulation: an innovative tool for health professions education. AEM Educ Train 2017 Apr;1(2):132-136 [FREE Full text] [doi: 10.1002/aet2.10015] [Medline: 30051023]
20. Coyne E, Calleja P, Forster E, Lin F. A review of virtual-simulation for assessing healthcare students' clinical competency. Nurse Educ Today 2021 Jan;96:104623. [doi: 10.1016/j.nedt.2020.104623] [Medline: 33125979]
21. Liaw SY, Carpio GA, Lau Y, Tan SC, Lim WS, Goh PS. Multiuser virtual worlds in healthcare education: a systematic review. Nurse Educ Today 2018 Jun;65:136-149. [doi: 10.1016/j.nedt.2018.01.006] [Medline: 29571002]
22. Liaw SY, Ooi SW, Rusli KD, Lau TC, Tam WW, Chua WL. Nurse-physician communication team training in virtual reality versus live simulations: randomized controlled trial on team communication and teamwork attitudes. J Med Internet Res 2020 Apr 08;22(4):e17279 [FREE Full text] [doi: 10.2196/17279] [Medline: 32267235]

23. Liaw SY, Wu LT, Soh SL, Ringsted C, Lau TC, Lim WS. Virtual reality simulation in interprofessional round training for health care students: a qualitative evaluation study. Clin Simul Nurs 2020 Aug;45:42-46. [doi: 10.1016/j.ecns.2020.03.013]
24. Liaw SY, Choo T, Wu LT, Lim WS, Choo H, Lim SM, et al. "Wow, woo, win"- Healthcare students' and facilitators' experiences of interprofessional simulation in three-dimensional virtual world: a qualitative evaluation study. Nurse Educ Today 2021 Oct;105:105018. [doi: 10.1016/j.nedt.2021.105018] [Medline: 34175564]
25. Chheang V, Fischer V, Buggenhagen H, Huber T, Huettl F, Kneist W, et al. Toward interprofessional team training for surgeons and anesthesiologists using virtual reality. Int J Comput Assist Radiol Surg 2020 Dec;15(12):2109-2118 [FREE Full text] [doi: 10.1007/s11548-020-02276-y] [Medline: 33083969]
26. Khanal P, Vankipuram A, Ashby A, Vankipuram M, Gupta A, Drumm-Gurnee D, et al. Collaborative virtual reality based advanced cardiac life support training simulator using virtual reality principles. J Biomed Inform 2014 Oct;51:49-59 [FREE Full text] [doi: 10.1016/j.jbi.2014.04.005] [Medline: 24732098]
27. Liaw SY, Ooi SL, Mildon R, Ang EN, Lau TC, Chua WL. Translation of an evidence-based virtual reality simulation-based interprofessional education into health education curriculums: an implementation science method. Nurse Educ Today 2022 Mar;110:105262. [doi: 10.1016/j.nedt.2021.105262] [Medline: 35063778]
28. King HB, Battles J, Baker DP, Alonso A, Salas E, Webster J, et al. TeamSTEPPS™: team strategies and tools to enhance performance and patient safety. In: Henriksen K, Battles JB, Keyes MA, Grady ML, editors. Advances in patient safety: new directions and alternative approaches (Volume 3: performance and tools). Rockville, MD: Agency for Healthcare Research and Quality (US); 2008.
29. Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. Intensive Care Med 2021 Nov;47(11):1181-1247 [FREE Full text] [doi: 10.1007/s00134-021-06506-y] [Medline: 34599691]
30. Unity. Unity Technologies. 2021. URL: https://unity.com/ [accessed 2021-09-11]
31. Kolb DA. Experiential learning: experience as the source of learning and development. Hoboken, NJ, USA: Prentice-Hall; 1984.
32. Vygotsky LS, Cole M. Mind in society: development of higher psychological processes. Cambridge, MA, USA: Harvard University Press; 1978.
33. Rhodes A, Evans LE, Alhazzani W, Levy MM, Antonelli M, Ferrer R, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016. Intensive Care Med 2017 Mar;43(3):304-377. [doi: 10.1007/s00134-017-4683-6] [Medline: 28101605]
34. IBM SPSS statistics for Windows (Version 26.0). IBM Corp. 2019. URL: https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-26 [accessed 2021-09-11]
35. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
36. Kirkpatrick JD, Kirkpatrick WK. Kirkpatrick's four levels of training evaluation. Alexandria, VA, USA: Association for Talent Development Press; 2016.
37. Rourke S. How does virtual reality simulation compare to simulated practice in the acquisition of clinical psychomotor skills for pre-registration student nurses? A systematic review. Int J Nurs Stud 2020 Feb;102:103466. [doi: 10.1016/j.ijnurstu.2019.103466] [Medline: 31783192]
38. Liaw SY, Wong LF, Chan SW, Ho JT, Mordiffi SZ, Ang SB, et al. Designing and evaluating an interactive multimedia web-based simulation for developing nurses' competencies in acute nursing care: randomized controlled trial. J Med Internet Res 2015 Jan 12;17(1):e5 [FREE Full text] [doi: 10.2196/jmir.3853] [Medline: 25583029]
39. Courteille O, Fahlstedt M, Ho J, Hedman L, Fors U, von Holst H, et al. Learning through a virtual patient vs. recorded lecture: a comparison of knowledge retention in a trauma case. Int J Med Educ 2018 Mar 28;9:86-92 [FREE Full text] [doi: 10.5116/ijme.5aa3.ccf2] [Medline: 29599421]
40. Farra S, Miller E, Timm N, Schafer J. Improved training for disasters using 3-D virtual reality simulation. West J Nurs Res 2013 May;35(5):655-671. [doi: 10.1177/0193945912471735] [Medline: 23287551]
41. Liaw SY, Chan SW, Chen FG, Hooi SC, Siau C. Comparison of virtual patient simulation with mannequin-based simulation for improving clinical performances in assessing and managing clinical deterioration: randomized controlled trial. J Med Internet Res 2014 Sep 17;16(9):e214 [FREE Full text] [doi: 10.2196/jmir.3322] [Medline: 25230684]
42. Ansquer R, Mesnier T, Farampour F, Oriot D, Ghazali DA. Long-term retention assessment after simulation-based-training of pediatric procedural skills among adult emergency physicians: a multicenter observational study. BMC Med Educ 2019 Sep 11;19(1):348 [FREE Full text] [doi: 10.1186/s12909-019-1793-6] [Medline: 31510979]
43. Colt HG, Davoudi M, Murgu S, Zamanian Rohani N. Measuring learning gain during a one-day introductory bronchoscopy course. Surg Endosc 2011 Jan;25(1):207-216 [FREE Full text] [doi: 10.1007/s00464-010-1161-4] [Medline: 20585964]
44. van Diggele C, Roberts C, Burgess A, Mellis C. Interprofessional education: tips for design and implementation. BMC Med Educ 2020 Dec 03;20(Suppl 2):455 [FREE Full text] [doi: 10.1186/s12909-020-02286-z] [Medline: 33272300]
45. Ruiz MG, Ezer H, Purden M. Exploring the nature of facilitating interprofessional learning: findings from an exploratory study. J Interprof Care 2013 Nov;27(6):489-495. [doi: 10.3109/13561820.2013.811640] [Medline: 23859380]

46. Hoffman KA, Aitken LM, Duffield C. A comparison of novice and expert nurses' cue collection during clinical decision-making: verbal protocol analysis. Int J Nurs Stud 2009 Oct;46(10):1335-1344. [doi: 10.1016/j.ijnurstu.2009.04.001] [Medline: 19555954]
47. Liaw SY, Chan SW, Scherpbier A, Rethans JJ, Pua GG. Recognizing, responding to and reporting patient deterioration: transferring simulation learning to patient care settings. Resuscitation 2012 Mar;83(3):395-398. [doi: 10.1016/j.resuscitation.2011.08.021] [Medline: 21946262]
48. Abouzeid E, Fouad S, Wasfy NF, Alkhadragy R, Hefny M, Kamal D. Influence of personality traits and learning styles on undergraduate medical students' academic achievement. Adv Med Educ Pract 2021;12:769-777 [FREE Full text] [doi: 10.2147/AMEP.S314644] [Medline: 34290542]
49. Urval RP, Kamath A, Ullal S, Shenoy AK, Shenoy N, Udupa LA. Assessment of learning styles of undergraduate medical students using the VARK questionnaire and the influence of sex and academic performance. Adv Physiol Educ 2014 Sep;38(3):216-220 [FREE Full text] [doi: 10.1152/advan.00024.2014] [Medline: 25179610]
50. Alkhasawneh E. Using VARK to assess changes in learning preferences of nursing students at a public university in Jordan: implications for teaching. Nurse Educ Today 2013 Dec;33(12):1546-1549. [doi: 10.1016/j.nedt.2012.12.017] [Medline: 23454890]
51. James S, D'Amore A, Thomas T. Learning preferences of first year nursing and midwifery students: utilising VARK. Nurse Educ Today 2011 May;31(4):417-423. [doi: 10.1016/j.nedt.2010.08.008] [Medline: 20826044]
52. Zhu HR, Zeng H, Zhang HY, Wan FJ, Guo HH, Zhang CH. The preferred learning styles utilizing VARK among nursing students with bachelor degrees and associate degrees in China. Acta Paul Enferm 2018 Mar;31(2):162-169. [doi: 10.1590/1982-0194201800024]
53. Tan AJ, Lee CC, Lin PY, Cooper S, Lau LS, Chua WL, et al. Designing and evaluating the effectiveness of a serious game for safe administration of blood transfusion: a randomized controlled trial. Nurse Educ Today 2017 Aug;55:38-44. [doi: 10.1016/j.nedt.2017.04.027] [Medline: 28521248]
54. Johnsen HM, Briseid HS, Brodtkorb K, Slettebø Å, Fossum M. Nursing students' perceptions of combining hands-on simulation with simulated patients and a serious game in preparing for clinical placement in home healthcare: a qualitative study. Nurse Educ Today 2021 Feb;97:104675 [FREE Full text] [doi: 10.1016/j.nedt.2020.104675] [Medline: 33302184]
55. Shumway JM, Harden RM, Association for Medical Education in Europe. AMEE Guide No. 25: the assessment of learning outcomes for the competent and reflective physician. Med Teach 2003 Nov;25(6):569-584. [doi: 10.1080/0142159032000151907] [Medline: 15369904]
56. Norcini JJ. Work based assessment. BMJ 2003 Apr 05;326(7392):753-755 [FREE Full text] [doi: 10.1136/bmj.326.7392.753] [Medline: 12676847]

Abbreviations

ABCDE: Airway, Breathing, Circulation, Disability, Exposure
EVA: embodied virtual agent
FGD: focus group discussion
IPE: interprofessional education
ISBAR: Identity, Situation, Background, Assessment and Recommendation

Edited by T Leung; submitted 18.11.21; peer-reviewed by A Fox-Robichaud, J Thompson, A Kononowicz; comments to author 09.01.22; revised version received 21.01.22; accepted 22.02.22; published 18.04.22.

Please cite as:
Chua WL, Ooi SL, Chan GWH, Lau TC, Liaw SY
The Effect of a Sepsis Interprofessional Education Using Virtual Patient Telesimulation on Sepsis Team Care in Clinical Practice: Mixed Methods Study
J Med Internet Res 2022;24(4):e35058
URL: https://www.jmir.org/2022/4/e35058
doi:10.2196/35058
PMID:35436237

Original Paper

Optimal Treatment Selection in Sequential Systemic and Locoregional Therapy of Oropharyngeal Squamous Carcinomas: Deep Q-Learning With a Patient-Physician Digital Twin Dyad

Elisa Tardini[1,2], BSc, MSc; Xinhua Zhang[1], BSc, MSc, PhD; Guadalupe Canahuate[3], BSc, MSc, PhD; Andrew Wentzel[1], BSc, MSc; Abdallah S R Mohamed[4,5], MSc, MD; Lisanne Van Dijk[4], PhD; Clifton D Fuller[4,5], BSc, MD, PhD; G Elisabeta Marai[1], PhD

[1]Department of Computer Science, University of Illinois at Chicago, Chicago, IL, United States

[2]Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan, Italy

[3]Department of Electrical and Computer Engineering, University Of Iowa, Iowa City, IA, United States

[4]MD Anderson Cancer Center, Houston, TX, United States

[5]Department of Radiation Oncology, The University of Texas, Austin, TX, United States

Corresponding Author:
Elisa Tardini, BSc, MSc
Department of Computer Science
University of Illinois at Chicago
851 S Morgan St
Chicago, IL, 60607
United States
Phone: 1 3127092826
Email: elisa.tardini@mail.polimi.it

Abstract

Background: Currently, selection of patients for sequential versus concurrent chemotherapy and radiation regimens lacks evidentiary support and it is based on locally optimal decisions for each step.

Objective: We aim to optimize the multistep treatment of patients with head and neck cancer and predict multiple patient survival and toxicity outcomes, and we develop, apply, and evaluate a first application of deep Q-learning (DQL) and simulation to this problem.

Methods: The treatment decision DQL digital twin and the patient's digital twin were created, trained, and evaluated on a data set of 536 patients with oropharyngeal squamous cell carcinoma with the goal of, respectively, determining the optimal treatment decisions with respect to survival and toxicity metrics and predicting the outcomes of the optimal treatment on the patient. Of the data set of 536 patients, the models were trained on a subset of 402 (75%) patients (split randomly) and evaluated on a separate set of 134 (25%) patients. Training and evaluation of the digital twin dyad was completed in August 2020. The data set includes 3-step sequential treatment decisions and complete relevant history of the patient cohort treated at MD Anderson Cancer Center between 2005 and 2013, with radiomics analysis performed for the segmented primary tumor volumes.

Results: On the test set, we found mean 87.35% (SD 11.15%) and median 90.85% (IQR 13.56%) accuracies in treatment outcome prediction, matching the clinicians' outcomes and improving the (predicted) survival rate by +3.73% (95% CI –0.75% to 8.96%) and the dysphagia rate by +0.75% (95% CI –4.48% to 6.72%) when following DQL treatment decisions.

Conclusions: Given the prediction accuracy and predicted improvement regarding the medically relevant outcomes yielded by this approach, this digital twin dyad of the patient-physician dynamic treatment problem has the potential of aiding physicians in determining the optimal course of treatment and in assessing its outcomes.

(J Med Internet Res 2022;24(4):e29455) doi:10.2196/29455

KEYWORDS

digital twin dyad; reinforcement learning; head and neck cancer

Introduction

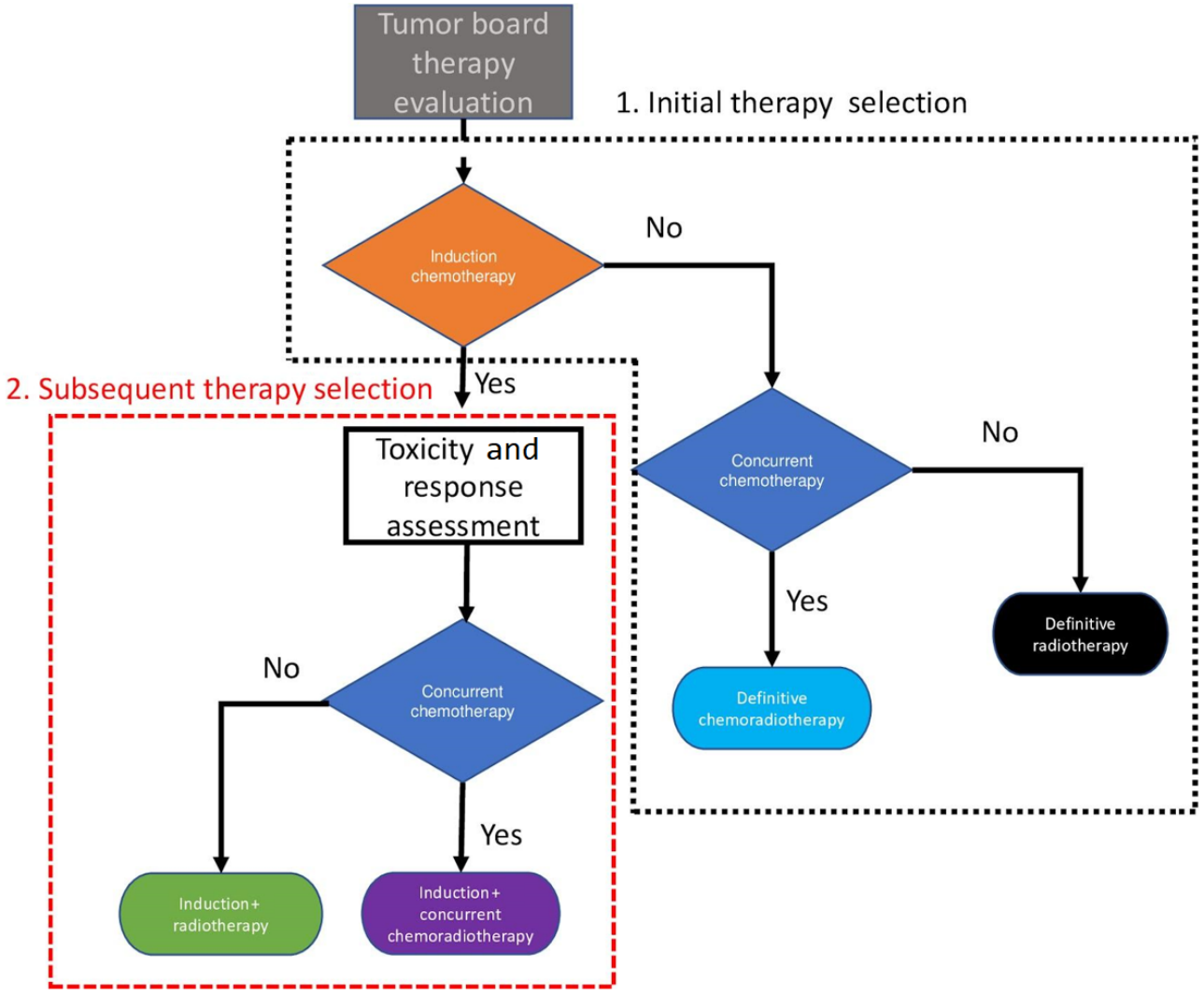
Background

Head and neck cancer, which includes cancers of the larynx, throat, lips, mouth, nose, and salivary glands, is now an epidemic, with 65,000 new cases in the United States annually [1], whose treatment is, as in many other types of cancers, a dynamic and complex process. This therapy process involves making multiple, patient-specific treatment decisions to maximize efficacy—for example, reduction in tumor size, time of local region control, and survival time—while minimizing side effects [2-4]. For example, a specific patient may undergo radiotherapy (RT) alone, RT with concurrent chemotherapy (CC), or induction chemotherapy (IC) [5]. After each round of IC, a decision must be made whether to continue IC or start either RT or CC. These decisions are currently taken by clinicians or multidisciplinary tumor boards based on pretherapy patient characteristics or crude heuristics. Notably, current risk-prediction models (eg, American Joint Committee on Cancer [AJCC] staging) incorporated in clinical decision support systems do not by themselves systematically direct clinicians to select an appropriate treatment that incorporates *both* oncologic and toxicity end points.

Furthermore, disposition to initial IC is then followed by a second responsive disposition to either RT or concurrent chemoradiotherapy. Inferring the optimal treatment policies for multistage decisions (eg, which treatment to administer initially and then after observing treatment response; Figure 1) post hoc is challenging because an optimal therapy sequence cannot be readily *pieced together* from several single-stage decisions.

For this reason, in the absence of rigorous clinical trials comparing adaptive IC permutations with concurrent RT, group comparison is exceedingly difficult because simple models that account for confounders at initial disposition (eg, propensity scores) are unequipped to incorporate sequential decision processes (eg, the choice of CC *after* IC).

Figure 1. Overview of the therapy selection process, which shows two distinct phases: initial therapeutic selection and subsequent therapeutic selection.



Digital Twinning

To address multistage models of therapy selection that incorporate both relevant cancer and side-effect considerations, we introduce an approach based on *digital twinning*, a new concept adapted to health research from the industrial world, where a digital replica (*digital twin*) of a physical entity or process is virtually recreated, with similar elements and dynamics, to perform real-time optimization and testing [6]. In health care, coupled digital twins, that is, *digital twin dyads*,

could be created for both patients and for the therapy process and used to inform in a quantitative manner adaptive therapy decision-making and allow personalization and optimization of health outcomes, prediction and prevention of adverse events, and planning interventions [7]. By leveraging a large number of head and neck cancer cases collected at a single institutional head and neck data tumor board at the MD Anderson Cancer Center (MDACC), we propose a methodology approach to leverage deep Q-learning (DQL) as a method to construct a *digital twin dyad* for simulation of therapy outcomes and potential implementation as a clinical decision aid. Q-learning is a recently developed machine learning method for supervised variable selection and weighting accounting for iterative processes [8].

In this paper, we apply for the first time Q-learning methodology to dynamically select treatment based on multiple clinically relevant outcomes from data specific to patients with head and neck cancer. We use these methods to construct and develop optimal dynamic treatment strategies, that is, digital twins of the therapy process. In conjunction with simulation models of patient data, the treatment prescription models form a patient-physician (prescriber) *digital twin dyad* in which the treatment prescription models act as a physician avatar by identifying the optimal treatment for the patient, whereas the simulation models represent the patient by predicting the outcome of the treatment sequence. We evaluate the results of this digital twin dyad approach on a curated data set of patients with head and neck cancer.

Methods

Overview

A state-of-the-art machine learning method applicable to the optimal therapy process problem is reinforcement learning (RL), in particular DQL [8-12]. DQL aims to solve problems in which a model has to choose among a series of options to maximize a certain goal in the given situation: the model observes a set of actions and the outcome these actions have, thus learning which choices are optimal and which are not. Q-learning is thus a type of machine learning that enables systems to automatically learn and improve from experience without being explicitly programmed. Q-learning has been shown to lead to valid results in a variety of medical problems, including the definition of a sequential multiple-assignment randomized trial [8,9], the optimal treatment of depression [8] and Attention-Deficit/Hyperactivity Disorder [9], and the breastfeeding habits that maximize child vocabulary development [10].

We used DQL to find a treatment policy that maximizes a linear combination of multiple patient outcomes; for example, toxicological and survival outcomes. We considered a 3-step Markov decision process (MDP), with 3 actions in each episode corresponding to the three treatment decision points for each patient:

1. Decision 1 (D1): *IC or not*
2. Decision 2 (D2): *CC or RT alone*
3. Decision 3 (D3): *neck dissection (ND) or not*

More details on the setup of the MDP are described in the following sections, including the reward functions and state variables.

Patient Data Set

We performed a retrospective review of 536 patients with oropharyngeal squamous cell carcinoma who were treated at the MDACC between 2005 and 2013 (Tables 1-3). Radiomics analysis was performed [13,14] for the segmented [15,16] primary tumor volumes. Only patients with a minimum follow-up of 4 years or who died within 4 years were included in the data set. The 536 examples were partitioned into 2 distinct sets for training and testing using a 75% (n=402)-25% (n=134) random split. To save space, in Table 1, the results of all binary features are shown only for 1 outcome; the others can be derived directly by subtracting from 100%. For example, the figures for sex being female are 65 (12.1%), 47 (11.7%), and 18 (13.4%), under the respective column headings.

Table 1. Demographics of pretreatment features (before decision 1 [D1]: induction chemotherapy or not; N=536).

Characteristics	All patients (N=536)	Training set (n=402)	Testing set (n=134)
Group 1: Pretreatment features (before D1)			
Age (years) at diagnosis, mean (SD)	58.9 (9.5)	58.5 (9.4)	60.2 (9.6)
Pathological grade, n (%)			
I	6 (1.1)	2 (0.5)	4 (3)
II	154 (28.7)	114 (28.4)	40 (29.9)
III	274 (51.1)	206 (51.2)	88 (65.7)
IV	3 (0.6)	1 (0.2)	2 (1.5)
Not available	99 (18.5)	79 (19.7)	20 (14.9)
Sex (male), n (%)	471 (87.9)	355 (88.3)	116 (86.6)
HPV[a] or P16[b] status, n (%)			
Negative	43 (8)	33 (8.2)	10 (7.5)
Positive	305 (56.9)	228 (56.7)	77 (57.5)
Unknown	188 (35.1)	141 (35.1)	47 (35.1)
T[c] category, n (%)			
T1	113 (21.1)	87 (21.6)	26 (19.4)
T2	219 (40.9)	156 (38.8)	63 (47)
T3	116 (21.6)	91 (22.6)	25 (18.7)
T4	86 (16)	67 (16.7)	19 (14.2)
Tx[d]	2 (0.4)	1 (0.2)	1 (0.7)
N[e] category (8th edition[f]), n (%)			
N0[g]	20 (3.7)	14 (3.5)	6 (4.5)
N1	249 (46.5)	181 (45)	68 (50.7)
N2	250 (46.6)	194 (48.3)	56 (41.8)
N3	17 (3.2)	13 (3.2)	4 (3)
AJCC[h] (8th edition), n (%)			
I	186 (34.7)	137 (34.1)	49 (36.6)
II	81 (15.1)	63 (15.7)	18 (13.4)
III	64 (11.9)	44 (10.9)	20 (14.9)
IV	203 (37.9)	157 (39.1)	46 (34.3)
Not available	2 (0.3)	1 (0.2)	1 (0.7)
Smoking status at diagnosis, n (%)			
Current	115 (21.5)	85 (21.1)	30 (22.4)
Former	203 (37.9)	151 (37.6)	52 (38.8)
Never	218 (40.7)	166 (41.3)	52 (38.8)
Smoking status			
Packs per year, mean (SD)	17.7 (23.7)	16.7 (22.9)	20.5 (26)
Not available, n (%)	28 (4.7)	21 (5.2)	7 (5.2)
Aspiration rate before therapy (no), n (%)	16 (3)	14 (3.5)	2 (1.5)
Number of affected lymph nodes, mean (SD)	2.0 (1.3)	2.1 (1.3)	1.8 (1)
Tumor laterality, n (%)			
Bilateral	21 (3.9)	16 (4)	5 (3.7)

Characteristics	All patients (N=536)	Training set (n=402)	Testing set (n=134)
Left	242 (45.1)	188 (46.8)	54 (40.3)
Right	273 (50.9)	198 (49.3)	75 (56)
Tumor subsite, n (%)			
Base of tongue	266 (49.6)	204 (50.7)	62 (46.3)
Tonsil	223 (41.6)	158 (39.3)	65 (48.5)
Other	47 (8.8)	40 (10)	7 (5.2)
Race, n (%)			
African American or Black	16 (3)	10 (2.5)	6 (4.5)
Asian	4 (0.7)	3 (0.7)	1 (0.7)
Hispanic or Latino	21 (3.9)	17 (4.2)	4 (3)
Native American	1 (0.2)	1 (0.2)	0 (0)
White or other	494 (92.2)	371 (92.3)	123 (91.8)

[a]HPV: human papillomavirus.

[b]P16: protein expression 16.

[c]T: primary tumor.

[d]Tx: no information about the primary tumor or it cannot be measured.

[e]N: lymph nodes.

[f]American Joint Committee on Cancer's Cancer Staging Manual, 8th edition.

[g]N0: nearby lymph nodes do not contain cancer.

[h]AJCC: American Joint Committee on Cancer.

Table 2. Feature demographics before and after decision junctions (N=536).

Characteristics	All patients (N=536), n (%)	Training set (n=402), n (%)	Testing set (n=134), n (%)
Group 2: Post– induction chemotherapy– decision features (after D1[a] and before D2[b])			
Prescribed chemotherapy			
None	342 (63.8)	250 (62.2)	92 (68.7)
Doublet	41 (7.6)	32 (8)	9 (6.7)
Triplet	143 (26.7)	111 (27.6)	32 (23.9)
Quadruplet	7 (1.3)	7 (1.7)	0 (0)
Not otherwise specified	3 (0.6)	2 (0.5)	1 (0.7)
Chemotherapy modification	85 (15.9)	65 (16.2)	20 (14.9)
Chemotherapy modification type			
No dose adjustment	451 (84.1)	336 (83.6)	115 (85.8)
Dose modified	21 (3.9)	16 (4)	5 (3.7)
Dose delayed	10 (1.9)	9 (2.2)	1 (0.7)
Dose cancelled	18 (3.4)	13 (3.2)	5 (3.7)
Dose delayed and modified	6 (1.1)	5 (1.2)	1 (0.7)
Regimen modification	29 (5.4)	22 (5.5)	7 (5.2)
Unknown	1 (0.2)	1 (0.2)	0 (0)
Dose-limiting toxicity	95 (17.7)	73 (18.2)	22 (16.4)
Dose-limiting toxicity g rade (also included for dermatological, neurological, gastrointestinal, hematological, nephrological, vascular, and infection [pneumonia])			
0	446 (83.2)	334 (83.1)	112 (83.6)
1	7 (1.3)	6 (1.5)	1 (0.7)
2	33 (6.2)	26 (6.5)	7 (5.2)
3	41 (7.6)	29 (7.2)	12 (9)
4	9 (1.7)	7 (1.7)	2 (1.5)
Imaging (yes)	194 (36.2)	152 (37.8)	42 (31.3)
Complete response, primary (1[c], as opposed to 0[d])	84 (15.7)	67 (16.7)	17 (12.7)
Complete response, nodal (1)	16 (3)	14 (3.5)	2 (1.5)
Parietal response, primary (1)	89 (16.6)	70 (17.4)	19 (14.2)
Parietal response, nodal (1)	156 (29.1)	125 (31.1)	31 (23.1)
Stable disease, primary (1)	11 (2.1)	8 (2)	3 (2.2)
Stable disease, nodal (1)	10 (1.9)	6 (1.5)	4 (3)
Group 3: Post–concurrent chemotherapy –decision features (after D2 and before D3[e])			
Concurrent chemotherapy regimen			
None	126 (23.5)	89 (22.1)	37 (27.6)
Platinum based	257 (47.9)	198 (49.3)	59 (44)
Cetuximab based	129 (24.1)	95 (23.6)	34 (25.4)
Other	24 (4.5)	20 (5)	4 (3)
Concurrent chemotherapy modification (1)	99 (18.5)	77 (19.2)	22 (16.4)
Complete response, primary 2 (1)	450 (84.1)	336 (83.8)	114 (85.1)
Complete response, nodal 2 (1)	247 (46.1)	186 (46.3)	61 (45.5)
Parietal response, primary 2 (1)	77 (14.4)	58 (14.4)	19 (14.2)
Parietal response, nodal 2 (1)	257 (47.9)	191 (47.5)	66 (49.3)

Characteristics	All patients (N=536), n (%)	Training set (n=402), n (%)	Testing set (n=134), n (%)
Stable disease, primary 2 (1)	2 (0.4)	2 (0.5)	0 (0)
Stable disease, nodal 2 (1)	10 (1.9)	6 (1.5)	4 (3)
Dose-limiting toxicity 2 (also included for dermatological, neurological, gastrointestinal, hematological, nephrological, vascular, and other)	102 (19)	80 (19.9)	22 (16.4)
Group 4: Primary outcomes after D3			
Four-year overall survival (alive)	457 (85.3)	344 (85.6)	113 (84.3)
Feeding tube 6 months (yes)	98 (18.3)	77 (19.2)	21 (15.7)
Aspiration rate after therapy (yes)	98 (18.3)	79 (19.7)	19 (14.2)
Dysphagia (yes)	154 (28.7)	122 (30.3)	32 (23.9)

[a]D1: decision 1 (induction chemotherapy or not).

[b]D2: decision 2 (concurrent chemotherapy or radiotherapy alone).

[c]The patient survived for at least four years after the treatment ended.

[d]All other events.

[e]D3: decision 3 (neck dissection or not).

Table 3. Demographics of physicians' decisions (N=536).

Characteristics	All patients (N=536), n (%)	Training set (n=402), n (%)	Testing set (n=134), n (%)
Treatment decisions (made by physicians)			
D1[a]: yes	194 (36.2)	152 (37.8)	42 (31.3)
D2[b]: yes	410 (76.5)	313 (77.9)	97 (72.4)
D3[c]: yes	111 (20.7)	84 (20.9)	27 (20.1)

[a]D1: decision 1 (induction chemotherapy or not).

[b]D2: decision 2 (concurrent chemotherapy).

[c]D3: decision 3 (neck dissection or not).

Ethics Approval

The data were collected after approval from the MDACC institutional review board (PA16-0303 and retrospective RCR03-0800).

Modeling

We focused on two outcome measures: (1) four-year *overall survival* (OS) as a single binary dichotomized outcome measure (ie, the patient survived for at least four years after the treatment ended, coded as 1, with all other events coded 0) and (2) the combination of *OS* and *dysphagia* (DP) as a multi-outcome measure. *DP* is defined as either *feeding-tube dependence* (FT) or *aspiration rate* (AR) 6 months after the end of treatment [17,18]. Note that although OS is encoded as a binary value, the outcome of treatment depends on the external situation and the treatment sequence applied; that is, both OS and DP are influenced by the treatment sequence applied. As a result, the whole problem is not a simple regression but bona fide RL with unknown transition probability. The combined outcome measure was computed only at the final step using the following formula:

$$\text{OS}-(\text{FT}+\text{AR After Therapy}-\text{AR Before Therapy}) \quad \textbf{(1)}$$

Equation (1) was used as the total reward in training the DQL models. For each of these scenarios, the models were trained with and without the inclusion of radiomics features [19,20]. The reward is 0 for D1 and D2 and equal to equation (1) for D3. As a result, there is no need for a discount factor (or, equivalently, set it to 1), and the total reward is exactly equation (1).

The state variables are illustrated in Tables 1 and 2, where all the features are divided into four groups separated by the state in which those features were used:

1. Group 1: pretreatment features (before D1)
2. Group 2: post–IC-decision features (after D1 and before D2)
3. Group 3: post–CC-decision features (after D2 and before D3)
4. Group 4: primary outcomes after ND decision (after D3)

The features in Table 1 were used for the initial state s0 of the MDP and are denoted as group 1. In Table 2, the features in group 2 combined with group 1 and a 1-hot vector of D1 were used for state s1; the features in group 3 combined with groups 1-2 along with a 1-hot vector of D1 and D2 were used for state s2. As a result, the features included at each decision point represent the *complete* history of the patient up to the current treatment decision. The features in group 4 were *only* used to evaluate the reward as formulated in equation (1) and *not* used as learning features. A detailed description of all variables is given in Table S1 in Multimedia Appendix 1, and we have

summarized the demographics of physicians' decisions in Table 3.

Preprocessing

The data set was randomly split into training (402/536, 75%) and testing (134/536, 25%) sets. To reduce the radiomic feature dimensionality (approximately 1000) [21], we applied principal component analysis and kept the 6 top components, which explain 90% of the overall feature variance. No blind assessment of the decisions or outcomes was made. Unknown human papillomavirus status was handled using a distinguished value (0). Missing values for all other covariates were handled using single imputation: median for numerical variables and mode for categorical variables. The ordinal covariates pathological grade, T (primary tumor) category, N (lymph nodes) category, AJCC staging, and prescribed chemotherapy (none, single, doublet, triplet, or quadruplet) were coded as numerical features. After these preprocessing steps, all features were rescaled to the (–1 to +1) range, as is standard in neural network (NN) training.

DQL Neural Modeling

Figure 2 shows an overview of the training process. A separate NN model was trained for each of the decision points D1-D3. Each model was constructed recursively based on the previous model results at the subsequent decision point or the outcome (single or combined) in the case of D3. The models were trained to optimize the total rewards *without any discounting factor*, that is, the combined outcome of equation (1).

The first model to be trained was Q3, which represents ND (D3), based on the final outcomes, the treatment decisions made in D3, and the patient's history before D3. We tuned the learning rate so that the mean reward converged smoothly instead of fluctuating drastically. The training for D3 was terminated when the NN weights had converged. Next, the model for D2 was trained based on the result of Q3 instead of the final outcomes, and D1 was trained based on the result of Q2. The models were constructed and trained using the PyTorch framework with graphics processing unit acceleration. Once the models had been trained, they were used in a forward order, as opposed to the training order, to prescribe the optimal treatment at each decision step. This is illustrated in Figure 3.

Figure 2. Overview of deep Q-learning model training. RL: reinforcement learning.

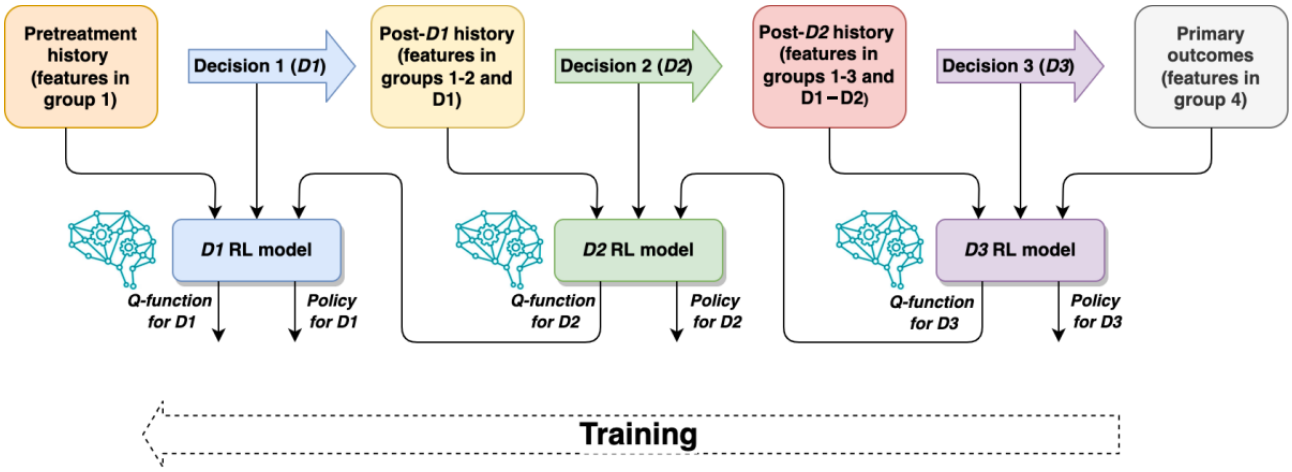
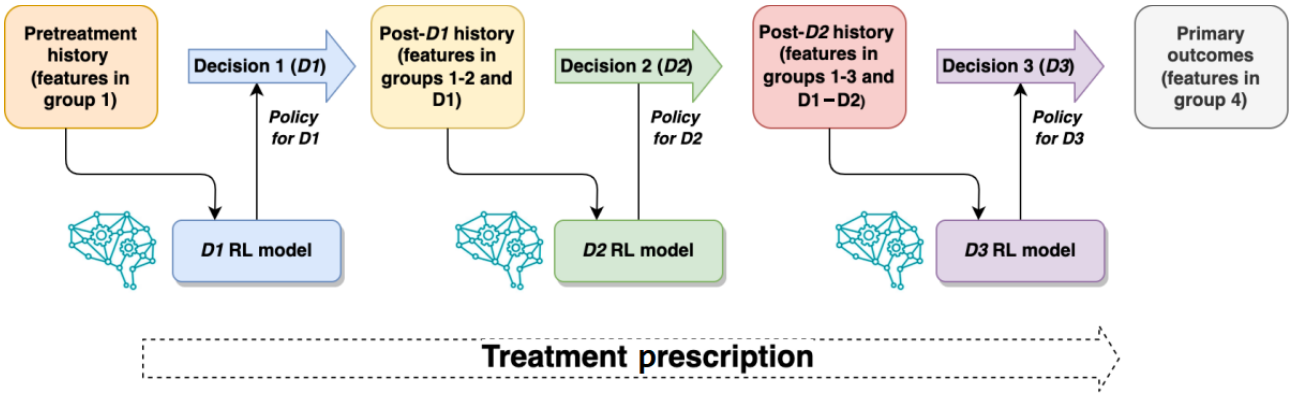


Figure 3. Overview of applying deep Q-learning model to make treatment prescriptions. RL: reinforcement learning.



We constructed multiple shallow-to-deep NNs with an increasing number of layers until the deepest model showed poor performance because of overfitting. We sampled 1000 separate training sets from the initial training data and trained a separate model on each of these sets, thus obtaining bootstrapped models with 95% CIs. Because of the high computational cost of bootstrapping, we will report in the *Results* section the performance of survival and toxicity under all possible numbers of hidden layers from 1 to 8, instead of performing the 5-fold cross-validation on all hyperparameters such as the number of nodes in each layer.

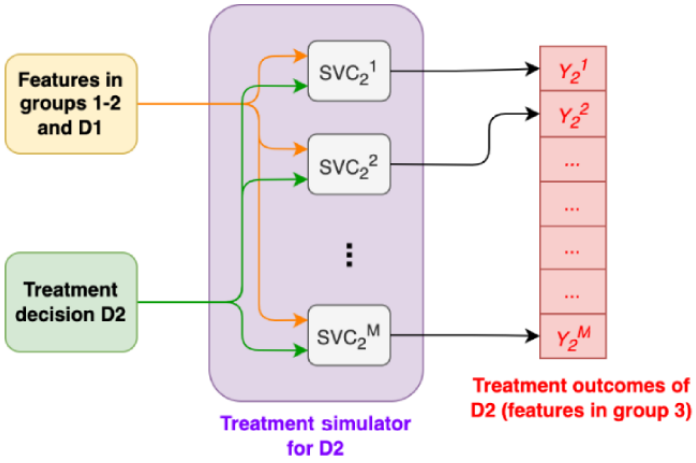
By prescribing an optimal treatment at each treatment junction, the DQL models constructed a *digital twin* of the decision process, with the goal of finding an optimal treatment plan that may differ from the physician's original decisions. The trained models and the code to compute the treatment decisions have been made available on GitHub [22].

Treatment Simulator for Evaluating DQL

As the DQL goal is not to replicate clinicians' decisions but to find an optimal, potentially different treatment, our evaluation includes building a treatment simulator (TS) model that, given a patient's history and the prescribed treatment, predicts the outcome of that treatment. The TS consists of a transition model for each intermediate and final outcome measure, built using a support vector classifier (SVC). For example, in the case of D1, an SVC was trained for each group 2 feature in Table 2 to predict, as output, that feature's value resulting from D1, and the input of these SVCs is all the Table 1 features, that is, group 1. For D2, SVC will predict for group 3 features in Table 2 using, as input, features from groups 1-2, along with a 1-hot vector of D1. The architecture is demonstrated in Figure 4. D3 was treated similarly with input features from groups 1-3, along with a 1-hot vector of D1 and D2. The full details of the SVC are provided in Table S2 in Multimedia Appendix 1. The C, γ, and class weights of each SVC were tuned through 5-fold cross-validation over the *F1* score on the training data because different values are needed for optimally predicting different features. Some features such as FT have quite imbalanced values, and to address this, we set the weights of each training example to be inversely proportional to the frequency of its class, hence placing more emphasis on less-common classes. The accuracy of the TS in predicting the next-stage feature value (instead of the treatment decision) was assessed with 95% CIs by using out-of-bag evaluation of 1000 models trained on stratified bootstrapped samples.

The TS serves as an in silico *digitaltwin* of the patient treatment because we can use it to dynamically simulate the patient's in vivo course as a function-given treatment policy, without having to physically treat the patient.

Figure 4. Illustration of the treatment simulator for D2. Those for D1 and D3 are similar, and their input features are from group 1 and groups 1-3, respectively. SVC: support vector classifier. D1: decision 1. D2: decision 2. D3: decision 3.



Protocol of Evaluating DQL

The DQL models were evaluated against the TS because our goal is not to replicate physicians' decisions but to learn from the final reward and then quantitatively evaluate the treatment decisions learned by the DQL model. Such *what-if* questions are standard in off-policy evaluation in RL (in *off-policy* evaluation, one evaluates a policy without being able to implement it in the real environment). The state-of-the-art approaches fall into three categories [23]: (1) direct approach where a model of the environment is fit (same as what we do), (2) importance sampling, and (3) a combination of the 2. Importance sampling is known to suffer from high variance and thus would require a large number of samples, whereas our test set consisted of 134 patients. Similarly, Gottesman et al [24] detailed this difficulty along with several possible scenarios, but no conclusion was drawn regarding which metric to use. Yauney and Shah [25] evaluated the learned policy through simulated clinical trials, an approach identical to ours.

At the same time, to the best of our knowledge, there is no existing rule-based approach (eg, decision trees) that is suitable for this task. We note that although very generic methods such as decision trees could be customized for a single-step prediction, they do not account for the sequential nature of this decision-making process. Furthermore, ultimately, evaluating such rule-based approaches would encounter the same *what-if* questions, that is, off-policy evaluation. Indeed, this has been a long-standing open problem in RL, and we hope that our digital twin approach may provide a new *partial* solution.

Although we tested the DQL models against the TS that allows on-policy evaluation, we emphasize two important considerations in the evaluation protocol:

1. TS was not used for training. Instead, we intentionally trained DQL on a tabular observation data set of 402 patients. This is because if we did train on the TS, the learned model would overfit the simulated environment, thereby overestimating the test performance (which is also measured from the TS). This deliberate decoupling of training and testing strategy, which is also adopted by Yauney and Shah [25], is aimed specifically to ensure the fairness of the comparison. Again, note that although the TS can be directly used to optimize the policy through any model-based RL, we intentionally refrained from doing so

and followed the model-free DQL. This ensures a fair testing, noting that the TS is also used in generating the testing trajectory.

2. The learned agent did not have access to the TS at test time, and the decisions were based solely on the current state. The TS was invoked only to simulate the environment, that is, generating the consequent state arising from the proposed decision and treatment, allowing the performance to be evaluated. This was consistent with the model-free nature of the DQL and ensured a fair evaluation by avoiding peeking into the real dynamics under which the test was conducted.

Incidentally, even if the TS *were* available for decision-making (that is, the planning setting), there would still be significant obstacles. In open-loop planning, a sequence of actions (3 treatment decisions in our application) is chosen *before* actually performing any of them, that is, later actions neither await nor respond to the outcome of the preceding actions. In this case, one only needs to compute 8 scores and select the optimal one. However, even in such an overly simplistic solution, one still needs to compute the expected reward, which relies on integration over the stochastic outcome of the actions, that is, states s1-s3. Mathematically, it solves

Although sampling is a natural approach to it, the high dimensionality of the state space demands a large amount of samples from the TS to accurately compute the expected reward.

In practice, closed-loop planning is clearly preferred, where later actions are chosen to best respond to the outcome of preceding decisions and treatments, leading to the mathematical optimization formulation as

As a result, we must compute the state value (V[s]-functions) or the state-action value (Q[s,a]-functions). Because of the complexity of state space, both of them are nontrivial, even given the TS. Compared with open-loop planning, an additional layer of difficulty is incurred here because one needs to estimate 8 *functions* instead of 8 real numbers.

To summarize, this *patient treatment digital twin* approach enables us to simulate the results of applying the Q-learning models to patients and to compare the outcomes with those resulting from the clinicians' decisions. Fairness was also upheld by not using the TS in either training or decision-making at testing time.

Evaluation Metrics

The TS performance was evaluated by 2 accuracies without running the DQL. The *1-step* accuracy follows the trajectory from the data set and, at each of the 3 decision junctions, predicts the resulting feature value after practicing the physician's treatment and then compares it with the ground truth outcome in the data set. In contrast, the *start-to-finish* accuracy is only concerned with the final outcome features in group 4 of Table 2. It uses the TS to generate a simulated *3-step* trajectory for each patient by following the 3 treatment decisions from the physician and compares the final outcome with the ground truth.

The DQL models were then evaluated by comparing the OS and DP rates (as computed by the TS) resulting from the DQL treatment decisions with the outcomes observed under physician treatment on a separate test set. To facilitate interpretation, we computed the similarity between each of the DQL model's decisions and the physicians' decisions, considering each decision point independently. This evaluation does not need the TS. To further support interpretation, the policy followed by each model was analyzed by computing the increase (or decrease) in prescription rate for each treatment decision compared with the physicians' ad hoc prescriptions to express whether the model was more (or less) likely to prescribe a certain treatment when compared with actual physicians.

We also evaluated the DQL treatment decisions by examining compliance with the National Comprehensive Cancer Network guidelines of acceptable care [26], which state that eligible patients with advanced-stage cancer (T3-4 or N1-3) must be prescribed chemotherapy, either IC (D1) or CC (D2). These guidelines or restrictions were not explicitly imposed during model training.

We first report the performance of the TS and the simulation performance of the DQL models and compare the DQL recommendations with the physician decision process, both in terms of per-decision similarity and overall similarity, that is, averaging the similarity for each decision point for each model. To report compliance, and to ensure quality and facilitate reproducibility, we provide a formal presentation of the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis checklist, formalized in Table S3 in Multimedia Appendix 1.

Results

Accuracy of TS Models

The complete *1-step* prediction accuracy of each TS model (with 95% CIs) is presented in Table 4. The average bootstrapped prediction accuracy of the individual TS models was 87.35% (SD 11.15%), and the median accuracy was 92.07% (IQR 13.56%). At the whole trajectory level, the average *start-to-finish* prediction accuracy on the test set outcomes was 83.21% (SD 1.54%), with 83.96% (SD 0.37%) accuracy for OS, 82.46% (SD 1.87%) for DP, 88.43% (SD 1.87%) for FT and 83.58% (SD 0.75%) for AR. Please note that these accuracy values are *neither* in terms of treatment prediction *nor* comparable with physician's treatment D1, D2, and D3. Instead, the TS predicts the patient's feature or state (eg, complete response or nodal) resulting from a treatment decision given in the data and compares it with the ground truth outcome in the data set. Therefore, this accuracy should not be compared with the frequency of matching physician decisions (which is 70.4%, as we will show in the *Similarity to Physicians* section).

Table 4. One-step prediction accuracy of treatment simulation (with 95% CIs) based on out-of-bag evaluation of 1000 stratified bootstrapped samples.

Predicted outcome	Accuracy without radiomics (%; 95% CI)	Accuracy with radiomics (%; 95% CI)
Overall survival (4 years)	78.23 (73.20-82.92)	*78.95 (74.29-83.09)* [a]
Feeding tube (6 months)	74.37 (68.81-79.40)	*74.74 (68.53-80)*
Aspiration rate after therapy	*75 (69.38-80)*	73.96 (68.04-78.76)
Prescribed chemotherapy (single, doublet, triplet, quadruplet, none, or not otherwise specified)	*83 (77.32-87.57)*	82.77 (78.06-87.13)
Chemotherapy modification (yes or no)	*82.09 (76.96-86.34)*	80.22 (75.98- 84.82)
Dose modified	92.39 (89.23-94.95)	*94.50 (92.31-96.39)*
Dose delayed	*92.39 (89.12-95.17)*	92.35 (88.56-95.52)
Dose cancelled	91.58 (87.68-94.77)	*93.37 (90.05-96.15)*
Regimen modification	*93.54 (84.36-95.88)*	91.79 (54.7-95.05)
DLT[b] (yes or no)	81.51 (77.25-85.42)	*81.77 (77.34-85.79)*
DLT: dermatological	*92.77 (23.95-95.29)*	90.58 (87.05-93.3)
DLT: neurological	92.17 (88.66-95.1)	*92.27 (88.83-95.26)*
DLT: gastrointestinal	89.60 (85.86-92.96)	*90.36 (86.8-93.36)*
DLT: hematological	90.10 (86.17-93.23)	*91.84 (88.02-94.47)*
DLT: nephrological	*99.03 (98-100)*	98.50 (96.55-99.52)
DLT: vascular	98.45 (96.45-100)	*98.50 (96.86-100)*
DLT: infection (pneumonia)	*98.98 (94.42-100)*	98.44 (96.37-99.50)
DLT: other	*95.08 (90.82-97.57)*	92.35 (83.17-96.98)
DLT: grade	73.85 (53.84-79.9)	*77.02 (72.55-81.48)*
No imaging (0=no and 1=yes)	100 (100-100)	100 (100-100)
Complete response, primary	83.51 (78.82-87.56)	*84.02 (79.58-88.05)*
Complete response, nodal	94.79 (90.5-97.03)	*94.82 (89.64-97.4)*
Parietal response, primary	*81.47 (76.84-86.27)*	80.32 (75.89-84.85)
Parietal response, nodal	92.93 (90-95.65)	92.93 (90.05-95.52)
Stable disease, primary	95.10 (91.96-97.84)	*96.35 (92.96-98.03)*
Stable disease, nodal	96.58 (94.47-98.05)	*97.50 (96.08-98.55)*
Concurrent chemotherapy regimen	*70 (64.68-75.27)*	65.99 (59.91-71.8)
Concurrent chemotherapy modification (yes or no)	70.53 (64.92-76.06)	*71.43 (65.68-76.68)*
Complete response, primary 2	*79.22 (23.03-85.22)*	77.35 (29.95-84.57)
Complete response, nodal 2	55.50 (49.01-61.54)	*56.25 (50-61.94)*
Parietal response, primary 2	78.92 (74.26-83.25)	*83.66 (79.90-86.60)*
Parietal response, nodal 2	52.50 (46.19-58.03)	*52.85 (46.46-58.62)*
Stable disease, primary 2	99.48 (98.46-100)	99.48 (98.41-100)
Stable disease, nodal 2	96.50 (94.12-98.04)	*96.92 (94.36-98.45)*
DLT: dermatological 2	91.99 (87.63-95.17)	*94.95 (91.53-97.07)*
DLT: neurological 2	*95.79 (5.96-97.46)*	91.97 (88.29-94.69)
DLT: gastrointestinal 2	89.74 (85.22-93.65)	*91.13 (87.50-94.06)*
DLT: hematological 2	92.71 (89.42-95.16)	*93.23 (90.10-95.57)*
DLT: nephrological 2	92.25 (88.17-97.94)	*96.53 (93.62-98.48)*
DLT: vascular 2	100 (99.45-100)	100 (99.02-100)
DLT: other 2	*93.97 (89.73-96.86)*	93.24 (89.23-96.14)

[a]Values in italics indicate whether higher accuracy is achieved by including or excluding radiomics.

[b]DLT: dose-limiting toxicity.

Performance of DQL in OS and DP

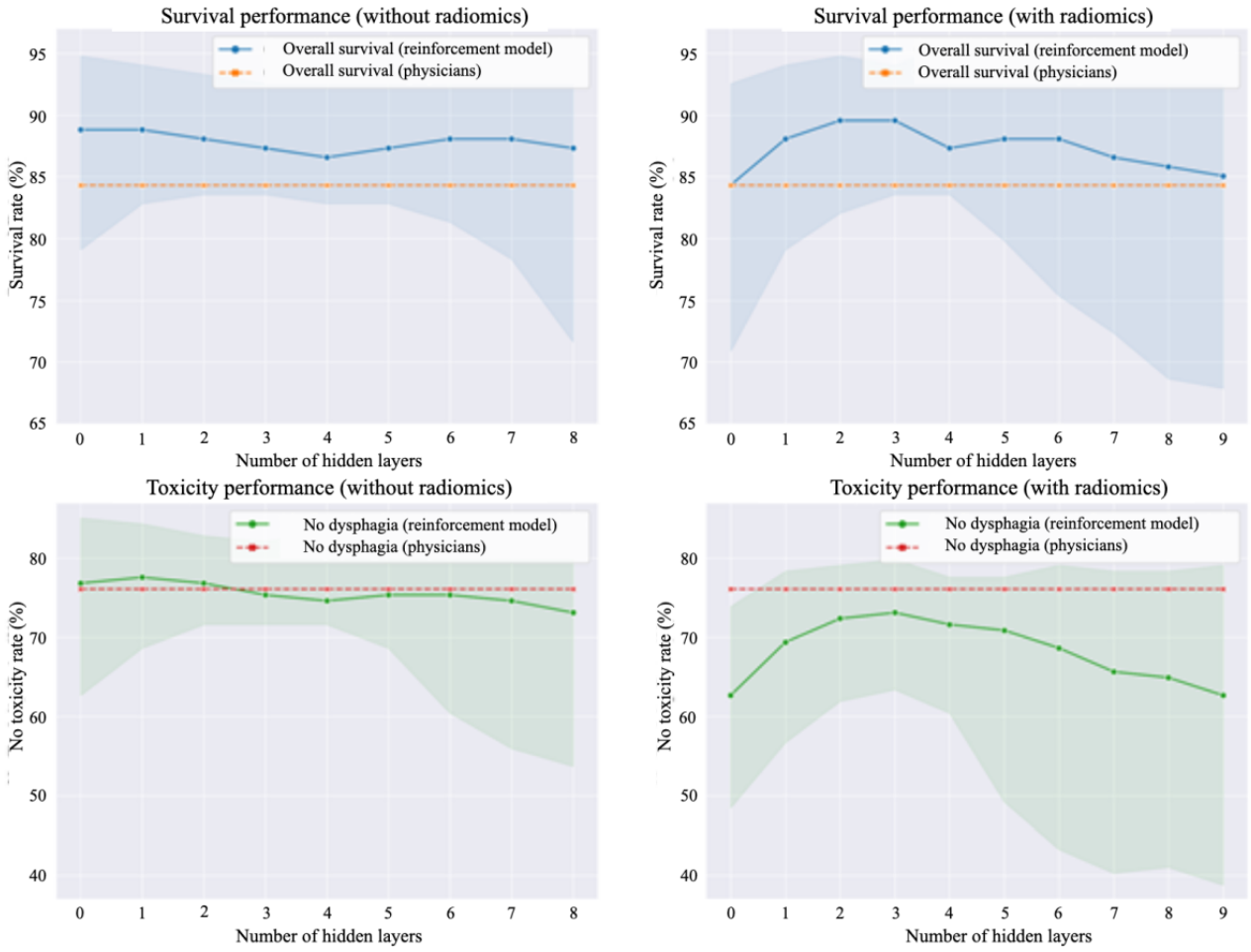
Recall from group 4 in Table 2 that the baseline outcomes observed under physician care are 85.57% (training set) and 84.33% (test set) OS rate of staying alive and 69.65% (training set) and 76.12% (test set; absence of) DP rate.

The complete performance of all DQL models on simulated patient outcomes is presented in Table S4 in Multimedia Appendix 1. For models trained to predict both OS and DP, the selected models were the ones with 2-3 hidden layers, which outperformed physician outcomes for OS. For predicting OS on the test data, the best model with radiomics had the highest average predicted OS rate but had higher variance and a worse lower bound of 95% CI (+5.22%, 95% CI –2.26% to 10.45%) compared with the best model without radiomics (+4.48%, 95% CI –1.49% to 9.7%). For DP, models without radiomics outperformed models with radiomics in terms of both average and lower bounds in terms of simulated patient outcomes.

For the purposes of this paper, we consider the *best* model to be the 2-layer NNs without radiomic features because it had the best lower bounds on predicted OS and DP for all models, while still affording a good average performance. This model yielded a median OS improvement, compared with physicians' results, of +2.74% (95% CI –0.25% to 6.47%; training set) and +3.73% (95% CI –0.75% to 8.96%; test set), with an absolute highest OS rate of 88.31% (95% CI 85.32%-92.04%; training set) and 88.06% (95% CI 83.58%-93.53%; test set). With respect to DP, the same 2-layer model showed a +3.98% (95% CI –1.24% to 9.2%) improvement on training data, with 73.63% (95% CI 68.41%-78.86%) of simulated patients not exhibiting DP under the model's treatment decisions. This 2-layer model yielded a +0.75% (95% CI –4.48% to 6.72%) improvement on the test set, from the baseline 76.12% to 76.87% (95% CI 71.64%-82.84%).

To assess model parsimony (ie, the minimum number of layers for maintaining equivalent predictive performance), Figure 5 shows a comparison of DQL models with different numbers of layers on simulated test data for the combined outcome models (OS and DP) and with and without radiomics features. Broadly, neither simpler models with <2 layers nor models with >4 layers performed as effectively as the 2-layer model. In Figure 5, continuous lines show the average performance of the bootstrapped models, whereas highlighted areas represent the 95% CIs. Dashed lines show the empirical patient outcomes observed under the physicians' decisions. Models with 1-4 layers had the highest performance and lower variance, whereas models with >4 layers overfit the data. Furthermore, models without radiomics had better overall performance for toxicity outcomes. Models with radiomics had slightly higher performance for OS outcomes but had higher variance and worse lower bounds than models trained without radiomics.

Figure 5. Model performance for the combined outcome (overall survival+dysphagia) models without (left) and with radiomics (right). The figure shows the performance for overall survival (top) and toxicity (dysphagia; bottom), with varying numbers of layers showing treatment simulation results on the test data.



Similarity to Physicians

The similar rates (with 95% CIs) with respect to physicians' treatment *twinning*, both per-decision and overall similarity, across all models are presented in Table S5 in Multimedia Appendix 1. We reiterate that the goal of DQL is not to replicate clinicians' decisions but to find an optimal, albeit potentially different, treatment. However, it is clearly of interest to measure the similarity as a reference. In terms of overall similarity of the in silico Q-learning treatment policies compared with those actually delivered ad hoc in vivo by physicians, our best model (OS+DP, 2 layers, and no radiomics) showed an overall 70.4% (95% CI 65.34%-73.63%) similarity to the physician decisions on the training set (ie, the model chooses the same treatment as the physicians for 70.4% of the considered treatment decisions) and 69.65% (95% CI 63.43%-73.38%) on the test set, although another model (the 3-layer OS+DP model without radiomics, which performed consistently worse in simulation for all outcomes) did show higher similarity rates.

Compliance With National Comprehensive Cancer Network Guidelines

The distributions (with 95% CIs) of the T and N stages of patients in the test set, separated by chemotherapeutic treatment prescribed by the best-performing model, are presented in Table 5.

The rates at which models choose a certain policy compared with the physicians' treatment rate at each decision point are shown in Figure 6. Gray lines represent the 95% CIs. The numbers on top of the bars show for how many patients (out of 134) the Q-learning model recommended that treatment. The IC prescription rate varies in a similar way between OS and OS+DP, CC is significantly more frequent in OS+DP models, whereas ND is consistently less frequent in OS+DP. The best-performing model in terms of simulated outcomes (OS+DP, 2 layers, and no radiomics) had a higher IC (D1) rate (2.99% increase in prescription rate compared with physicians' prescriptions for 46 patients, 95% CI –14.93% to 26.88%) and one of the highest CC (D2) rates (21.64% increase, 126 patients, 95% CI –2.99% to 27.61%), as well as the lowest ND (D3) rate (20.15% decrease, 0 patients, 95% CI –20.15% to –11.19%).

Table 5. Tumor stage demographics of patients based on the chemotherapeutic treatment decisions of the best-performing model (n=134).

Demographics	Chemotherapy			No chemotherapy, no induction chemotherapy, radiotherapy alone (%; 95% CI)
	Induction chemotherapy		No induction chemotherapy, concurrent chemotherapy (%; 95% CI)	
	Concurrent chemotherapy (%; 95% CI)	Radiotherapy alone (%; 95% CI)		
T[a] category				
T1	23.08 (0-65.38)	3.85 (0-26.92)	69.23 (26.92- 96.15)	0 (0-23.08)
T2	25.40 (6.35-55.56)	3.17 (0-22.22)	66.67 (38.06-88.89)	0 (0-20.63)
T3	32 (8-64.1)	4 (0-28)	60 (28-88)	0 (0-16)
T4	36.84 (5.26-84.21)	5.26 (0-31.58)	52.63 (10.53-94.74)	0 (0-15.79)
Tx[b]	0 (0-100)	0 (0-100)	100 (0-100)	0 (0-100)
N[c] category				
N0[d]	20 (0-100)	0 (0-40)	60 (0-100)	0 (0-40)
N1	17.39 (0-73.91)	0 (0-30.43)	73.91 (17.39-95.65)	0 (0-21.74)
N2	29.41 (9.80-56.89)	3.92 (0-22.55)	62.75 (36.27-81.37)	0 (0-17.65)
N3	25 (0-100)	0 (0-50)	50 (0-100)	0 (0-25)
N category (8th edition[e])				
N0	16.67 (0-100)	0 (0-33.33)	66.67 (0-100)	0 (0-33.33)
N1	22.06 (2.94-60.29)	2.94 (0-26.47)	70.59 (30.88-92.65)	0 (0-22.06)
N2	33.93 (8.93-69.64)	3.57 (0-26.79)	58.93 (25-83.93)	0 (0-14.29)
N3	25 (0-100)	0 (0-50)	50 (0-100)	0 (0-25)

[a]T: primary tumor.

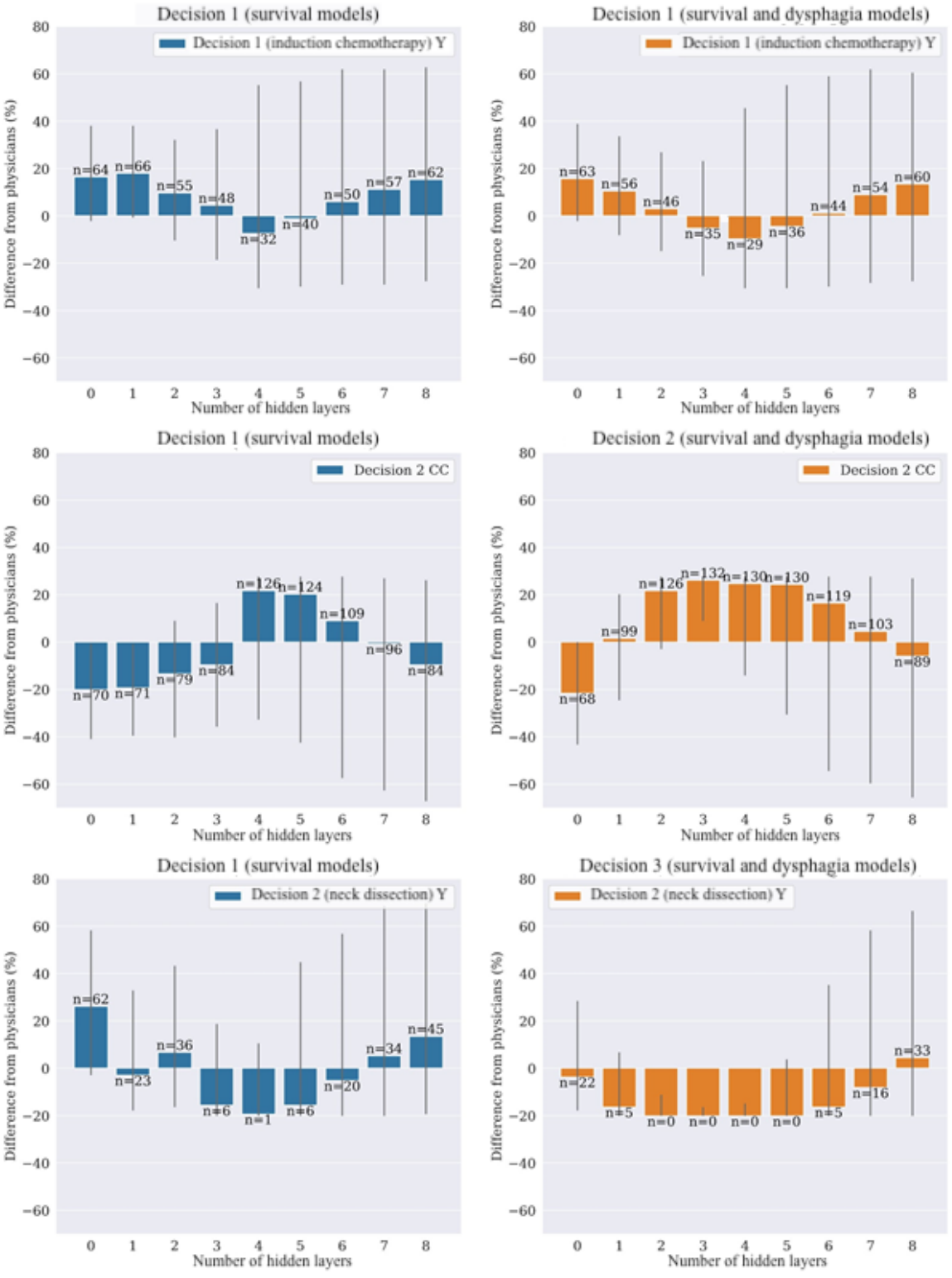
[b]Tx: no information about the primary tumor or it cannot be measured.

[c]N: lymph nodes.

[d]N0: nearby lymph nodes do not contain cancer.

[e]American Joint Committee on Cancer's Cancer Staging Manual, 8th edition.

Figure 6. Absolute increase (or decrease) of treatment decision rate compared with physicians' decisions. The plots refer to decisions 1 (top), 2 (middle), and 3 (bottom) on the test set and for models considering only overall survival as an outcome measure (left) or overall survival+dysphagia (right) without radiomics. Y: yes.



Computational Cost

The training time for a single DQL model did not significantly vary between shallower and deeper NNs and was just a few minutes on average for a complete model. With 1000-sample bootstrapping, the training time was accordingly longer, costing >24 hours to generate the results shown in Figure 5. However, these models are computationally inexpensive and can be deployed virtually in real time. Because of the computational cost of bootstrapping, we only used 5-fold cross-validation to tune the TS hyperparameters using SVCs, whereas for the other backbone hyperparameters such as the number of layers, we opted to report the performance of OS and DP under all possible numbers of hidden layers from 1 to 8 instead of performing cross-validation.

Clinical Case Review

As there is no practical way of verifying counterfactual *what-if* scenarios in actual patient care, we performed a post hoc case review of selected divergence of the policy prediction from delivered care. We herein report 2 case representative studies of patient-specific DQL treatment decisions—one differed from, whereas the other mostly concurred with, the original decisions of the treatment team of physicians—with discussion and input from oncologists at the MDACC. These 2 case studies were selected by performing 1000 bootstrapped trials on the entire data set (randomly partitioning it into training and testing), then recording the frequency with which the DQL prediction agreed with the physician's (among the 1000 bootstrapped trials). We used the smallest geometric mean value to select the first case study. The second case study was selected for high values of agreement rates in D2 and D3 but a low value of agreement rates in D1.

The patient in the first case study differed in every decision: the treatment sequence prescribed by their clinician team was D1: IC, D2: RT, and D3: ND, whereas the DQL sequence was D1: not IC, D2: CC, and D3: not ND. During our discussion, upon retrieving and examining the medical records, the oncologists described this case as having "a very unique and strange presentation" with bilateral disease involving the retropharyngeal lymph node (RPN). As the MDACC has historically associated RPN involvement with increased metastatic risk in published series [27,28], the patient was prescribed IC in D1 as part of an informal local policy for cases with perceived high risk of distant metastases or unsalvageable nodal failure (eg, retropharyngeal recurrence). The patient exhibited a substantive response with respect to the response evaluation criteria in solid tumors, wherein the primary tumor volume and index RPN had clinical complete response; therefore, in D2, the physicians prescribed RT alone. The patient exhibited a sizable response again, this time to the primary tumor. After RT, there was a notation by the radiologist of a negative positron emission tomography scan (eg, no evidence of metabolic uptake); however, the nodal remnant was evident and "malignancy could not be excluded." Consequently, in D3, the physicians prescribed, as a precaution, a completion ND on the lymph node but found no cancer, only necrotic tissue. In the oncologists' assessment, the DQL sequence would approximate typical standard of care more than the delivered treatment in terms of general community practice. In this case, the MDACC team altered the treatment based on additional *local* information related to the RPN. However, in their assessment, most other centers would not alter treatment based on this *typically not collected* information because RPN status is not a formal component of staging materials or risk categories for oropharyngeal cancer [29,30] nor of AJCC staging systems [31], and many, if not most, practices would treat this case as concurrent chemoradiotherapy. In summary, this was an extremely unusual, unique case where additional unannotated *local* information made the difference between the DQL and the prescriber's sequence. Future work that includes nodal involvement methodology [3,4] not currently reflected in the AJCC 8th Edition could address this type of borderline case.

The patient we considered in the second case study featured disagreement only in the first decision. The treatment sequence prescribed by their clinician team was D1: not IC, D2: CC, and D3: not ND, whereas the DQL sequence was D1: IC, D2: CC, and D3: not ND. Upon examining the medical records, the oncologists noted that the patient had only 1 functioning kidney; therefore, in the first stage, the team decided to prescribe a low-dose chemotherapy regimen treatment as a precaution to prevent renal injury [32]. In their assessment, our dyad system performed well, given the input specifications of this case, and the difference in therapy selection was attributed to occult (but clinically meaningful) comorbid disease variables not included in the decision platform that influenced the physicians' process.

Overall, the physician review in both these instances that we investigated in detail suggests that, in the absence of specific *local* practices or occult clinical features not included in this decision platform, the DQL recommendation would have been a good strategy and that the dyad provided "clinically acceptable recommendations."

Discussion

Principal Findings

The high average, median, and overall accuracies provided by the TS in predicting the outcomes of treatments indicate that the TS is a valid digital twin for the treatment process when predicting the outcome of a treatment sequence. Our results also indicate that the Q-learning models indeed capture the nature of the dynamic treatment problem and provide a valid solution. Our models showed consistent improvements for all the outcome features taken into account, as well as moderate similarity to physicians' decisions. Overall, these results indicate that DQL modeling can serve as a digital twin of the treatment decision process and TS modeling can serve as a digital twin of the patient treatment. When combined, DQL and TS constitute a valid patient-physician *digital twin dyad* for optimal policy determination in sequential systemic and locoregional therapy of oropharyngeal squamous carcinomas.

Furthermore, our results show that the DQL models that consider OS+DP outperform models considering only OS in terms of simulated survival rate. As the absence of DP (FT or AR) symptoms is positively correlated with OS, maximizing these indirectly helps maximize OS-model performance as well.

Moreover, OS+DP models show higher similarity to actual physician decisions because they represent a finer-grained approximation of the decision process than models that include only OS as an outcome, including more of the features considered by the physician when choosing an optimal treatment.

Surprisingly, given the abundance of data on radiomics models for head and neck cancer [13,33-40], Q-learning models *without* radiomics yielded a better performance than models that included radiomics, in terms of simulated outcomes and variance, for both training and testing data. The most evident example is given by the simulated DP: none of the models trained with radiomics features managed to improve the outcome observed after physicians' treatment in the test set (Figure 5,

bottom right). From these results, the addition of textural features *failed* to improve model performance and instead significantly increased the performance variance of the bootstrapped models.

Our findings also justify the choice of a deep NN model instead of a regular linear model: whereas by using DQL we reduce model parsimony, we can see that the results of the linear models (ie, the 0–hidden-layers NNs) are comparatively suboptimal to deeper models in terms of simulated performance, CI variance, and similarity.

Furthermore, per Table 5, the best-performing model did not violate the standard of care regarding chemotherapeutic treatment of patients because all patients with stages T3-4 or N1-3 cancer were prescribed either IC (D1) or CC (D2), with most of them being prescribed at least CC, showing that clinical applicability was maintained.

When comparing OS-only models with OS+DP models, the prescription rates presented in Figure 5 showed 2 separate trends for the 2 categories: whereas the IC (D1) prescription rate varied similarly for the OS and OS+DP models, the rates of CC (D2) and ND (D3) were significantly different between the 2 categories. In general, OS+DP models tended to have a higher rate of CC (D2) and a lower rate of ND (D3). In other words, *models that also consider toxicity as an outcome measure balanced a more aggressive chemotherapeutic treatment with a lower rate of ND*, which is consistent with the known positive correlation between surgery and DP symptoms such as FT and AR.

Limitations

Although the proposed approach was shown to be effective in dynamically selecting optimal treatment strategy for patients with oropharyngeal squamous cell carcinoma, it is not without limitations. Because of the retrospective nature of the data set, our Q-learning models had to be evaluated through the TS, a supervised learning model, which might be seen as self-referential. However, the TS is a necessary approach before prospective application because evaluating the models based on physician similarity alone would not reflect the purpose of DQL. Intuitively, the goal of DQL is not to replicate the decisions taken by physicians in the data set but to learn from these decisions and their effect to discern between optimal and nonoptimal choices with respect to a given outcome measure.

Furthermore, because we train our *digital twin dyad* on a representative cohort from a single cancer center (MDACC), the physician decision or prescribing heuristics reflected may not be fully generalizable to other facilities with other practitioners. However, the physician prescriptions at the MDACC are aligned with the state of the art in the field. In particular, we note that whereas 2 studies [41,42] have questioned the relevance of IC to treatment, both studies have failed to accrue and thus are null. Our modeling approach could conceivably be implemented and extended to generate similar *digital twin dyad*s across other or multiple institutions.

Conclusions

In conclusion, we constructed a DQL modeling approach to make optimized sequential treatment decisions based on a set of desired outcomes in head and neck cancer therapy and paired it with a simulation of the treatment process for evaluation purposes. This modeling approach represents, to our knowledge, the first application of DQL with simulation as a *digital twin dyad* to simultaneously represent both s*tate-specific patient data* and *physician or prescriber policies* for head and neck squamous carcinoma. Furthermore, this work is the first reported implementation of DQL for DP and OS composite-outcome modeling. Our approach further demonstrates the technical feasibility of such a *digital twin dyad* and provides a benchmarking data set and relevant code for model dissemination, site-specific implementation, and iterative model improvement. Carrying out a prospective clinical study application could further confirm the validity of this approach as part of the standard of care.

Acknowledgments

The authors thank all members of the Electronic Visualization Laboratory, members of the MD Anderson Head and Neck Cancer Quantitative Imaging Collaborative Group, and the authors' collaborators at the University of Iowa and University of Minnesota. This work was directly supported by the National Institutes of Health (NCI-R01-CA214825 and NCI-R01CA225190) and the National Science Foundation (CNS-1625941 and CNS-1828265). Direct infrastructure support was provided for this project by the multidisciplinary Stiefel Oropharyngeal Research Fund of the University of Texas MD Anderson Cancer Center Charles and Daneen Stiefel Center for Head and Neck Cancer, the MD Anderson Cancer Center Support Grant (P30CA016672), and the MD Anderson Program in Image Guided Cancer Therapy. CDF received funding and salary support *unrelated* to this project during the period of study execution from the National Institutes of Health ([R25EB025787-01, 1R01DE025248/R56DE025248, and R01DE028290], 1R01CA218148, P30CA016672, and P50 CA097007), NSF (1933369), and the Sabin Family Foundation. CDF has received unrelated direct industry grant support as well as honoraria and travel funding from Elekta AB that are not relevant to this project.

Data Availability

The data set used in the data analysis is publicly available [43].

Authors' Contributions

ET, XZ, GC, CDF, AW, and GEM designed and developed the machine learning models and were responsible for data extraction and curation, statistical analysis, and interpretation. LVD, ASRM, and CDF were responsible for direct patient care provision, direct clinical data collection, interpretation, and analytic support. GC supervised statistical analysis, data extraction, and analytic support and is the guarantor of statistical quality. ET, XZ, AW, GC, LVD, ASRM, CDF, and GEM were responsible for manuscript writing and editing. XZ, GC, CDF, and GEM, as the primary investigators, conceived, coordinated, and directed all study activities and were responsible for data collection, project integrity, manuscript content, editorial oversight, and correspondence. All authors made substantial contributions to the conception or design of the work or the acquisition, analysis, or interpretation of data; drafted the manuscript or revised it critically for important intellectual content; gave final approval of the version to be published; and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary Tables 1-5.
[DOCX File , 52 KB - jmir_v24i4e29455_app1.docx]

References

1. Adelstein DJ, Ridge JA, Gillison ML, Chaturvedi AK, D'Souza G, Gravitt PE, et al. Head and neck squamous cell cancer and the human papillomavirus: summary of a National Cancer Institute State of the Science Meeting, November 9-10, 2008, Washington, D.C. Head Neck 2009 Nov;31(11):1393-1422. [doi: 10.1002/hed.21269] [Medline: 19787782]
2. Multidisciplinary Larynx Cancer Working Group. Conditional survival analysis of patients with locally advanced laryngeal cancer: construction of a dynamic risk model and clinical nomogram. Sci Rep 2017 Mar 09;7:43928 [FREE Full text] [doi: 10.1038/srep43928] [Medline: 28276466]
3. Luciani T, Wentzel A, Elgohari B, Elhalawani H, Mohamed A, Canahuate G, et al. A spatial neighborhood methodology for computing and analyzing lymph node carcinoma similarity in precision medicine. J Biomed Inform 2020;112S:100067 [FREE Full text] [doi: 10.1016/j.yjbinx.2020.100067] [Medline: 34417010]
4. Wentzel A, Luciani T, van Dijk LV, Taku N, Elgohari B, Mohamed AS, Spatial-Non-spatial Multi-Dimensional Analysis of Radiotherapy Treatment/Toxicity Team SMART3. Precision association of lymphatic disease spread with radiation-associated toxicity in oropharyngeal squamous carcinomas. Radiother Oncol 2021 Aug;161:152-158. [doi: 10.1016/j.radonc.2021.06.016] [Medline: 34126138]
5. Marai GE, Ma C, Burks AT, Pellolio F, Canahuate G, Vock DM, et al. Precision risk analysis of cancer therapy with interactive nomograms and survival plots. IEEE Trans Vis Comput Graph 2019 Apr;25(4):1732-1745 [FREE Full text] [doi: 10.1109/TVCG.2018.2817557] [Medline: 29994094]
6. Fagherazzi G. Deep digital phenotyping and digital twins for precision health: time to dig deeper. J Med Internet Res 2020 Mar 03;22(3):e16770 [FREE Full text] [doi: 10.2196/16770] [Medline: 32130138]
7. Palmer M. The potential of the digital twin as a disruptor of healthcare: perspective from medical devices. National Institute of Biomedical Imaging and Bioengineering Interagency Modeling and Analysis Group Digital Twin Meeting Keynote. 2019. URL: https://www.imagwiki.nibib.nih.gov/sites/default/files/2020-01/2019-10-24%20Palmer_IMAG_DigitaTwinKeynote-final_1.pdf [accessed 2022-04-06]
8. Schulte PJ, Tsiatis AA, Laber EB, Davidian M. Q- and A-learning methods for estimating optimal dynamic treatment regimes. Stat Sci 2014 Nov;29(4):640-661 [FREE Full text] [doi: 10.1214/13-STS450] [Medline: 25620840]
9. Nahum-Shani I, Qian M, Almirall D, Pelham WE, Gnagy B, Fabiano GA, et al. Q-learning: a data analysis method for constructing adaptive interventions. Psychol Methods 2012 Dec;17(4):478-494 [FREE Full text] [doi: 10.1037/a0029373] [Medline: 23025434]
10. Moodie EE, Chakraborty B, Kramer MS. Q-learning for estimating optimal dynamic treatment rules from observational data. Can J Stat 2012 Dec 01;40(4):629-645 [FREE Full text] [doi: 10.1002/cjs.11162] [Medline: 23355757]
11. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. Nature 2015 Feb 26;518(7540):529-533. [doi: 10.1038/nature14236] [Medline: 25719670]
12. Sutton RS, Barto AG. Reinforcement learning: an introduction. 2nd edition. Cambridge, MA, USA: MIT Press; 2018.
13. M. D. Anderson Cancer Center Head and Neck Quantitative Imaging Working Group. Investigation of radiomic signatures for local recurrence using primary tumor texture analysis in oropharyngeal head and neck cancer patients. Sci Rep 2018 Jan 24;8(1):1524 [FREE Full text] [doi: 10.1038/s41598-017-14687-0] [Medline: 29367653]
14. Elhalawani H, Lin TA, Volpe S, Mohamed AS, White AL, Zafereo J, et al. Machine learning applications in head and neck radiation oncology: lessons from open-source radiomics challenges. Front Oncol 2018 Aug 17;8:294 [FREE Full text] [doi: 10.3389/fonc.2018.00294] [Medline: 30175071]

15. -. 4. Definition of volumes. J ICRU 2010 Apr;10(1):41-53. [doi: 10.1093/jicru/ndq009] [Medline: 24173326]
16. MICCAI/M.D. Anderson Cancer Center Head and Neck Quantitative Imaging Working Group. Matched computed tomography segmentation and demographic data for oropharyngeal cancer radiomics challenges. Sci Data 2017 Jul 04;4:170077 [FREE Full text] [doi: 10.1038/sdata.2017.77] [Medline: 28675381]
17. Wentzel A, Hanula P, Luciani T, Elgohari B, Elhalawani H, Canahuate G, et al. Cohort-based T-SSIM visual computing for radiation therapy prediction and exploration. IEEE Trans Vis Comput Graph 2020 Jan;26(1):949-959 [FREE Full text] [doi: 10.1109/TVCG.2019.2934546] [Medline: 31442988]
18. Wentzel A, Hanula P, van Dijk LV, Elgohari B, Mohamed AS, Cardenas CE, et al. Precision toxicity correlates of tumor spatial proximity to organs at risk in cancer patients receiving intensity-modulated radiotherapy. Radiother Oncol 2020 Jul;148:245-251 [FREE Full text] [doi: 10.1016/j.radonc.2020.05.023] [Medline: 32422303]
19. Tosado J, Zdilar L, Elhalawani H, Elgohari B, Vock DM, Marai GE, et al. Clustering of largely right-censored oropharyngeal head and neck cancer patients for discriminative groupings to improve outcome prediction. Sci Rep 2020 Mar 02;10(1):3811 [FREE Full text] [doi: 10.1038/s41598-020-60140-0] [Medline: 32123193]
20. Patel H, Vock DM, Marai GE, Fuller CD, Mohamed AS, Canahuate G. Oropharyngeal cancer patient stratification using random forest based-learning over high-dimensional radiomic features. Sci Rep 2021 Jul 07;11(1):14057 [FREE Full text] [doi: 10.1038/s41598-021-92072-8] [Medline: 34234160]
21. Zdilar L, Vock DM, Marai GE, Fuller CD, Mohamed AS, Elhalawani H, et al. Evaluating the effect of right-censored end point transformation for radiomic feature selection of data from patients with oropharyngeal cancer. JCO Clin Cancer Inform 2018 Dec;2:1-19 [FREE Full text] [doi: 10.1200/CCI.18.00052] [Medline: 30652615]
22. HNC-Digital-Twin. GitHub. URL: https://github.com/ElisaTardini/HNC-Digital-Twin [accessed 2022-04-12]
23. Voloshin C, Le HM, Jiang N, Yue Y. Empirical study of off-policy policy evaluation for reinforcement learning. arXiv (forthcoming) 2019.
24. Gottesman O, Johansson F, Meier J, Dent J, Lee D, Srinivasan S, et al. Evaluating reinforcement learning algorithms in observational health settings. arXiv (forthcoming) 2018.
25. Yauney G, Shah P. Evaluating reinforcement learning algorithms in observational health settings. In: Proceedings of the 3rd Machine Learning for Healthcare Conference. 2018 Presented at: MLHC '18; August 16-18, 2018; Stanford, CA, USA p. 1-49.
26. Pfister DG, Spencer S, Adelstein D, Adkins D, Anzai Y, Brizel DM, et al. Head and neck cancers, version 2.2020, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw 2020 Jul;18(7):873-898. [doi: 10.6004/jnccn.2020.0031] [Medline: 32634781]
27. Lin TA, Garden AS, Elhalawani H, Elgohari B, Jethanandani A, Ng SP, et al. Radiographic retropharyngeal lymph node involvement in HPV-associated oropharyngeal carcinoma: patterns of involvement and impact on patient outcomes. Cancer 2019 May 01;125(9):1536-1546 [FREE Full text] [doi: 10.1002/cncr.31944] [Medline: 30620385]
28. Gunn GB, Debnam JM, Fuller CD, Morrison WH, Frank SJ, Beadle BM, et al. The impact of radiographic retropharyngeal adenopathy in oropharyngeal cancer. Cancer 2013 Sep 01;119(17):3162-3169 [FREE Full text] [doi: 10.1002/cncr.28195] [Medline: 23733178]
29. Fakhry C, Zhang Q, Gillison ML, Nguyen-Tân PF, Rosenthal DI, Weber RS, et al. Validation of NRG oncology/RTOG-0129 risk groups for HPV-positive and HPV-negative oropharyngeal squamous cell cancer: Implications for risk-based therapeutic intensity trials. Cancer 2019 Jun 15;125(12):2027-2038 [FREE Full text] [doi: 10.1002/cncr.32025] [Medline: 30913305]
30. Ang KK, Harris J, Wheeler R, Weber R, Rosenthal DI, Nguyen-Tân PF, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. N Engl J Med 2010 Jul 01;363(1):24-35 [FREE Full text] [doi: 10.1056/NEJMoa0912217] [Medline: 20530316]
31. Machczyński P, Majchrzak E, Niewinski P, Marchlewska J, Golusiński W. A review of the 8th edition of the AJCC staging system for oropharyngeal cancer according to HPV status. Eur Arch Otorhinolaryngol 2020 Sep;277(9):2407-2412 [FREE Full text] [doi: 10.1007/s00405-020-05979-9] [Medline: 32342197]
32. Hoek J, Bloemendal KM, van der Velden LA, van Diessen JN, van Werkhoven E, Klop WM, et al. Nephrotoxicity as a dose-limiting factor in a high-dose cisplatin-based chemoradiotherapy regimen for head and neck carcinomas. Cancers (Basel) 2016 Feb 16;8(2):21 [FREE Full text] [doi: 10.3390/cancers8020021] [Medline: 26891330]
33. Wong AJ, Kanwar A, Mohamed AS, Fuller CD. Radiomics in head and neck cancer: from exploration to application. Transl Cancer Res 2016 Aug;5(4):371-382 [FREE Full text] [doi: 10.21037/tcr.2016.07.18] [Medline: 30627523]
34. Bogowicz M, Leijenaar RT, Tanadini-Lang S, Riesterer O, Pruschy M, Studer G, et al. Post-radiochemotherapy PET radiomics in head and neck cancer - the influence of radiomics implementation on the reproducibility of local control tumor models. Radiother Oncol 2017 Dec;125(3):385-391. [doi: 10.1016/j.radonc.2017.10.023] [Medline: 29122358]
35. Vallières M, Kay-Rivest E, Perrin LJ, Liem X, Furstoss C, Aerts HJ, et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. Sci Rep 2017 Aug 31;7(1):10117 [FREE Full text] [doi: 10.1038/s41598-017-10371-5] [Medline: 28860628]
36. Parmar C, Leijenaar RT, Grossmann P, Rios Velazquez E, Bussink J, Rietveld D, et al. Sci Rep 2015 Jun 05;5:11044 [FREE Full text] [doi: 10.1038/srep11044] [Medline: 26251068]

37. Parmar C, Grossmann P, Rietveld D, Rietbergen MM, Lambin P, Aerts HJ. Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer. Front Oncol 2015;5:272 [FREE Full text] [doi: 10.3389/fonc.2015.00272] [Medline: 26697407]
38. Jethanandani A, Lin TA, Volpe S, Elhalawani H, Mohamed AS, Yang P, et al. Exploring applications of radiomics in magnetic resonance imaging of head and neck cancer: a systematic review. Front Oncol 2018;8:131 [FREE Full text] [doi: 10.3389/fonc.2018.00131] [Medline: 29868465]
39. van Dijk LV, Langendijk JA, Zhai TT, Vedelaar TA, Noordzij W, Steenbakkers RJ, et al. Delta-radiomics features during radiotherapy improve the prediction of late xerostomia. Sci Rep 2019 Aug 28;9(1):12483 [FREE Full text] [doi: 10.1038/s41598-019-48184-3] [Medline: 31462719]
40. van Dijk LV, Thor M, Steenbakkers RJ, Apte A, Zhai TT, Borra R, et al. Parotid gland fat related Magnetic Resonance image biomarkers improve prediction of late radiation-induced xerostomia. Radiother Oncol 2018 Sep;128(3):459-466 [FREE Full text] [doi: 10.1016/j.radonc.2018.06.012] [Medline: 29958772]
41. Haddad R, O'Neill A, Rabinowits G, Tishler R, Khuri F, Adkins D, et al. Induction chemotherapy followed by concurrent chemoradiotherapy (sequential chemoradiotherapy) versus concurrent chemoradiotherapy alone in locally advanced head and neck cancer (PARADIGM): a randomised phase 3 trial. Lancet Oncol 2013 Mar;14(3):257-264. [doi: 10.1016/S1470-2045(13)70011-1] [Medline: 23414589]
42. Cohen EE, Karrison TG, Kocherginsky M, Mueller J, Egan R, Huang CH, et al. Phase III randomized trial of induction chemotherapy in patients with N2 or N3 locally advanced head and neck cancer. J Clin Oncol 2014 Sep 01;32(25):2735-2743 [FREE Full text] [doi: 10.1200/JCO.2013.54.6309] [Medline: 25049329]
43. Tardini E, Zhang X, Canahuate G, Wentzel A, Mohamed A, van Dijk L, et al. Data deposition for "Optimal policy determination in sequential systemic and locoregional therapy of oropharyngeal squamous carcinomas: a patient-physician digital twin dyad with deep Q-learning". figshare. 2020. URL: https://doi.org/10.6084/m9.figshare.13235453.v2 [accessed 2022-04-06]

Abbreviations

AJCC: American Joint Committee on Cancer
AR: aspiration rate
CC: concurrent chemotherapy
DP: dysphagia
DQL: deep Q-learning
FT: feeding-tube dependence
IC: induction chemotherapy
MDACC: MD Anderson Cancer Center
MDP: Markov decision process
ND: neck dissection
NN: neural network
OS: overall survival
RL: reinforcement learning
RPN: retropharyngeal lymph node
RT: radiotherapy
SVC: support vector classifier
TS: treatment simulator

Edited by G Eysenbach; submitted 07.04.21; peer-reviewed by G Lim, S Kim, S You; comments to author 09.07.21; revised version received 03.09.21; accepted 09.02.22; published 20.04.22.

Please cite as:
Tardini E, Zhang X, Canahuate G, Wentzel A, Mohamed ASR, Van Dijk L, Fuller CD, Marai GE
Optimal Treatment Selection in Sequential Systemic and Locoregional Therapy of Oropharyngeal Squamous Carcinomas: Deep Q-Learning With a Patient-Physician Digital Twin Dyad
J Med Internet Res 2022;24(4):e29455
URL: https://www.jmir.org/2022/4/e29455
doi:10.2196/29455
PMID:35442211

Original Paper

Distracting Through Procedural Pain and Distress Using Virtual Reality and Guided Imagery in Pediatric, Adolescent, and Young Adult Patients: Randomized Controlled Trial

Jennifer A Hoag[1], PhD; Jeffrey Karst[1], PhD; Kristin Bingen[1], PhD; Akasha Palou-Torres[1], MS; Ke Yan[1], PhD

Division of Pediatric Psychology and Developmental Medicine, Department of Pediatrics, Medical College of Wisconsin, Milwaukee, WI, United States

Corresponding Author:
Jennifer A Hoag, PhD
Division of Pediatric Psychology and Developmental Medicine
Department of Pediatrics
Medical College of Wisconsin
8701 W Watertown Plank Road
MFRC 3018
Milwaukee, WI, 53226
United States
Phone: 1 414 955 5738
Email: jhoag@mcw.edu

Abstract

Background: Children with acute and chronic illness undergo frequent, painful, and distressing procedures.

Objective: This randomized controlled trial was used to evaluate the effectiveness of guided imagery (GI) versus virtual reality (VR) on the procedural pain and state anxiety of children and young adults undergoing unsedated procedures. We explored the role of trait anxiety and pain catastrophizing in intervention response.

Methods: Children and young adults were recruited from the hematology, oncology, and blood and marrow transplant clinics at a children's hospital. Each study participant completed the GI and VR intervention during separate but consecutive unsedated procedures. Self-report measures of pain and anxiety were completed before and after the procedures.

Results: A total of 50 participants (median age 13 years) completed both interventions. GI and VR performed similarly in the management of procedural pain. Those with high pain catastrophizing reported experiencing less nervousness about pain during procedures that used VR than those using GI. State anxiety declined pre- to postprocedure in both interventions; however, the decrease reached the level of significance during the VR intervention only. Those with high trait anxiety had less pain during GI.

Conclusions: In our sample, VR worked as well as GI to manage the pain and distress associated with common procedures experienced by children with acute or chronic illnesses. Children who are primed for pain based on beliefs about pain or because of their history of chronic pain had a better response to VR. GI was a better intervention for those with high trait anxiety.

Trial Registration: ClinicalTrials.gov NCT04892160; https://clinicaltrials.gov/ct2/show/NCT04892160

(J Med Internet Res 2022;24(4):e30260) doi:10.2196/30260

KEYWORDS

virtual reality; procedural; pain; anxiety; pediatric; guided imagery

Introduction

Pain

According to the International Association for the Study of Pain, "pain is always a personal experience that is influenced to varying degrees by biological, psychological, and social factors" [1]. Studies estimate that as many as half of children with acute or chronic illnesses experience procedure-related pain and distress [2]. In the short term, pain can manifest as withdrawal, clinginess, moodiness, or anger. There is substantial evidence, however, that inadequately addressed pain in childhood is associated with neurological and behavioral outcomes, including increased pain sensitivity, over the life course [3]. For example, children with sickle cell disease with a higher frequency of

painful vaso-occlusive episodes are more likely to report heightened pain responses during venipuncture [4].

Children with cancer, sickle cell disease, and other blood disorders undergo routine procedures over many months or years. Unsurprisingly, pain from diagnostic procedures and treatment is one of the most frequently cited physical problems in children undergoing cancer treatment [5]. Over the past 40 years, there has been a trend toward increased pain control through the use of sedation and analgesia; however, there are risks to sedation, including hypoxia, that outweigh the benefits in recurrent and routine procedures. The identification and use of nonpharmacological interventions to manage pain could mitigate the risk of neurological and behavioral changes that result from poorly managed pain without the risks of sedation.

Guided Imagery

Distraction is an effective and readily available nonpharmacologic tool for pain management [6,7]. It suppresses the highly salient sensations of pain and anxiety by consciously shifting attention to a more pleasant activity or thought. Guided imagery (GI) is a powerful nonimmersive distraction that involves describing in detail a situation incompatible with the experience of pain and is meant to evoke feelings of calm. GI scripts often begin with brief relaxation exercises, such as diaphragmatic breathing, followed by a vivid description of a relaxing activity, such as walking along a beach, flying among the clouds, or participating in a campfire. It is widely regarded as useful in decreasing pain and anxiety during procedures that do not warrant pharmacologic intervention [8-11].

Virtual Reality

Virtual reality (VR) is an immersive, 3D, interactive technology that engages multiple senses and creates an artificial environment that the user can inhabit. VR has been used to assist with pediatric procedural distress in several contexts over the past 20 years such as burn care [12,13], dental procedures [14-16], intravenous needle sticks [17-25], and port access [26-28]. Reviews of VR use have been positive, with most suggesting that VR is a feasible and efficacious method of distraction that can reduce patient-reported pain and distress [29].

While there have been numerous studies comparing VR to no intervention [20,28] and VR to standard of care (primarily access to television or tablets [17,21,22]), there have been no studies directly comparing the widely accepted nonimmersive distraction of GI to the promising immersive distraction of VR. Since procedure-related pain cannot be avoided, it is important to investigate which intervention provides the most relief and whether there are subcategories of children who respond better to one intervention over another. For example, research has demonstrated that individuals who are primed for pain and hold catastrophic beliefs about pain have more difficulty being distracted during painful experiences [30-32]. Similarly, state anxiety, a fluid variable that describes one's current level of anxiety, is predictive of pain tolerance and pain-related anxiety [33]. In their experiment using noxious electrical stimuli, Tang and Gibson [34] found that even when state anxiety was lower, individuals with high trait anxiety (ie, a stable variable that indicates greater disposition to experience anxiety) still reported higher subjective pain intensity ratings than those with low trait anxiety. Johnson [35] posited that the more distracting the stimuli, the greater the reduction in an individual's capacity to process pain and feel distressed. There is also evidence that active distraction techniques are more beneficial for pain management than passive approaches [36].

In this study, we directly compare the effects of VR and GI during an unsedated procedure on subjective and objective measures of pain and anxiety. We hypothesize that the VR intervention will be associated with decreased experiences of procedural pain and distress as compared to the nonimmersive GI intervention. We further hypothesize that the impact of VR on reducing procedural pain and anxiety will be greater in pediatric patients who have higher levels of pain catastrophizing and greater state and trait anxiety.

Methods

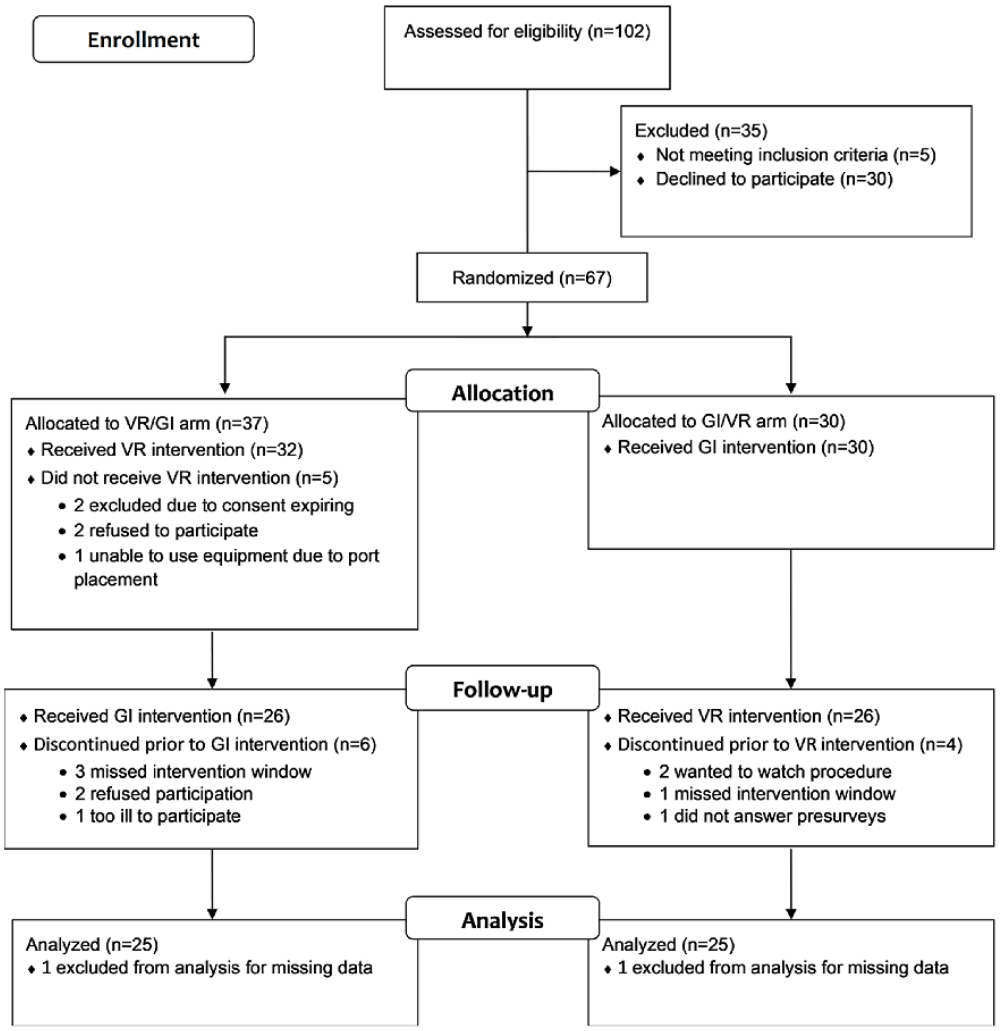
Study Design

This was a single-site, crossover, randomized controlled trial (RCT) used to evaluate the effectiveness of GI versus VR on the procedural pain and distress of children and young adults undergoing unsedated procedures. A convenience sample of participants was recruited from the hematology, oncology, and blood and marrow transplant services at a large tertiary children's hospital in Wisconsin. Data were collected between February 2018 and April 2019, at which point the threshold of enrolled patients had been reached. The interventions included (1) nonimmersive distraction via a 15-minute audio recording of a guided imagery script and (2) immersive distraction using KindVR Aqua (KindVR LLC), a virtual reality game that runs over 15 minutes. Conditions were counterbalanced so that there were 2 possible condition orders (VR/GI and GI/VR).

Participants

Children and young adults aged 8 to 25 years seen by the hematology, oncology, or blood and marrow transplant services were eligible if they were at least 1-month postdiagnosis and undergoing one of the following unsedated procedures: venipuncture, port access, or peripherally inserted central-line catheter or central venous line dressing change. Patients were excluded if they were not able to read or speak English proficiently, had identified physical impairments (eg, blindness, active infection of the skin, history of seizure disorder) that would have prevented them from using VR equipment, or had significant developmental delays that would have prevented them from completing required study questionnaires. There were 102 children and young adults screened for eligibility. Of these, 34.3% (35/102) were excluded for the reasons indicated in the CONSORT (Consolidated Standards on Reporting Trials) flow diagram (Figure 1). Of the remaining 67 participants, 37 were randomized into the VR/GI arm and 30 were randomized into the GI/VR arm. A total of 52 participants completed both interventions, and 2 were excluded from the final analysis due to missing data, resulting in a final sample of 50 participants.

Figure 1. Consolidated Standards on Reporting Trials flow diagram. GI: guided imagery; VR: virtual reality.



Study personnel identified eligible patients via the clinic schedule and inpatient census. After enrollment, study participants were randomly assigned to one of 2 possible condition orders using an online random number generator [37]. All attempts were made to ensure the study conditions (VR, GI) took place over the course of 2 consecutive procedures, excluding unplanned or emergent procedures. Conditions were separated by a minimum of 5 days and a maximum of 40 days to minimize the threat of treatment artifacts and extraneous events. The minimum time limit ensured that patients would not participate in 2 conditions within the same calendar week, while the maximum time limit allowed participation of patients who receive treatment on approximately a monthly basis. Both time points for an individual participant involved the same procedure type (ie, venipuncture, port access, or dressing change). Participants completed preprocedure questionnaires and, 3 to 5 minutes before the nurse entered the room for the procedure, the intervention was started. Each intervention lasted approximately 15 minutes. At the end of the procedure, participants completed their postprocedure questionnaires. Study personnel remained in the room to provide technical assistance and complete an observational measure of distress (ie, Children's Emotional Manifestation Scale [CEMS]).

Interventions

VR Condition

The VR intervention consisted of an interactive audio and visual underwater experience. The VR software used was KindVR Aqua, a research-based game focused on reducing pain and distress during medical procedures. Aqua offered both passive and active gameplay. In the passive experience, the software moved participants through an ocean filled with sea creatures and allowed them to observe the underwater scene. In the active experience, participants launched balls at the sea creatures. When hit, the creatures turned a variety of bright colors and points were earned. Participants were encouraged but not required to actively participate to increase the level of distraction. Study personnel recorded whether the participant participated in the active portion of the VR program. We used an off-the-shelf consumer headset (Gear VR, Samsung) powered by a Samsung smartphone with over-ear, noise-cancelling headphones. A wireless controller was used to interact with the underwater environment; the controller could be used with one hand if the procedure necessitated. The equipment did not require internet capability.

GI Condition

The GI script used in this study described an underwater scene that closely mimicked the VR condition. Similar to other GI scripts, ours began with instructions to take a few deep breaths to aid in relaxation. We then offered vivid descriptions of swimming underwater, which were similar to those seen in the VR intervention (eg, "You hear off in the distance the faint, yet beautiful songs of friendly whales talking to one another while making their way through the clear blue water. A sea turtle then glides past you. His face and flippers are patterned with spots of deep tan and brown, reflecting the rays of sunlight streaming through the water"). The script was audiorecorded on a tablet. Participants listened to the recording using over-ear, noise-cancelling headphones to approximate the headphones used during the VR condition.

Measures

Patient perceptions of pain were assessed prior to the first procedure using either the child (for participants aged 8-16 years) or adult (for participants aged 17 years and older) version of the Pain Catastrophizing Scale (PCS) [38]. The PCS is a 13-question survey that assesses thoughts and feelings related to pain, specifically catastrophic thinking about pain, on a 5-point scale ranging from 1 (not at all) to 5 (extremely). The measure includes a total score and rumination, magnification, and helplessness subscales. Scores range from 0 to 52 with scores >30 considered to indicate an elevated level of catastrophic thinking.

A visual analog scale (VAS) was used to assess pain and distress after each procedure. This measure asked about 4 domains of pain, including worst pain, average amount of pain, nervousness about pain, and time spent thinking about pain [39]. Scores range from 0 to 100 with higher scores indicating worsening symptoms.

The CEMS [40] was completed by study personnel during each of the study visits. This observational measure offers an objective way to measure distress during difficult medical experiences. It includes the following domains: anxiety score, facial expression, vocalization, activity, interaction, and level of cooperation. Each domain is scored using a 5-point scale. Total scores range from 1 to 25 with higher scores indicating more distress.

Trait (underlying or baseline) anxiety was assessed prior to the first procedure using the trait portion of either the child (for participants aged 8-16 years) or adult (for participants aged 17 years and older) version of the State-Trait Anxiety Inventory (STAI), while state (in the moment) anxiety was assessed prior to and following each procedure using the state portion of the STAI [41]. The STAI includes 40 items in 2 subscales (state and trait anxiety). Subscale scores for children range from 20 to 60, and scores for adults range from 20 to 80. Higher scores indicate greater anxiety. Scores were transformed to z-scores for the purposes of comparison.

A demographic survey was completed by the patient or patient's caregiver prior to study completion. Relevant treatment history was collected from the patient's medical record. This information included date of birth, diagnosis date, dates of procedures, treatment type, and relapse status. All dates were removed once calculations were made (eg, time between visits, age at diagnosis).

Ethical Approval

The institutional review board at Children's Wisconsin reviewed and approved all study documents and protocols (1110230-13), and the study was registered at ClinicalTrials.gov [NCT04892160]. All study participants and their caregivers were informed about the study in person by a clinical research coordinator prior to completing any study measures. Caregivers signed consent forms for their child's participation, while all patients aged 7 years and older signed assent forms. Study recruitment was conducted by clinical research coordinators and principal investigator (JAH). Randomization and assessment were conducted and intervention fidelity was assessed by the clinical research coordinators.

Statistical Methods

We performed a power analysis when designing the study. For simplicity, we used a 2-sided paired *t* test at a Bonferroni corrected =.025. With 50 participants, we have at least 80% power to detect a difference of .45 standard deviations. Categorical data are summarized as frequency and percentage and continuous data as median and IQR. Study groups (young adults vs children, groups with different orders of interventions, and diagnosis groups) were compared using chi-square or Fisher exact tests for categorical variables and Mann-Whitney or Kruskal-Wallis tests for continuous variables. Paired data pre- versus postintervention and GI versus VR were analyzed using a Wilcoxon signed-rank test. Pearson correlations assessed relationships between continuous variables such as pain and anxiety. Statistical software used included SAS (version 9.4, SAS Institute Inc), SPSS (version 26, IBM Corp), and R (version 3.6.0, R Foundation for Statistical Computing). Unadjusted $P<.05$ was considered statistically significant.

Results

Demographics

Participant demographics are presented in Table 1. A total of 50 participants completed both interventions and were included in the analyses. The median time from diagnosis to study participation was 2.1 (IQR 0.2-8.1) years. There were no differences in randomization groups by participant age, gender, race, type of procedure, or household income. Participant age and gender did not differ across the 3 diagnostic groups (ie, cancer, sickle cell disease, other). Race ($P<.001$) and type of procedure ($P<.001$) were significantly different with a larger percentage of White individuals and port access procedures in participants with cancer and a larger percentage of African Americans and venipuncture procedures in participants with sickle cell disease. These results were expected and reflect racial differences in risk of disease and differences in standard treatment. Nearly all participants (47/50, 94%) engaged in active play during the VR intervention.

Table 1. Participant demographics (n=50).

	Value
Male, n (%)	26 (52)
Age (years), median (IQR)	13 (11-16)
Race, n (%)	
White	26 (52)
Black	18 (36)
Other	6 (12)
Procedure, n (%)	
Venipuncture	13 (26)
Port access	26 (52)
Dressing change	11 (22)
Diagnosis, n (%)	
Cancer	31 (62)
Sickle cell disease	12 (24)
Other	7 (14)
Parent education, n (%)	
High school	13 (26)
Some college	16 (32)
Bachelor degree	7 (14)
Graduate degree	5 (10)
Unknown	9 (18)
Household income (US $), n (%)	
<25,000	8 (16)
25,000-49,999	9 (18)
50,000-74,999	6 (12)
75,000-99,999	3 (6)
>100,000	9 (18)
Unknown	15 (30)

Procedural Pain Outcomes

Self-reported pain scores on the VAS ranged from 0 to 100 across interventions. Scores for worst pain, average pain, nervousness about pain, and time spent thinking about pain did not differ between GI and VR. Similarly, there were no significant differences between interventions in CEMS score. There were no differences between the pain ratings of children and young adults in either intervention.

State Anxiety Outcomes

The preprocedure state anxiety scores did not differ between GI and VR (median z-scores –0.38 vs –0.34, respectively, *P*=.24), nor did postprocedure state anxiety scores (median z-scores –0.53 vs –0.69, respectively, *P*=.44). When comparing the change from pre- to postprocedure, there was a significant decline in state anxiety reported for the VR intervention (median z-scores –0.34 vs –0.69, *P*<.001) and no significant change in the GI intervention (*P*=.07). There were no differences between children and young adults in state anxiety scores in either the GI or VR intervention.

Relationship Between Procedural Pain and State Anxiety

In the GI intervention, there was a significant relationship between preprocedure state anxiety and nervousness about pain and time spent thinking about pain but not worst pain or average pain (Table 2). Postprocedure state anxiety following the GI intervention was significantly related to all areas of self-reported pain, including worst pain, average pain, nervousness about pain, and time spent thinking about pain. In the VR intervention, ratings of preprocedure state anxiety were significantly related to worst pain, average pain, nervousness about pain, and time spent thinking about pain (Table 2). Similarly, in the VR intervention, ratings of postprocedure state anxiety were significantly related to worst pain, nervousness about pain, and time spent thinking about pain. Following the VR intervention, state anxiety was no longer related to ratings of average pain.

Table 2. Relationship between pre- and postprocedural pain and state anxiety.

	State anxiety			
	GI[a]		VR[b]	
	Pre	Post	Pre	Post
Worst pain				
r[c]	0.28	0.46	0.47	0.33
P value	.05	<.001	<.001	.02
Average pain				
r	0.18	0.31	0.49	0.26
P value	.20	.03	<.001	.07
Nervousness about pain				
r	0.45	0.38	0.48	0.40
P value	.001	.01	<.001	.004
Time spent thinking about pain				
r	0.45	0.38	0.56	0.51
P value	.001	.01	<.001	<.001

[a]GI: guided imagery.

[b]VR: virtual reality.

[c]*r*: estimate of Pearson product-moment correlation coefficient.

Impact of Pain Catastrophizing on Procedural Pain and Anxiety

Of the participants, 14% (7/50) had an elevated total pain catastrophizing score (>30). Greater levels of pain catastrophizing were associated with worst pain experienced during the procedure for both interventions (Table 3). Increased helplessness was associated with worst pain for participants in the GI intervention but not the VR intervention. Rumination and magnification were not related to worst pain.

Greater levels of pain catastrophizing were associated with higher average pain experienced during the procedure for both interventions (Table 3). Increased magnification and helplessness were associated with higher average pain for participants in both interventions, whereas rumination was not related to average pain in either intervention.

Greater levels of pain catastrophizing were associated with more nervousness about experiencing pain in the GI but not the VR intervention (Table 3). Increased rumination and helplessness were associated with more nervousness for participants in the GI but not the VR intervention. Magnification was not related to nervousness about pain in either intervention.

Table 3. Relationship between procedural pain and pain catastrophizing.

	Rumination		Magnification		Helplessness		PCS[a] total	
	GI[b]	VR[c]	GI	VR	GI	VR	GI	VR
Worst pain								
r[d]	0.22	0.26	0.22	0.28	0.33	0.25	0.3	0.29
P value	.13	.07	.13	.06	.02	.09	.04	.04
Average pain								
r	0.28	0.25	0.32	0.4	0.4	0.37	0.39	0.38
P value	.06	.09	.03	.005	.005	.01	.007	.009
Nervousness about pain								
r	0.29	0.05	0.21	0.11	0.35	0.08	0.34	0.09
P value	.05	.72	.16	.46	.02	.60	.02	.56
Time spent thinking about pain								
r	0.17	0.08	0.07	0.16	0.16	0.21	0.16	0.17
P value	.26	.60	.63	.29	.27	.15	.27	.24
CEMS[e]								
r	0.16	0.25	0.18	0.21	0.22	0.17	0.22	0.23
P value	.28	.08	.21	.15	.14	.26	.14	.11

[a]PCS: Pain Catastrophizing Scale.

[b]GI: guided imagery.

[c]VR: virtual reality.

[d]*r*: estimate of Pearson product-moment correlation coefficient.

[e]CEMS: Children's Emotional Manifestation Scale.

There was no relationship between pain catastrophizing and time spent thinking about pain in either intervention (Table 3). There was no relationship between pain catastrophizing and CEMS ratings of pain in either intervention (Table 3). There was no relationship between pain catastrophizing and pre- or postprocedure state anxiety (Table 4). There was a significant relationship between pain catastrophizing and trait anxiety (r=0.44, P=.002).

Table 4. Relationship between anxiety and pain catastrophizing.

	Rumination		Magnification		Helplessness		PCS[a] total	
	GI[b]	VR[c]	GI	VR	GI	VR	GI	VR
Preprocedure state anxiety								
r[d]	0.16	0.18	0.20	0.10	0.21	0.14	0.22	0.16
P value	.27	.24	.18	.51	.15	.36	.14	.28
Postprocedure state anxiety								
r	0.11	0.09	0.14	0.05	0.27	0.23	0.21	0.16
P value	.47	.56	.33	.73	.07	.12	.15	.28

[a]PCS: Pain Catastrophizing Scale.

[b]GI: guided imagery.

[c]VR: virtual reality.

[d]*r*: estimate of Pearson product-moment correlation coefficient.

Impact of Trait Anxiety on Procedural Pain and Anxiety

There was a significant relationship between trait anxiety and all areas of self-reported procedural pain in the VR intervention (Table 5). In the GI intervention, trait anxiety was significantly related to worst pain and nervousness about pain but not average pain or time spent thinking about pain (Table 5). There was a significant relationship between trait anxiety and CEMS score during the GI but not the VR intervention (Table 5).

During both interventions, higher levels of trait anxiety were significantly correlated with higher pre- and postprocedure measures of state anxiety (GI preprocedure: r=0.58, P<.001; GI postprocedure: r=0.43, P=.002; VR preprocedure: r=0.42, P=.003; VR postprocedure: r=0.37, P=.01).

Table 5. Relationship between procedural pain and anxiety.

	Trait anxiety	
	GI[a]	VR[b]
Worst pain		
r[c]	0.30	0.43
P value	.03	.002
Average pain		
r	0.24	0.48
P value	.09	<.001
Nervousness about pain		
r	0.30	0.51
P value	.04	<.001
Time spent thinking about pain		
r	0.10	0.51
P value	.48	<.001
CEMS[d]		
r	0.41	0.21
P value	.003	.14

[a]GI: guided imagery.

[b]VR: virtual reality.

[c]*r*: estimate of Pearson product-moment correlation coefficient.

[d]CEMS: Children's Emotional Manifestation Scale.

Disease Group Differences

When participating in GI, there were no differences between diagnostic groups in procedure pain scores on the VAS. When participating in VR, there were no differences between diagnostic groups in worst pain (P=.61), average pain (P=.57), time spent thinking about pain (P=.27), or CEMS score (P=.70). There were, however, significant differences between diagnostic groups in the level of nervousness reported during the VR intervention (P=.04). Pairwise comparisons revealed that the significant differences were between cancer and other diagnoses (median scores 12 vs 0, P=.01), with more nervousness reported by those with cancer. There were no differences between cancer and sickle cell (P=.43) or sickle cell and other (P=.13). There were no significant differences between the diagnostic groups on the PCS total score or the PCS subscales.

Across interventions, there were no pre- or postprocedure anxiety score differences for patients in any disease group. The change of pre- to postanxiety scores differed significantly in the sickle cell disease group with greater declines in anxiety during VR compared to GI (median change in z-scores –0.14 vs 0.07, P=.03). Trait anxiety was comparable across diagnoses (P=.53).

Discussion

Principal Findings

Poorly managed procedure-related pain is acutely distressing and can lead to increased pain sensitivity throughout the lifespan, lowering the likelihood of seeking medical care as an adult [3]. This RCT was designed to identify low-risk, nonpharmacological options to manage pain and distress experienced during recurrent procedures where sedation is unwarranted. Specifically, we compared a highly effective distraction strategy, GI, with a less well known but more immersive strategy, VR. In general, we found the interventions performed similarly in their management of procedural pain. Across interventions, the majority of participants rated the procedures as causing a low level of pain; however, there was a subset of participants who self-reported high levels of pain, demonstrating the need to identify patients at higher risk of pain and distress.

Similar to previous literature [30-32], we found that the subgroup that held unhelpful beliefs about pain (ie, those with high pain catastrophizing scores) reported higher levels of worst pain and average pain across both interventions. Confirming our hypothesis, those with high pain catastrophizing reported experiencing less nervousness about pain during procedures

that used VR than those using GI. Our findings suggest that the mechanism by which VR lessened the impact of pain catastrophizing on the experience of pain was twofold—by acting on feelings of helplessness and rumination about pain. We attribute these findings to the more immersive nature of VR, which increases the cognitive distraction to pain [35].

We had anticipated that those with more anxiety, whether state or trait, would have a more powerful response to VR. We did not expect there to be differences between how those with high state and trait anxiety responded to the interventions. State anxiety declined pre- to postprocedure during the VR intervention, specifically in participants who started with higher state anxiety. State anxiety is transient and situational. A good distractor consumes most of one's cognitive energies, leaving little capacity to process pain and anxiety [42]. It makes sense, then, that those participants who were highly anxious about the procedure had a powerful response to the immersive nature of VR.

As a more global and stable construct, those with high trait anxiety responded differently. Contrary to our hypothesis, GI disrupted the relationships between trait anxiety and the variables of average pain and time spent thinking about pain but VR did not. Previous research has suggested that those with high trait anxiety experience greater increases in physiological arousal when presented with stress and more accurately perceive these changes compared to those with low trait anxiety [43,44]. The GI intervention offered 2 components that may have been a better fit for those with high trait anxiety: brief guidance in diaphragmatic breathing to reduce physiological arousal and the ability to watch the procedure. This finding suggests that VR may be more efficacious for children who are generally well adjusted but evidence a high degree of distress around a procedure, whereas GI may be more efficacious for those who have preexisting or chronic anxiety.

All study participants, regardless of their diagnosis, had a similar pain response during GI. Children and young adults with cancer, however, were more nervous about experiencing pain during the VR intervention than those in the other category. This may be a function of length of disease or procedure type (ie, port access vs dressing change). Participants with sickle cell disease responded with a more powerful reduction in anxiety when using VR than GI. While it is unclear why, we know that children and adolescents with sickle cell disease have a different pain trajectory than those with other diseases [45]. The transition from acute and intermittent vaso-occlusive pain crises in childhood to chronic pain in adolescence is well documented but remains poorly understood [46]. There were no differences in trait anxiety or beliefs about pain by diagnosis.

Limitations and Future Research

This RCT crossover study compared 2 distraction interventions used during unsedated procedures in a real-world pediatric medical population. Although this study has several strengths, limitations must be noted. First, the selected procedures elicited low levels of pain and anxiety, which may have diluted differences between the interventions. Future research that includes more painful or frightening procedures, such as lumbar puncture or nasogastric tube placement, would advance the findings of this study. Had we known that the procedures in this study were so well tolerated for most participants, we would have screened for distress and excluded those who did not meet a minimum threshold.

A second limitation of this study is the lack of a control intervention. We determined that providing supportive distraction is the standard of care during procedures at our institution; therefore, we did not feel it was ethical to withhold. However, a control intervention with marginal distraction, such as engaging in conversation or encouraging the child to watch television, could have been used and would have served as a useful comparison to better elucidate the benefits of more distracting interventions such as GI and VR.

A third limitation of this study is the heterogeneity of the sample selected, including different diagnoses and procedures. While this yielded more efficient recruitment and realistic representation of children seen in a hematology, oncology, and blood and marrow transplant setting, it also introduced confounding variables that threatened the internal validity of the study. Future research should include either a more homogenous sample (eg, only children with sickle cell disease, only children with high pain catastrophizing) or a larger number of participants such that subgroups can be examined with greater confidence and a smaller margin of error.

Clinical Implications and Conclusion

This study shows that, in general, VR works as well as GI to manage the pain and distress associated with common procedures experienced by children with an acute or chronic illness. We found that children who are primed for pain, based on beliefs about pain or because of their history of chronic pain, have a better response to VR. GI is a better intervention for those with high trait anxiety who may benefit from a greater sense of control when able to watch the procedure. As medical treatments are increasingly tailored at the individual level, mental health providers too need to give more thought to the power of individualized interventions. This new information advances our understanding of who may benefit more from GI and VR.

Acknowledgments

This research was supported in part by grants from Northwestern Mutual Foundation and the Midwest Athletes Against Childhood Cancer Fund.

Authors' Contributions

JAH, JK, and KB contributed to the design and implementation of the research, analysis of the results, and writing of the manuscript. APT collected the data. KY analyzed the data.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 647 KB - jmir_v24i4e30260_app1.pdf]

References

1. Merskey H, Bogduk N. Classification of Chronic Pain. 2nd Edition. Seattle: IASP Press; 1994.
2. Sobol-Kwapińska M, Sobol M, Woźnica-Niesobska E. Parental behavior and child distress and pain during pediatric medical procedures: systematic review and meta-analysis. Health Psychol 2020 Jul;39(7):558-572. [doi: 10.1037/hea0000864] [Medline: 32281822]
3. Young KD. Pediatric procedural pain. Ann Emerg Med 2005 Feb;45(2):160-171. [doi: 10.1016/j.annemergmed.2004.09.019] [Medline: 15671974]
4. Schlenz AM, McClellan CB, Mark TRM, McKelvy AD, Puffer E, Roberts CW, et al. Sensitization to acute procedural pain in pediatric sickle cell disease: modulation by painful vaso-occlusive episodes, age, and endothelin-1. J Pain 2012 Jul;13(7):656-665 [FREE Full text] [doi: 10.1016/j.jpain.2012.04.001] [Medline: 22633685]
5. Hedström M, Haglund K, Skolin I, von Essen L. Distressing events for children and adolescents with cancer: child, parent, and nurse perceptions. J Pediatr Oncol Nurs 2003;20(3):120-132. [doi: 10.1053/jpon.2003.76] [Medline: 12776260]
6. Vessey JA, Carlson KL, McGill J. Use of distraction with children during an acute pain experience. Nurs Res 1994;43(6):369-372. [Medline: 7971302]
7. Sahiner NC, Bal MD. The effects of three different distraction methods on pain and anxiety in children. J Child Health Care 2016 Sep;20(3):277-285. [doi: 10.1177/1367493515587062] [Medline: 26040282]
8. Huth MM, Broome ME, Good M. Imagery reduces children's post-operative pain. Pain 2004 Jul;110(1-2):439-448. [doi: 10.1016/j.pain.2004.04.028] [Medline: 15275797]
9. Landier W, Tse AM. Use of complementary and alternative medical interventions for the management of procedure-related pain, anxiety, and distress in pediatric oncology: an integrative review. J Pediatr Nurs 2010 Dec;25(6):566-579 [FREE Full text] [doi: 10.1016/j.pedn.2010.01.009] [Medline: 21035021]
10. Powers SW. Empirically supported treatments in pediatric psychology: procedure-related pain. J Pediatr Psychol 1999 Apr;24(2):131-145. [doi: 10.1093/jpepsy/24.2.131] [Medline: 10361392]
11. Uman LS, Birnie KA, Noel M, Parker JA, Chambers CT, McGrath PJ, et al. Psychological interventions for needle-related procedural pain and distress in children and adolescents. Cochrane Database Syst Rev 2013 Oct 10(10):CD005179. [doi: 10.1002/14651858.CD005179.pub3] [Medline: 24108531]
12. Das DA, Grimmer KA, Sparnon AL, McRae SE, Thomas BH. The efficacy of playing a virtual reality game in modulating pain for children with acute burn injuries: a randomized controlled trial [ISRCTN87413556]. BMC Pediatr 2005 Mar 03;5(1):1 [FREE Full text] [doi: 10.1186/1471-2431-5-1] [Medline: 15745448]
13. Jeffs D, Dorman D, Brown S, Files A, Graves T, Kirk E, et al. Effect of virtual reality on adolescent pain during burn wound care. J Burn Care Res 2014;35(5):395-408. [doi: 10.1097/BCR.0000000000000019] [Medline: 24823326]
14. Asl Aminabadi N, Erfanparast L, Sohrabi A, Ghertasi Oskouei S, Naghili A. The impact of virtual reality distraction on pain and anxiety during dental treatment in 4-6 year-old children: a randomized controlled clinical trial. J Dent Res Dent Clin Dent Prospects 2012;6(4):117-124 [FREE Full text] [doi: 10.5681/joddd.2012.025] [Medline: 23277857]
15. Atzori B, Lauro Grotto R, Giugni A, Calabrò M, Alhalabi W, Hoffman HG. Virtual reality analgesia for pediatric dental patients. Front Psychol 2018;9:2265 [FREE Full text] [doi: 10.3389/fpsyg.2018.02265] [Medline: 30532720]
16. Shetty V, Suresh LR, Hegde AM. Effect of virtual reality distraction on pain and anxiety during dental treatment in 5 to 8 year old children. J Clin Pediatr Dent 2019;43(2):97-102. [doi: 10.17796/1053-4625-43.2.5] [Medline: 30730798]
17. Caruso TJ, George A, Menendez M, De Souza E, Khoury M, Kist MN, et al. Virtual reality during pediatric vascular access: a pragmatic, prospective randomized, controlled trial. Paediatr Anaesth 2020 Feb;30(2):116-123. [doi: 10.1111/pan.13778] [Medline: 31785015]
18. Dumoulin S, Bouchard S, Ellis J, Lavoie KL, Vézina M, Charbonneau P, et al. A randomized controlled trial on the use of virtual reality for needle-related procedures in children and adolescents in the emergency department. Games Health J 2019 Aug;8(4):285-293. [doi: 10.1089/g4h.2018.0111] [Medline: 31135178]
19. Gerçeker G, Binay Ş, Bilsin E, Kahraman A, Yılmaz HB. Effects of virtual reality and external cold and vibration on pain in 7- to 12-year-old children during phlebotomy: a randomized controlled trial. J Perianesth Nurs 2018 Dec;33(6):981-989. [doi: 10.1016/j.jopan.2017.12.010] [Medline: 29559294]

20. Özalp Gerçeker G, Ayar D, Özdemir EZ, Bektaş M. Effects of virtual reality on pain, fear and anxiety during blood draw in children aged 5-12 years old: a randomised controlled study. J Clin Nurs 2020 Apr;29(7-8):1151-1161. [doi: 10.1111/jocn.15173] [Medline: 31889358]
21. Gold JI, Mahrer NE. Is virtual reality ready for prime time in the medical space? A randomized control trial of pediatric virtual reality for acute procedural pain management. J Pediatr Psychol 2018 Apr 01;43(3):266-275. [doi: 10.1093/jpepsy/jsx129] [Medline: 29053848]
22. Goldman RD, Behboudi A. Virtual reality for intravenous placement in the emergency departmenta: a randomized controlled trial. Eur J Pediatr 2021 Mar;180(3):725-731. [doi: 10.1007/s00431-020-03771-9] [Medline: 32779029]
23. Piskorz J, Czub M. Effectiveness of a virtual reality intervention to minimize pediatric stress and pain intensity during venipuncture. J Spec Pediatr Nurs 2018 Jan;23(1):e12201. [doi: 10.1111/jspn.12201] [Medline: 29155488]
24. Walther-Larsen S, Petersen T, Friis SM, Aagaard G, Drivenes B, Opstrup P. Immersive virtual reality for pediatric procedural pain: a randomized clinical trial. Hosp Pediatr 2019 Jul;9(7):501-507. [doi: 10.1542/hpeds.2018-0249] [Medline: 31160472]
25. Wong CL, Li CK, Chan CWH, Choi KC, Chen J, Yeung MT, et al. Virtual reality intervention targeting pain and anxiety among pediatric cancer patients undergoing peripheral intravenous cannulation: a randomized controlled trial. Cancer Nurs 2021;44(6):435-442. [doi: 10.1097/NCC.0000000000000844] [Medline: 32511154]
26. Birnie KA, Kulandaivelu Y, Jibb L, Hroch P, Positano K, Robertson S, et al. Usability testing of an interactive virtual reality distraction intervention to reduce procedural pain in children and adolescents with cancer [formula: see text]. J Pediatr Oncol Nurs 2018;35(6):406-416. [doi: 10.1177/1043454218782138] [Medline: 29950139]
27. Gershon J, Zimand E, Lemos R, Rothbaum BO, Hodges L. Use of virtual reality as a distractor for painful procedures in a patient with pediatric cancer: a case study. Cyberpsychol Behav 2003 Dec;6(6):657-661. [doi: 10.1089/109493103322725450] [Medline: 14756933]
28. Nilsson S, Finnström B, Kokinsky E, Enskär K. The use of Virtual Reality for needle-related procedural pain and distress in children and adolescents in a paediatric oncology unit. Eur J Oncol Nurs 2009 Apr;13(2):102-109. [doi: 10.1016/j.ejon.2009.01.003] [Medline: 19230769]
29. 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]
30. Crombez G, Eccleston C, Baeyens F, Eelen P. Attentional disruption is enhanced by the threat of pain. Behav Res Ther 1998 Feb;36(2):195-204. [doi: 10.1016/s0005-7967(97)10008-0] [Medline: 9613025]
31. Crombez G, Eccleston C, Baeyens F, Eelen P. When somatic information threatens, catastrophic thinking enhances attentional interference. Pain 1998 Apr;75(2-3):187-198. [doi: 10.1016/s0304-3959(97)00219-4] [Medline: 9583754]
32. Van Damme S, Crombez G, Eccleston C. Disengagement from pain: the role of catastrophic thinking about pain. Pain 2004 Jan;107(1-2):70-76. [doi: 10.1016/j.pain.2003.09.023] [Medline: 14715391]
33. Cimpean A, David D. The mechanisms of pain tolerance and pain-related anxiety in acute pain. Health Psychol Open 2019;6(2):2055102919865161 [FREE Full text] [doi: 10.1177/2055102919865161] [Medline: 31827877]
34. Tang J, Gibson SJ. A psychophysical evaluation of the relationship between trait anxiety, pain perception, and induced state anxiety. J Pain 2005 Sep;6(9):612-619. [doi: 10.1016/j.jpain.2005.03.009] [Medline: 16139780]
35. Johnson MH. How does distraction work in the management of pain? Curr Pain Headache Rep 2005 Apr;9(2):90-95. [doi: 10.1007/s11916-005-0044-1] [Medline: 15745617]
36. Wohlheiter KA, Dahlquist LM. Interactive versus passive distraction for acute pain management in young children: the role of selective attention and development. J Pediatr Psychol 2013 Mar;38(2):202-212. [doi: 10.1093/jpepsy/jss108] [Medline: 23092971]
37. Research Randomizer. URL: https://www.randomizer.org/ [accessed 2022-03-17]
38. Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: development and validation. Psychol Assess 1995;7(4):524-532. [doi: 10.1037/1040-3590.7.4.524]
39. Hoffman HG, Doctor JN, Patterson DR, Carrougher GJ, Furness TA. Virtual reality as an adjunctive pain control during burn wound care in adolescent patients. Pain 2000 Mar;85(1-2):305-309. [doi: 10.1016/s0304-3959(99)00275-4] [Medline: 10692634]
40. Li HCW, Lopez V. Children's Emotional Manifestation Scale: development and testing. J Clin Nurs 2005 Feb;14(2):223-229. [doi: 10.1111/j.1365-2702.2004.01031.x] [Medline: 15669931]
41. Spielberger C, Gorsuch R, Lushene R, Vagg P, Jacobs G. Manual for the State-Trait Anxiety Inventory. Palo Alto: Consulting Psychologists Press; 1983.
42. McCaul KD, Malott JM. Distraction and coping with pain. Psychol Bull 1984 May;95(3):516-533. [Medline: 6399756]
43. Noteboom JT, Barnholt KR, Enoka RM. Activation of the arousal response and impairment of performance increase with anxiety and stressor intensity. J Appl Physiol (1985) 2001 Nov;91(5):2093-2101 [FREE Full text] [doi: 10.1152/jappl.2001.91.5.2093] [Medline: 11641349]
44. Richards JC, Bertram S. Anxiety sensitivity, state and trait anxiety, and perception of change in sympathetic nervous system arousal. J Anxiety Disord 2000;14(4):413-427. [doi: 10.1016/s0887-6185(00)00031-1] [Medline: 11043889]

45. Takaoka K, Cyril AC, Jinesh S, Radhakrishnan R. Mechanisms of pain in sickle cell disease. Br J Pain 2021 May;15(2):213-220. [doi: 10.1177/2049463720920682] [Medline: 34055342]
46. Brandow AM, Zappia KJ, Stucky CL. Sickle cell disease: a natural model of acute and chronic pain. Pain 2017 Apr;158 Suppl 1:S79-S84 [FREE Full text] [doi: 10.1097/j.pain.0000000000000824] [Medline: 28221286]

Abbreviations

CEMS: Children's Emotional Manifestation Scale
CONSORT: Consolidated Standards on Reporting Trials
GI: guided imagery
PCS: Pain Catastrophizing Scale
RCT: randomized controlled trial
STAI: State-Trait Anxiety Inventory
VAS: visual analog scale
VR: virtual reality

Edited by R Kukafka; submitted 18.05.21; peer-reviewed by S Said Yekta-Michael, D Lin, M Linnett, C Son; comments to author 01.09.21; revised version received 06.12.21; accepted 03.01.22; published 18.04.22.

Please cite as:
Hoag JA, Karst J, Bingen K, Palou-Torres A, Yan K
Distracting Through Procedural Pain and Distress Using Virtual Reality and Guided Imagery in Pediatric, Adolescent, and Young Adult Patients: Randomized Controlled Trial
J Med Internet Res 2022;24(4):e30260
URL: https://www.jmir.org/2022/4/e30260
doi:10.2196/30260
PMID:35436209

Original Paper

Pre2Pub—Tracking the Path From Preprint to Journal Article: Algorithm Development and Validation

Lisa Langnickel[1,2], MSc; Daria Podorskaja[1,3], MSc; Juliane Fluck[1,4], Prof Dr

[1]ZB MED - Information Centre for Life Sciences, Cologne, Germany

[2]Graduate School Digital Infrastructure in the Life Sciences, Bielefeld Institute for Bioinformatics Infrastructure, Faculty of Technology, Bielefeld University, Bielefeld, Germany

[3]Bonn-Aachen International Center for Information Technology, University of Bonn, Bonn, Germany

[4]The Agricultural Faculty, University of Bonn, Bonn, Germany

Corresponding Author:
Lisa Langnickel, MSc
ZB MED - Information Centre for Life Sciences
Gleueler Straße 60
Cologne, 50931
Germany
Phone: 49 2287360355
Email: langnickel@zbmed.de

Abstract

Background: The current COVID-19 crisis underscores the importance of preprints, as they allow for rapid communication of research results without delay in review. To fully integrate this type of publication into library information systems, we developed preview: a publicly available, central search engine for COVID-19–related preprints, which clearly distinguishes this source from peer-reviewed publications. The relationship between the preprint version and its corresponding journal version should be stored as metadata in both versions so that duplicates can be easily identified and information overload for researchers is reduced.

Objective: In this work, we investigated the extent to which the relationship information between preprint and corresponding journal publication is present in the published metadata, how it can be further completed, and how it can be used in preVIEW to identify already republished preprints and filter those duplicates in search results.

Methods: We first analyzed the information content available at the preprint servers themselves and the information that can be retrieved via Crossref. Moreover, we developed the algorithm Pre2Pub to find the corresponding reviewed article for each preprint. We integrated the results of those different resources into our search engine preVIEW, presented the information in the result set overview, and added filter options accordingly.

Results: Preprints have found their place in publication workflows; however, the link from a preprint to its corresponding journal publication is not completely covered in the metadata of the preprint servers or in Crossref. Our algorithm Pre2Pub is able to find approximately 16% more related journal articles with a precision of 99.27%. We also integrate this information in a transparent way within preVIEW so that researchers can use it in their search.

Conclusions: Relationships between the preprint version and its journal version is valuable information that can help researchers finding only previously unknown information in preprints. As long as there is no transparent and complete way to store this relationship in metadata, the Pre2Pub algorithm is a suitable extension to retrieve this information.

(J Med Internet Res 2022;24(4):e34072) doi:10.2196/34072

KEYWORDS

preprints; information retrieval; COVID-19; metadata; BERT; Bidirectional Encoder Representations from Transformers

Introduction

The publication of non–peer-reviewed research manuscripts, called preprints, has gained popularity in recent years. Several studies demonstrate the benefits of publishing such preprints; for example, based on the number of citations [1,2]. In addition, the current COVID-19 pandemic shed new light on this type of publication because it allows researchers to communicate new findings quickly, since the publication process is not slowed by peer review.

Although preprints can be a very valuable source of information, especially in times of a pandemic, there is a wide range of quality that cannot be assessed without close examination of the content. The World Health Organization titles the overabundance of both correct and incorrect information during a disease outbreak an "infodemic" [3]. Therefore, preprints must be carefully integrated into existing information infrastructures. Several search portals already include preprints in their database: Europe PMC has been including them for several years [4]; in response to the current health crisis, PubMed launched a pilot project and is integrating COVID-19–related preprints from various preprint servers [5].

In order to ensure rapid central access to COVID-19 preprints and to clearly separate this publication type from reviewed prints, we developed a new preprint service: preVIEW COVID-19 [6,7]. This semantic search engine currently combines more than 40,000 COVID-19–related preprints from 7 different sources. As preprints seem to be established as common first publication type, the problem of content information duplication arises. Consequently, more preprints and journal articles have almost the same content. Especially for articles on COVID-19, where all published information is considered, it is of great importance that as soon as a preprint has been published in a peer-reviewed journal, this information is contained in the metadata and displayed on search portals. Otherwise, this leads to an unnecessary overload for information specialist reading and selecting relevant articles twice. For example, Europe PMC included a filter function where the information is queried from Crossref [8,9]. Crossref is a not-for-profit organization that—besides the allocation of digital object identifiers (DOIs)—offers tools and methods "to help the research community discover, link, cite, and assess scholarly content" [10]. Even though Crossref stores metadata for a total amount of 93,035 journals, books, and conference proceedings [11], Fraser et al [12] recently showed that the information about a corresponding journal article is not completely represented in available metadata. The authors developed an algorithm—based on fuzzy string matching—to automatically query Scopus to find a corresponding journal article.

In this paper, we present the new quality-checked and open source algorithm Pre2Pub, which queries the open access search engine PubMed to find the appropriate journal publications for preprints on the basis of text and author matches. Pre2Pub makes use of recent advances in natural language processing and integrates a Bidirectional Encoder Representations from Transformers (BERT)-based model (sentence-BERT) [13], is evaluated, and is subjected to error analysis. We show that this information is missing in 16% of preprints, making the work of information seekers more difficult, especially during the COVID-19 pandemic. Furthermore, using the semantic search engine preVIEW as an example, this paper presents how the Pre2Pub results can be used as additional information and filtering options in a user-friendly way.

Methods

Methods Overview

First, we describe our approach to collect the data from bioRxiv and medRxiv and provide an overview of the generated data sets. Afterward, the developed Pre2Pub algorithm is explained. Finally, we describe the implementation of Pre2Pub into our semantic search engine preVIEW.

Data

We retrieved data from 2019, 2020, and 2021 from bioRxiv and medRxiv [14]. The servers provide a shared application programming interface (API) for COVID-19–related preprints [15], which is used to distinguish between COVID-19 and non–COVID-19 documents. For further analysis, we stored the following metadata: preprint's DOI, title, authors list, publication date, published server (bioRxiv or medRxiv), abstract, and (if available) the DOI for the corresponding peer reviewed journal article. Second, we accessed the Crossref API [16] for all retrieved preprints and used the *is-preprint-of* function to determine whether a link to a corresponding peer reviewed version is stored there.

The data consist of a total of 132,339 preprints. Of these, a corresponding journal article link can be found for 51,957 preprints via Crossref and for 62,748 preprints in the metadata of the preprint server. For training and evaluation, we randomly selected 4000 preprints from 2019 and 2020 for which a corresponding journal article is referenced in the preprint server. These were split into training and testing data, with 2000 preprints each. An overview of the retrieved data is provided in Table 1.

Table 1. Overview of collected data from bioRxiv and medRxiv from 2019, 2020, and 2021.

Collected data	Preprints, n	Journal digital object identifiers in bioRxiv or medRxiv, n	Journal digital object identifiers in Crossref, n	Distribution (COVID-19/ non–COVID-19), n/n
Total collected data	132,339	62,748	51,957	21,846/110,493
Training data	2000	2000	1433	194/1806
Test data	2000	2000	1452	173/1827

Pre2Pub Algorithm

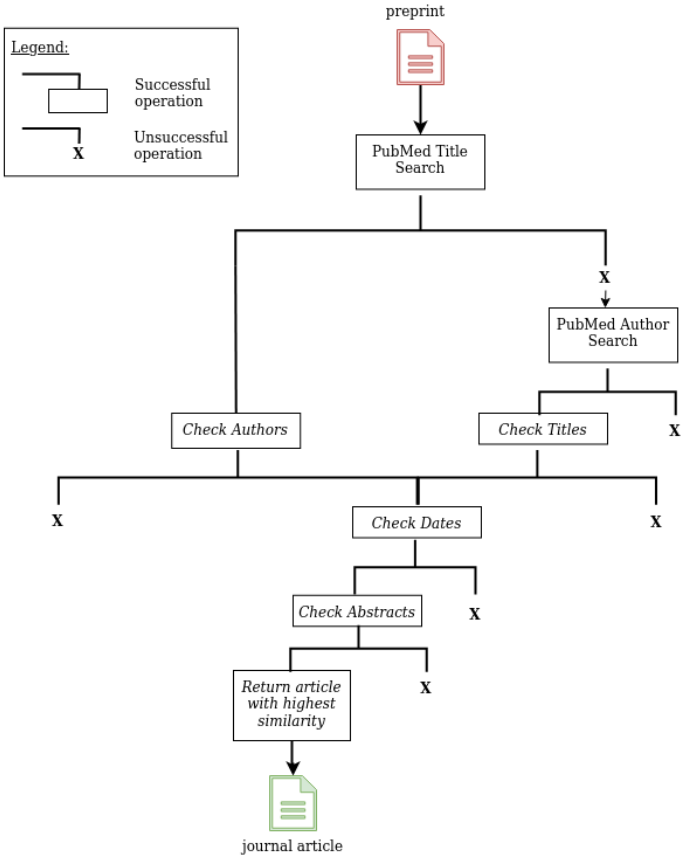
This algorithm uses a preprint DOI as input and searches for a corresponding journal article in PubMed. We use the E-utilities NCBI service to retrieve PubMed articles on the basis of a defined search query [17]. Pre2Pub consists of 5 steps that are shown in a simplified workflow overview in Figure 1 and described in detail as follows:

1. For all preprints, we used the cleaned title (ie, stop words are removed using the natural language toolkit [18]) as the search term to retrieve a list of PubMed IDs in the *title* field. Retrieval maximum is set to 5 to avoid false-positive results.

2. If the search using the preprint title was successful, the authors of the preprint and the authors of the matched PubMed journal articles are compared. The preprint servers provide the author names in different formats. For example, some servers provide full names with all first names; others provide only last names and first names' initials, as bioRxiv does. Because PubMed usually stores the complete names (last name, middle name, first name), we queried Crossref as well and extracted the complete preprint's author names there. To compare the standardized author names between preprint and journal formats, we used the Levenshtein ratio [19] and determined matches using several criteria iteratively developed with the help of the training data: first, we browsed the two lists of authors simultaneously and determine the Levenshtein ratio for each pair. If the value is greater than 0.9, we assume that the authors are identical. If this is not the case, we distinguish the following four cases to consider the authors of the preprints and the authors of the journal article as the same: first, if we find more matched than unmatched author pairs when iterating over the author lists (ie, based on consensus); second, if the first 3 authors of both lists are the same; third, if the first and the last author of preprint and journal article are identical; and fourth, if the first and last authors of the preprint are found in the author list of the journal article (regardless of position).
3. If the title search was not successful, we searched the list of authors in the *author* field. Retrieval maximum was set to 5 in both cases to avoid false-positive results. If successful, we compared the titles of the preprint and the fetched journal articles. To accomplish this, we generated embeddings by making use of Sentence-BERT, "a modification of the pretrained BERT network that use siamese and triplet network structures to derive semantically meaningful sentence embeddings that can be compared using cosine-similarity" [13]. As a pretrained model, we load BioBERT (BioBERT-Base v1.0 [+ PubMed 200K + PMC 270K]) [20]. The threshold is set to 0.95.
4. We compared the dates of publication to ensure that the date of the preprint is older than the date of the peer-reviewed article.
5. We compared the abstract of the preprint to those of the PubMed articles that passed the date check and either the author or title check—depending on which search was successful in the beginning. We generated embeddings and applied the same method as for the title check described in step (3).

Finally, we chose the article with the highest abstract similarity value and, under the condition that it exceeds the threshold of 0.9, linked the found article to the preprint.

Figure 1. Simplified overview of Pre2Pub's working mechanism. The input is the metadata of a preprint. Then, by querying the PubMed application programming interface, a title search is performed. If this is successful, the authors are compared by our 5-step method. If not, we perform an author search directly in PubMed and compare then the titles. For the matching articles, the dates are checked and the pairwise abstract similarity is determined. Finally, the article with the highest similarity is returned—if it exceeds the threshold of 0.9.



Evaluation Metrics

To evaluate the algorithm, we used the independent test set and determined the precision, recall, and F_1-score. If our algorithm finds the same DOI, it is considered a true positive; if our algorithm does not yield a matching journal article, we consider it a false negative occurrence; if our algorithm yields another DOI, this is considered both a false positive and a false negative.

Implementation in preVIEW

As quality differences can exist between a preprint and a peer-reviewed journal article, the availability of a filter option in search portals is important. To display information on whether a preprint is already published in a journal, we integrated Pre2Pub in our semantic search engine preVIEW. Therefore, we updated the information on a weekly basis in the following way: for each preprint, we determined whether a corresponding journal link can be retrieved via the corresponding preprint server (currently only implemented for bioRxiv and medRxiv). If we did not obtain a result, we queried Crossref. If this also failed, we used Pre2Pub. This workflow is applied for all preprints where no information about a corresponding journal article is available. In contrast, if a publication link is already found via the preprint server or Crossref, no update is carried out. For the preprints where either no publication was found previously or the information was only found by Pre2Pub, we checked for updates in the preprint server and Crossref to ensure high quality.

Results

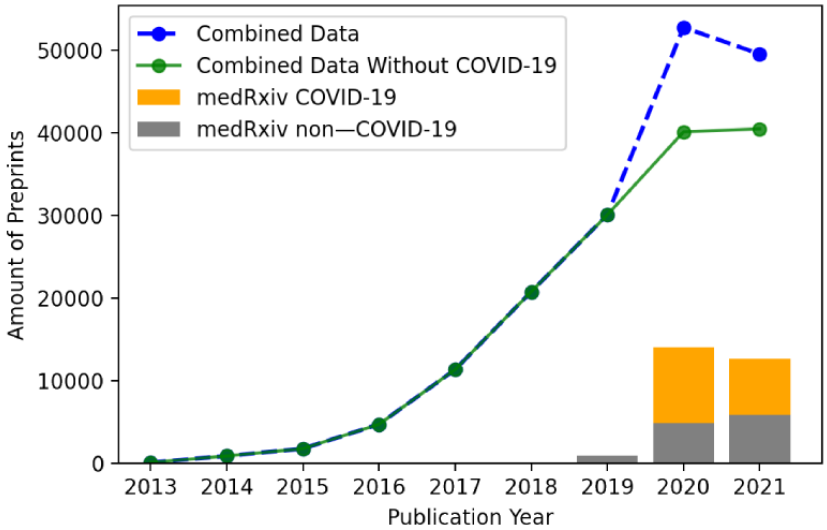
Results Overview

This section first provides a general overview of the growth of preprints during the last decade. We then analyzed the collected data set in terms of the number and amount of corresponding journal publications for both preprint servers in the last 2 years. Afterward, we evaluated our developed algorithm on the basis of the independent test set, including error analyses. We further investigated differences in the path from preprint to journal publication for COVID-19–related and non–COVID-19 publications. Finally, we describe the integration into preVIEW's user interface.

Above-Average Growth of Preprints in 2020

The COVID-19 crisis led to an increased use of preprint servers because they allow rapid dissemination of research results. The preprint server bioRxiv was launched in 2013 and comprised both biological and medical topics [21]. In the mid-2019, the server medRxiv has been launched additionally to separate these two fields [21]. While only 109 preprints were submitted to bioRxiv in 2013, a total of 30,094 preprints were already submitted to bioRxiv or medRxiv in 2019. As depicted in Figure 2, the publication of COVID-19–related preprints steepens the publication curve and leads to an above-average growth of preprints in 2020.

Figure 2. Growth of preprint publications. Combined data include data from both bioRxiv and medRxiv. The blue line comprises all available data, and the green line indicates the amount of preprints minus COVID-19–related papers. Only the medRxiv proportion is depicted as a bar diagram, thereby distinguishing COVID-19 and non–COVID-19 articles.



Amount of Preprints and Corresponding Journal Articles

Significantly more papers are uploaded on bioRxiv than on medRxiv. However, an increase in popularity can be seen in both cases. Whereas only 913 preprints have been posted on medRxiv in 2019, a total of 14,070 preprints were uploaded in 2020. This is partly because medRxiv has been launched in June 2019—before clinical and epidemiological papers were submitted to bioRxiv as well. Figure 3 provides an overview of the retrieved data sets from the preprint servers for the whole years 2019 and 2020.

We independently determined the absolute number of links to a corresponding journal article found by three different methods: (1) the bioRxiv or medRxiv API itself, (2) Crossref, and (3) our algorithm. For every category, the lowest number of articles can be found on Crossref. For 2021, Pre2Pub yielded the most matching publications. For example, for bioRxiv preprints, 9672 corresponding journal articles are referenced in their API. However, Pre2Pub yielded a total of 11,184 matching journal articles.

While we determined the absolute numbers in the first step, we further investigated the overlap among these three different methods: the Venn diagram in Figure 4 indicates the unique amount that each of the applied methods generates and the respective overlaps, determined on the joint data set (ie, bioRxiv and medRxiv from 2019, 2020, and 2021). Pre2Pub matches a total of 63,590 journal papers—51,473 of which could also be found via the APIs of bioRxiv or medRxiv and Crossref. Hence, 12,117 journal articles were only found by Pre2Pub. This makes up 16.2% of all found journal articles (74,880).

Figure 3. Overview of found journal articles in different sources and through different methods. In the 6 plots, the first bar indicates the total amount of preprints deposited in the preprint servers. Afterward, the absolute amount of found peer-reviewed publications is indicated for all three methods investigated: via the preprint server application programming interface itself, via Crossref, and via Pre2Pub. DOI: digital object identifier.

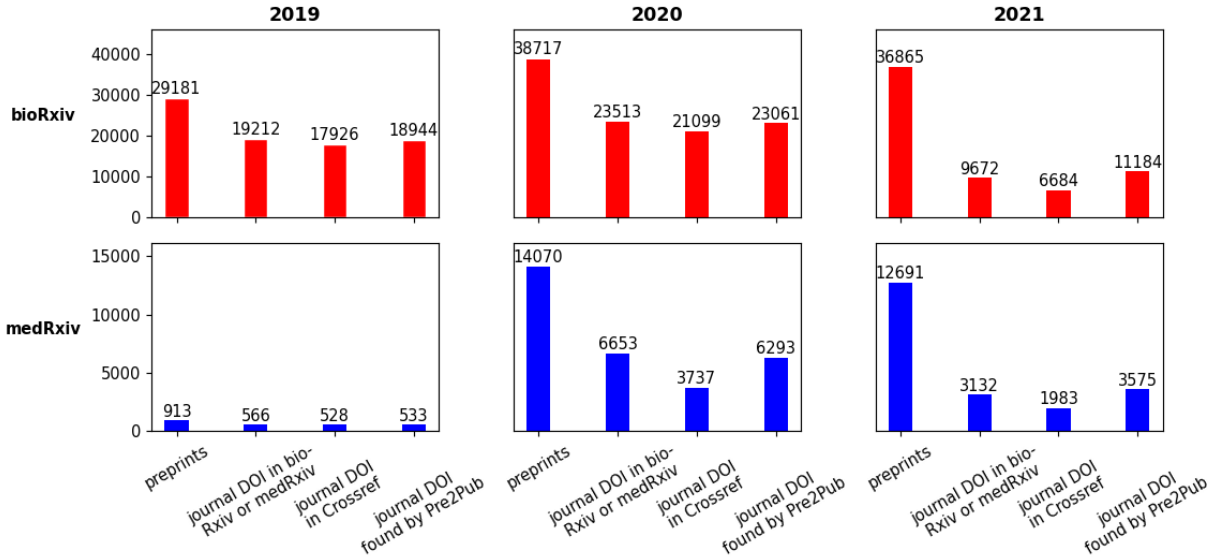
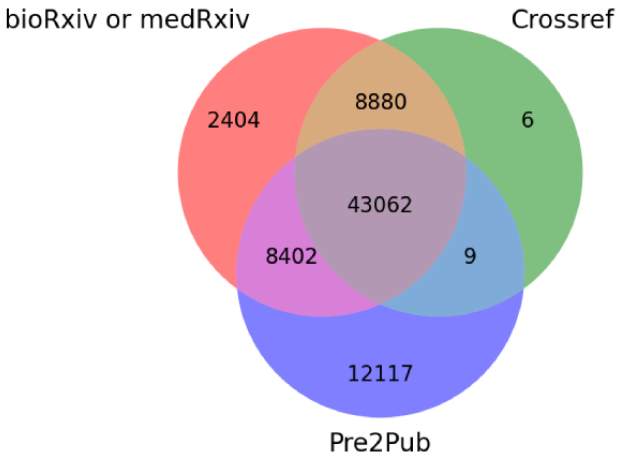


Figure 4. Amount and intersection of found journal articles for 3 different methods. For the combined data set from bioRxiv and medRxiv for 2019, 2020, and 2021, the amount of journal articles is shown for each method. While 43,062 articles were found by all methods, 12,117 articles were only found by Pre2Pub.



Evaluation of Pre2Pub

To investigate how reliably Pre2Pub identifies the corresponding journal articles for the preprints, we evaluated Pre2Pub using a test set of 2000 documents; thus, the same document count was used for training. The evaluation results are summarized in Table 2. Similar to training, the evaluation on the independent test set shows promising results with an F_1-score of 89.29%, resulting from a precision of 99.27% and a recall of 83.29%. On both, the training and the test sets, we achieve a much higher precision than recall; this result is desired because this ensures a higher probability of a false-negative prediction than a false-positive one. The latter would lead to misinformation and therefore needs to be avoided.

An overview of some false-positive and some false-negative matches can be found in Table 3. For some of the false-positive matches, the bioRxiv gold standard was not correct. In the first example, bioRxiv links to an excerpt of conference abstracts and Pre2Pub provides the correct journal article. Similarly, the second example depicts an error on the side of the preprint server. The DOI that is stored in the metadata does not exist—the last 2 digits are missing. In contrast, Pre2Pub finds the correct one.

Table 2. Evaluation results of Pre2Pub on training and test data.

Data set	Precision, %	Recall, %	F_1-score, %
Training data	99.10	82.64	90.13
Test data	99.27	81.14	89.29

Table 3. Examples of false-positive and false-negative matches found by Pre2Pub.

Preprint digital objective identifier	Journal digital object identifier found		Error Analysis
	bioRxiv/medRxiv	Pre2Pub	
10.1101/669713	10.1016/j.bpj.2019.11.1890	10.1016/j.bpj.2020.02.011	bioRxiv links only to conference abstracts, whereas Pre2Pub finds the corresponding article
10.1101/845933	10.1016/j.cub.2020.03.005	10.1111/ejn.15056	The correct article is not indexed in PubMed; the found article has the same first and last author and describes a related topic
10.1101/503763	10.1016/j.neuroimage.2019.1161	10.1016/j.neuroimage.2019.116186	Broken link at bioRxiv (incomplete digital object identifier); Pre2Pub finds the correct article
10.1101/549840	10.1128/mBio.00388-19	—[a]	No results via the application programming interface[b]
10.1101/2020.03.06.980631	10.1039/D0GC00903B	—	Article is not indexed in PubMed
10.1101/405597	10.1523/JNEUROSCI.0555-20.2020	—	Titles differ too much

[a]—: not determined.

[b]The search result retrieved from the application programming interface occasionally differing from a manual PubMed search

In the third example, Pre2Pub predicted an incorrect article with a related topic published by the same authors. The correct article is not indexed in PubMed and could therefore not be found. This is one of the major error sources for false negatives since our search is restricted to PubMed; hence, journal articles that are not indexed in PubMed cannot be found. The availability in PubMed has also an influence on false-positive results. If another similar article is found in PubMed with at least partly the same authors, this can lead to a false-positive match. If the correct article is present in PubMed as well, it would be a higher-ranked match; therefore, a false-positive entry would have been omitted.

The last example provides another reason for false negatives: if the titles differ too much, it may happen that the corresponding article is not within the search result, which is restricted to 5 articles.

Comparison of Preprint to Journal Traversal for COVID-19–Related and Non–COVID-19 Articles

To gain further insight into publication activity before and during the COVID-19 pandemic, we investigated the percentage of preprints that are republished in a journal in 2019, 2020, and 2021. For 2020 and 2021, we differentiated between COVID-19–related and non–COVID-19 articles. While approximately 75% of preprints were republished in a journal in 2019, only a total of 69% of them were published in 2020. Percentage wise, in 2020, more non–COVID-19 preprints than COVID-19–related preprints were published in a journal (69% vs 58%, respectively). However, in 2021, the trend was reversed and approximately 5% more COVID-19–related articles are published in comparison to non–COVID-19 articles. Owing to the large difference in sample amounts when comparing COVID-19 and non–COVID-19 articles, no significant conclusion could be drawn. Moreover, the numbers of published articles are much lower for 2021. This can be explained by the fact that it can take more than a year to finally publish the paper in the journal and the published preprints still undergo review. The results are summarized in Table 4.

Table 4. Overview of the amount of bio- and medRxiv preprints that are republished in a journal in 2019, 2020, and 2021.

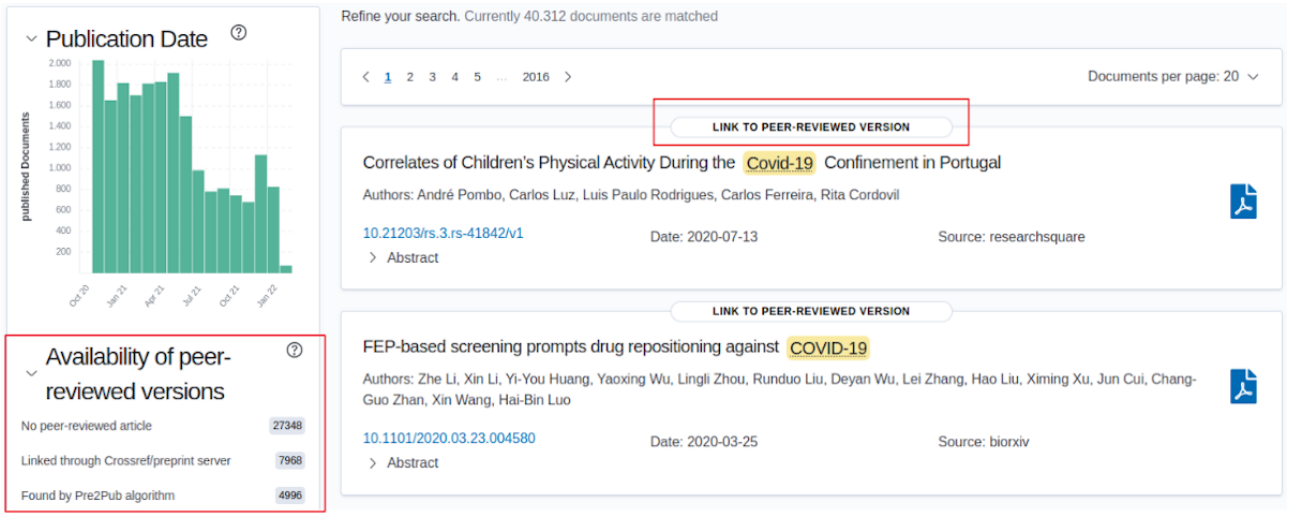
Year	COVID-19	Amount, n	Published, n (%)
2019	No	30,094	22,776 (75.68%)
2020	No	40,862	28,177 (68.96%)
2020	Yes	11,918	6920 (58.06%)
2021	No	40,441	13,489 (33.35%)
2021	Yes	9024	3518 (38.98%)

Resulting User Interface Integration

The information about corresponding republished journal versions of the preprint as well as filter functions for search were integrated within our search engine preVIEW. Thus, in our interface, we decided to distinguish between automatically mapped journal links and those provided by the preprint server or Crossref (Figure 5). This yielded three filter options found under the heading “Availability of peer-reviewed version”: (1) no reviewed article, (2) linked through preprint server or Crossref, and (3) found by Pre2Pub algorithm. Importantly, we ran Pre2Pub only for those preprints, where we found a linked journal article neither via the preprint server's API nor via Crossref.

Currently, out of 40,312 preprints, 12,964 have been already published in a journal (approximately 32%). Of them, 7968 corresponding journal publications were found through Crossref or the bioRxiv or medRxiv preprint server. Additional 4996 publications were found by Pre2Pub only—which makes up more than one-third of all found journal articles.

Figure 5. Integration in the search engine preVIEW, publicly available [22]. On the left, an overview of the amount of available peer-reviewed articles can be seen, which can be directly added to the search. On top of each article, if found, the link to the journal publication is provided and opens in a new tab by clicking.



Discussion

Principal Findings

In this paper, we first show that preprints have found their place in researchers' publishing behavior, and second that there is currently an information gap when a peer-reviewed version of a preprint appears, making the preprint irrelevant. Third, we provide a solution—the Pre2Pub algorithm—for identifying corresponding journal articles in PubMed and the integration of this information in the preVIEW search engine.

Despite the differences in quality, which may exist between a peer-reviewed and a non–peer-reviewed article, preprints are an appropriate and fast way to publish new results. Preprints have found their place in the publication workflow, as can be clearly seen in Figure 2, which shows an overview of the growth of preprints between 2013 and 2021. This is also true for medical preprints, with a marked increase in medRxiv publications in 2020—the beginning of the COVID-19 pandemic. The main reason for this increase was the high demand for information to quickly gain knowledge to address this crisis.

However, the increasing tendency to publish a preprint first and the peer-reviewed version afterward implies that similar information is published twice, making it difficult to search for relevant information—especially during the COVID-19 pandemic, when preprints have become more relevant. Therefore, it is important to make the information search as transparent as possible and to facilitate the identification and filtering of duplicates.

Consequently, for the publication workflow, it becomes imperative that information about links between publication versions such as preprints and corresponding peer reviewed versions is collected at each replication and is offered by services such as Crossref and DataCite.

During our research, and as also described by Fraser et al [12], we faced the problem that often the information of a journal publication cannot be found in the preprint metadata or via Crossref, although the article can be found in PubMed by a researcher. The underlying reason is that the process is not well established. Often, authors have the option to update the preprint article information to indicate the new journal publication. Some journals also take care of updating this information themselves, ensuring that the preprint is no longer modified. In addition, bioRxiv has its own update process. However, there is neither a uniform solution nor a fixed procedure that an individual researcher must follow.

Until such a process is in place, the presented algorithm Pre2Pub can be used: it searches for the corresponding publication in PubMed to find additional links—in case no results were retrieved via the metadata. Pre2Pub performed robustly as seen in our evaluation. It approached an F_1-score of 89.92%, resulting from a higher precision than recall (99.27% vs 81.14%, respectively). The significantly higher and near-perfect precision is desirable to minimize the number of false positives. We find

more than 12,000 additional publication links for our collected data from the last 3 years (Figure 4). Especially, for 2021, Pre2Pub finds more corresponding journal links, also in absolute numbers. This shows that the process for this type of publication is not well defined and highlights the incompleteness of the metadata provided by the preprint servers.

As a result, duplicated information increases the workload for researchers, which should be avoided, especially in times of a pandemic.

Currently, as shown in our service preVIEW, approximately 32% of COVID-19 preprints are already published in a journal. This result indicates the relevance of such a filtering option in information retrieval to help users of such search portals; for example, information specialists, to find relevant knowledge quickly. As more than one-third of these links were only found by Pre2Pub, it is not enough to simply query Crossref, as is, for example, done by Europe PMC [8].

Next, we plan to extend our algorithm to other literature search portals to enhance our recall further and to integrate the resulting information into the Livivo literature search engine [23].

Conclusions

Owing to the increased importance of preprints, there is a need to update search engines toward changed information retrieval needs. The need to be able to link a preprint to its corresponding journal article and to filter duplicates was addressed with the retrieval of this information from different resources and the integration within the search frontend of preVIEW.

In the future, we need the commitment of the publishers to request this information from the authors and to store it in the metadata. Until this becomes a reality, services such as Pre2Pub—a freely available algorithm able to find a journal publication for a given preprint—are necessary to retrieve the missing links.

Acknowledgments

Development of the search engine preVIEW was carried out as part of the NFDI4Health Task Force COVID-19 [24]. We gratefully acknowledge the financial support of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; project 451265285).

Data and Code Availability

Our code is publicly available on GitHub [25].

Authors' Contributions

LL and JF both contributed to the conceptional ideas of this work. DP developed the initial software. LL revised the software and is the primary author of the manuscript. DP participated in drafting the manuscript; JF and LL revised the manuscript and all authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Fu D, Hughey J. Releasing a preprint is associated with more attention and citations for the peer-reviewed article. Elife 2019 Dec 06;8:e52646 [FREE Full text] [doi: 10.7554/eLife.52646] [Medline: 31808742]
2. Davis PM, Fromerth MJ. Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles? Scientometrics 2007 May;71(2):203-215. [doi: 10.1007/s11192-007-1661-8]
3. The Lancet Infectious Diseases. The COVID-19 infodemic. The Lancet Infectious Diseases 2020 Aug;20(8):875. [doi: 10.1016/s1473-3099(20)30565-x]
4. Levchenko M, Gou Y, Graef F, Hamelers A, Huang Z, Ide-Smith M, et al. Europe PMC in 2017. Nucleic Acids Res 2018 Jan 04;46(D1):D1254-D1260 [FREE Full text] [doi: 10.1093/nar/gkx1005] [Medline: 29161421]
5. PubMed. URL: https://pubmed.ncbi.nlm.nih.gov/ [accessed 2021-10-04]
6. Langnickel L, Baum R, Darms J, Madan S, Fluck J. COVID-19 preVIEW: Semantic Search to Explore COVID-19 Research Preprints. Stud Health Technol Inform 2021 May 27;281:78-82. [doi: 10.3233/SHTI210124] [Medline: 34042709]
7. Langnickel L, Darms J, Baum R, Fluck J. preVIEW: from a fast prototype towards a sustainable semantic search system for central. J EAHIL 2021;17(3). [doi: 10.32384/jeahil17484]
8. Michaud F. Using the Crossref REST API. Part 12 (with Europe PMC). Crossref. 2018 Oct 10. URL: https://www.crossref.org/blog/using-the-crossref-rest-api.-part-12-with-europe-pmc/ [accessed 2021-10-04]
9. Crossref. URL: https://www.crossref.org/ [accessed 2021-10-04]
10. Introducing Crossref, the basics. Journals & Authors. 2018 Nov 13. URL: https://jasolutions.com.co/crossref/introducing-crossref-basics/ [accessed 2022-03-30]
11. Browsable title list. Crossref. URL: https://www.crossref.org/documentation/reports/browsable-title-list/ [accessed 2021-10-04]

12. Fraser N, Momeni F, Mayr P, Peters I. The relationship between bioRxiv preprints, citations and altmetrics. Quantitative Science Studies 2020 Apr 01:1-21. [doi: 10.1162/qss_a_00043]
13. Reimers N, Gurevych I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. arXiv. Preprint posted online August 27, 2019 [FREE Full text] [doi: 10.18653/v1/d19-1410]
14. bioRxiv API Summary. bioRxiv. URL: https://api.biorxiv.org/details/biorxiv/2013-01-01/2013-12-31/0 [accessed 2022-03-30]
15. COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv. medRxiv. URL: https://api.biorxiv.org/covid19/0 [accessed 2022-03-30]
16. REST API. Crossref. URL: https://www.crossref.org/documentation/retrieve-metadata/rest-api/ [accessed 2022-03-30]
17. Sayers E. The E-utilities In-Depth: Parameters, Syntax and More. Bethesda, MD: National Center for Biotechnology Information; 2010.
18. Natural Language Toolkit. NLTK Project. URL: https://www.nltk.org [accessed 2022-03-30]
19. Levenshtein V. Binary codes capable of correcting deletions, insertions and reversals. Soviet Physics Doklady 1966;10(8):707-710.
20. Lee J, Yoon W, Kim S, Kim D, Kim S, So C, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 2020 Feb 15;36(4):1234-1240 [FREE Full text] [doi: 10.1093/bioinformatics/btz682] [Medline: 31501885]
21. Sever R, Roeder T, Hindle S, Sussman L, Black KJ, Argentine J, et al. bioRxiv: the preprint server for biology. bioXriv. Preprint posted online November 6, 2019 . [doi: 10.1101/833400]
22. preVIEW: COVID-19. URL: https://preview.zbmed.de [accessed 2022-03-31]
23. Müller B, Poley C, Pössel J, Hagelstein A, Gübitz T. LIVIVO - the Vertical Search Engine for Life Sciences. Datenbank Spektrum 2017;17(1):29-34 [FREE Full text] [doi: 10.1007/s13222-016-0245-2] [Medline: 29368755]
24. NFDI4Health: National Research Data Infrastructure for Personal Health Data. German Research Foundation. URL: https://www.nfdi4health.de/en/ [accessed 2022-03-31]
25. Pre2Pub. GitHub. URL: https://github.com/zbmed/Pre2Pub [accessed 2022-03-31]

Abbreviations

API: application programming interface
BERT: Bidirectional Encoder Representations from Transformers
DOI: digital object identifier

Edited by C Lovis, C Basch; submitted 15.10.21; peer-reviewed by A Tomar, C Bakker; comments to author 19.01.22; revised version received 09.02.22; accepted 01.03.22; published 08.04.22.

Please cite as:
Langnickel L, Podorskaja D, Fluck J
Pre2Pub—Tracking the Path From Preprint to Journal Article: Algorithm Development and Validation
J Med Internet Res 2022;24(4):e34072
URL: https://www.jmir.org/2022/4/e34072
doi:10.2196/34072
PMID:35285808

Original Paper

Predicting Sepsis Mortality in a Population-Based National Database: Machine Learning Approach

James Yeongjun Park[1], MSc; Tzu-Chun Hsu[2], BSc; Jiun-Ruey Hu[3], MPH, MD; Chun-Yuan Chen[4], MD; Wan-Ting Hsu[5,6], MSc; Matthew Lee[6], JD, PhD; Joshua Ho[7], MSc; Chien-Chang Lee[2,7], MD, SCD

[1]Department of Biostatistics, Harvard TH Chan School of Public Health, Boston, MA, United States

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

[3]Department of Internal Medicine, Yale School of Medicine, New Haven, CT, United States

[4]Department of Medicine, National Taiwan University, Taipei, Taiwan

[5]Department of Epidemiology, Harvard TH Chan School of Public Health, Boston, MA, United States

[6]Medical Wizdom, LLC, Brookline, MA, United States

[7]Center of Intelligent Healthcare, National Taiwan University Hospital, Taipei, Taiwan

Corresponding Author:
Chien-Chang Lee, MD, SCD
Department of Emergency Medicine
National Taiwan University Hospital
Number 7, Chung-Shan South Road
Taipei, 100
Taiwan
Phone: 886 223123456
Email: hit3transparency@gmail.com

Abstract

Background: Although machine learning (ML) algorithms have been applied to point-of-care sepsis prognostication, ML has not been used to predict sepsis mortality in an administrative database. Therefore, we examined the performance of common ML algorithms in predicting sepsis mortality in adult patients with sepsis and compared it with that of the conventional context knowledge–based logistic regression approach.

Objective: The aim of this study is to examine the performance of common ML algorithms in predicting sepsis mortality in adult patients with sepsis and compare it with that of the conventional context knowledge–based logistic regression approach.

Methods: We examined inpatient admissions for sepsis in the US National Inpatient Sample using hospitalizations in 2010-2013 as the training data set. We developed four ML models to predict in-hospital mortality: logistic regression with least absolute shrinkage and selection operator regularization, random forest, gradient-boosted decision tree, and deep neural network. To estimate their performance, we compared our models with the Super Learner model. Using hospitalizations in 2014 as the testing data set, we examined the models' area under the receiver operating characteristic curve (AUC), confusion matrix results, and net reclassification improvement.

Results: Hospitalizations of 923,759 adults were included in the analysis. Compared with the reference logistic regression (AUC: 0.786, 95% CI 0.783-0.788), all ML models showed superior discriminative ability (P<.001), including logistic regression with least absolute shrinkage and selection operator regularization (AUC: 0.878, 95% CI 0.876-0.879), random forest (AUC: 0.878, 95% CI 0.877-0.880), xgboost (AUC: 0.888, 95% CI 0.886-0.889), and neural network (AUC: 0.893, 95% CI 0.891-0.895). All 4 ML models showed higher sensitivity, specificity, positive predictive value, and negative predictive value compared with the reference logistic regression model (P<.001). We obtained similar results from the Super Learner model (AUC: 0.883, 95% CI 0.881-0.885).

Conclusions: ML approaches can improve sensitivity, specificity, positive predictive value, negative predictive value, discrimination, and calibration in predicting in-hospital mortality in patients hospitalized with sepsis in the United States. These models need further validation and could be applied to develop more accurate models to compare risk-standardized mortality rates across hospitals and geographic regions, paving the way for research and policy initiatives studying disparities in sepsis care.

(J Med Internet Res 2022;24(4):e29982) doi:10.2196/29982

KEYWORDS

sepsis; mortality; machine learning; SuperLearner

Introduction

Background

Sepsis is a life-threatening condition caused by a dysregulated response of the body to infection. Sepsis is associated with high morbidity and mortality, increased health care expenditures, and long-term consequences [1-4]. It is a leading cause of hospitalization and death, with an estimated 850,000 emergency department visits per year and 59.6 deaths per 100,000 individuals in the United States [2,3]. The annual medical costs associated with sepsis are approximately US $24 billion in the United States [4]. There are clinical and economic incentives to improve and measure the quality of sepsis care in the United States [5]. Given the significant geographic disparities in sepsis outcomes, the development of robust severity adjustment tools is essential for objective sepsis mortality comparisons between hospitals.

Several tools to adjust for sepsis severity have been proposed by consensus conferences [6-8] using traditional statistical methods [9-11]. More recently, machine learning (ML) algorithms have improved the accuracy of sepsis mortality prediction models [12-16]. These tools were largely designed to incorporate the point-of-care risk stratification of patients into the clinical workflow [17-19]. Interhospital comparisons of sepsis care quality and evaluation of risk-adjusted sepsis outcomes have been difficult as the extraction of necessary data from each electronic medical record (EMR) system is time-consuming and not cost-effective [20,21]. Consequently, hospital administrative databases have gradually played a more prominent role and become more widely used by health service researchers because of their easy accessibility and inexpensiveness.

Existing efforts to test and refine sepsis mortality prediction models using hospital administrative data [20-23] have largely used logistic regression models and achieved satisfactory discrimination and calibration. More recent models adjusting for risk factors have made use of national administrative databases to compare risk-adjusted sepsis mortality between hospitals [24]. Although most of them achieved a good area under the receiver operating characteristic curve (AUC) in the range of 0.70-0.80, there is still room to improve their performance. In addition, these studies have focused on select academic centers with limited generalizability to other types of hospitals. Among them, the Severe Sepsis Mortality Prediction Model achieved the best performance with an AUC of 0.838 and was used to generate an integer-based score for risk adjustment in administrative data [21].

Objectives

ML models have a better ability to automatically select variables, handle large sets of variables, and detect complex multi-way interactions as well as nonlinear relationships [25]. These features enable ML models to improve on conventional regression models in predicting health-related outcomes [26]. In this study, we compare the outcomes of several ML algorithms to predict sepsis mortality using the full range of variables provided in the US Nationwide Inpatient Sample (NIS) database [27,28]. We also determine the accuracy among different derivation and validation models. Our objective is to provide an accurate and reliable tool to compare sepsis-related mortality between hospitals in the United States.

Methods

Identification of Cases

Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis was used in this study. The sepsis cohort was identified using the Martin implementation [29,30], which identifies cases with explicit codes from the International Classification of Diseases, Ninth Revision, Clinical Modification for sepsis or systemic fungal infection (038 septicemia, 020.0 septicemic, 790.7 bacteremia, 117.9 disseminated fungal infection, 112.5 disseminated Candida infection, or 112.81 disseminated fungal endocarditis) and a diagnosis of acute organ dysfunction. Seven acute organ or system dysfunctions were evaluated in this study: cardiovascular or shock, respiratory, central nervous system, hematologic, hepatic, renal, and metabolic system dysfunction. To reduce self-prophecy bias, we removed cases with cardiac arrest and ventricular fibrillation, respiratory failure, and respiratory insufficiency.

We split the data into a training set (NIS 2010-NIS 2013) and a testing set (NIS 2014). As the random forest and neural network models could not handle missing values, we removed patients with any missing values from the predictor variables. After removing patients with any missing values, the training data set included 726,918 adult patients, and the validation cohort included 196,841 adult patients.

Ethical Considerations

Our study involved analysis of de-identified patients from publicly available data. Therefore, no ethics approval was required by the Institutional Review Board (IRB).

Variables

We used 5-dimensional data as predictors (demographic characteristics, pre-existing comorbidities, hospital characteristics, diagnosis, and procedure performed on the first day of admission). A total of 1331 variables were included in the ML models. We compared our ML models with the reference model using the conventional logistic regression model with predictors reported in a previous study [21]. In the random forest model, we used the Gini Impurity to compute variable importance, where the improvement in the split criterion is the importance attributed to the splitting variable, and identified the top 50 variables based on the variable of importance values [31]. In addition, we calculated the Shapley Additive

Explanations (SHAP) values from the xgboost model. SHAP is a popular model-agnostic, local explanation approach designed to explain any given classifier. Lundberg and Lee [32] proposed the SHAP value as a united approach to explaining the output of any ML model. We calculated the SHAP values of each feature for each sample and extracted the top 50 variables based on the mean SHAP values.

Model Development

We developed four models using ML approaches: (1) logistic regression with least absolute shrinkage and selection operator (LASSO) regularization (LASSO regression), (2) random forest, (3) gradient-boosted decision tree, and (4) deep neural network. In these ML models, we used several methods to minimize potential overfitting in each model: (1) LASSO regularization, (2) out-of-bag estimation, (3) cross-validation, (4) dropout, (5) ridge regularization, and (6) batch normalization.

Finally, we compared our results with the Super Learner model, which is an algorithm that uses cross-validation to estimate the performance of multiple ML models and summarizes the prediction of those models using the ensemble method [33]. In addition, we trained a logistic regression model that used the same features as the ML models. The main analytic script can be found in Multimedia Appendix 1.

Conventional Logistic Regression (Severe Sepsis Mortality Prediction Model)

Logistic regression uses a function ranging between 0 and 1 to describe the probability that the outcome belongs to one of 2 particular categories. In contrast to linear regression, logistic regression does not require predicted variables to have a linear relationship with the outcome. Logistic regression is well suited for classification problems, such as problems involving describing the risk of developing a disease or the risk of mortality. In this study, we used the previously published Severe Sepsis Mortality Prediction Model as a reference for benchmarking.

Logistic Regression Model With LASSO Regularization

LASSO regularization is a model that shrinks regression coefficients toward 0, thereby effectively selecting important predictors and improving the interpretability of the model [34]. The coefficients of the LASSO regression are the values that minimize the residual sum of squares plus shrinkage penalty. The regularization was tuned by minimizing λ to minimize the mean squared error. We used 10-fold cross-validation to yield the optimal regularization parameter minimizing the sum of least squares plus shrinkage penalty using the R glmnet package (R Foundation for Statistical Computing).

Random Forest Models

Random forest is an ensemble of decision trees from bootstrapped training samples. Random forests modify the bagged tree procedure by only allowing a random number of the predictor variables to be considered at each split of each tree [26,35,36]. For this study, the Gini Impurity was used to determine the optimal variable and location of the split at each node in the tree. To optimize the AUC of the resulting tree, a cost complexity parameter, which penalizes larger trees, was used to control the size of the final tree. To improve the accuracy and stability of the decision tree model, a procedure called bagging was used to fit a bagged tree model [37]. This involved taking random bootstrap samples of patient data with replacement and fitting an unpruned tree model to each sample. The number of bagged trees in the final model was determined in the training data set using 10-fold cross-validation to maximize the training set AUC. We considered from 100 to 2000 trees and performed a pairwise statistical test to choose the best number of trees (Table 1). This results in trees that are less correlated with each other compared with bagged trees, thus potentially increasing accuracy. The optimal number of trees and predictor variables to be considered at each split was determined using 10-fold cross-validation, and the combination with the highest training set AUC was denoted as the final model. Table 2 shows the association between the number of variables allowed to be considered at each split in the random forest model with discrimination. The final random forest model was fitted with 400 trees with 50 variables at each split. We used the ranger package in R to construct the random forest models.

Table 1. Sensitivity analysis of tree numbers in the random forest algorithm.

Number of trees allowed	AUC[a] (95% CI)	Pairwise significant comparison of AUC	*P* value
100	0.876 (0.874-0.878)	100 trees versus 200 trees	<.001[b]
200	0.877 (0.876-0.879)	200 trees versus 300 trees	<.001[b]
300	0.878 (0.877-0.880)	300 trees versus 400 trees	<.001[b]
400	0.878 (0.877-0.880)	400 trees versus 500 trees	.30

[a]AUC: area under the curve.

[b]Values are significant at $P<.001$.

Table 2. Association between the number of variables allowed to be considered at each split in the random forest model and model discrimination.

Number of variables allowed	AUC[a] (95% CI)	Number of variables	Pairwise significant comparison of AUC (*P* value)
3	0.852 (0.850-0.854)	3 variables versus 5 variables	<.001[b]
5	0.860 (0.858-0.862)	5 variables versus 9 variables	<.001[b]
9	0.868 (0.866-0.869)	9 variables versus 15 variables	<.001[b]
15	0.874 (0.872-0.875)	15 variables versus 20 variables	<.001[b]
20	0.875 (0.874-0.877)	20 variables versus 25 variables	<.001[b]
25	0.877 (0.875-0.879)	25 variables versus 40 variables	<.001[b]
40	0.878 (0.876-0.880)	40 variables versus 50 variables	.02[c]
50	0.878 (0.877-0.880)	50 variables versus 70 variables	.53
70	0.878 (0.877-0.880)	N/A[d]	N/A

[a]AUC: area under the curve.

[b]Values are significant at $P<.001$.

[c]Values are significant at $P<.05$.

[d]N/A: not applicable.

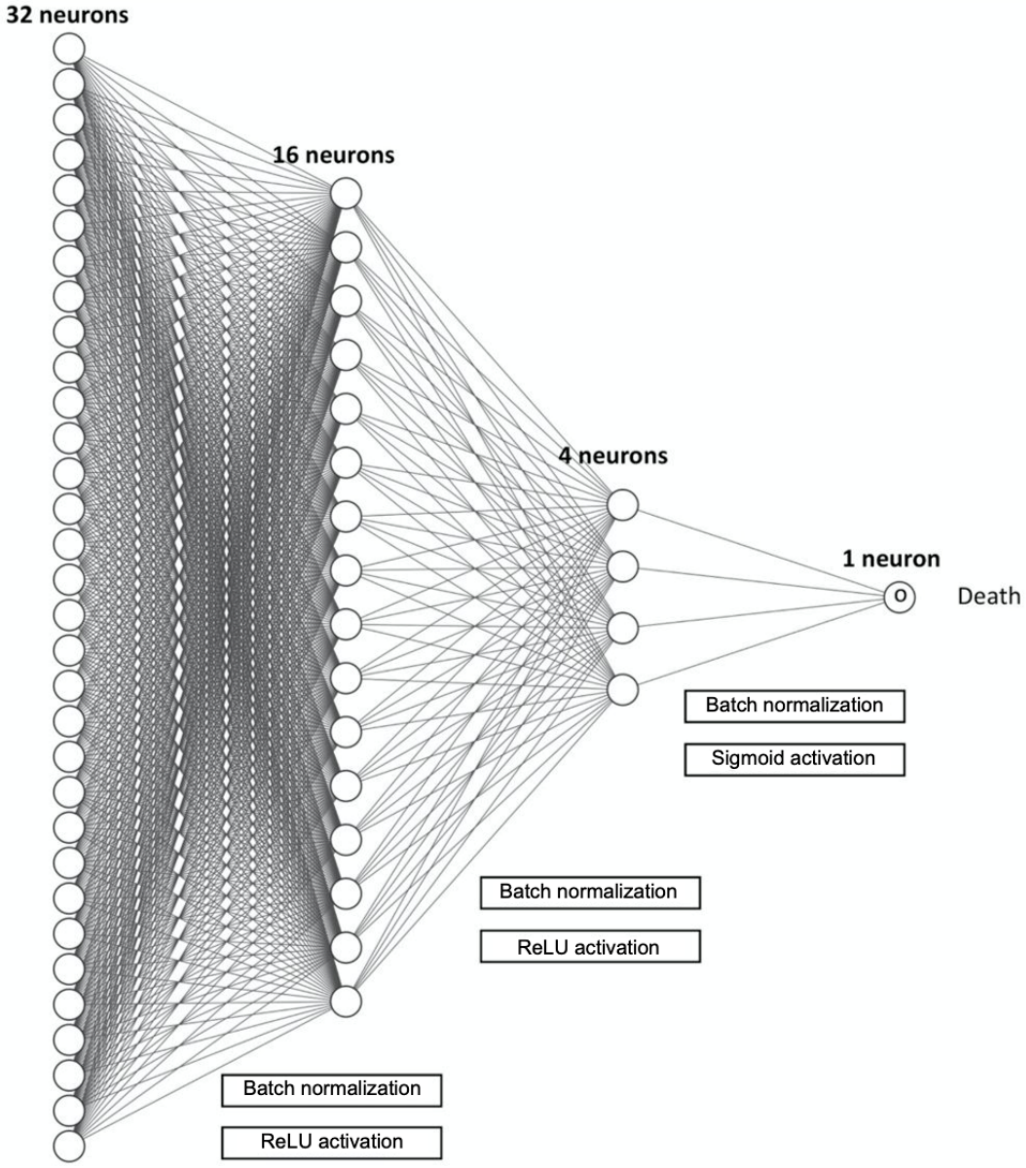
Xgboost

Gradient-boosted decision trees are also an ensemble method that constructs new tree models predicting the errors and residuals of previous models [38]. When adding the new models, this model uses a gradient descent algorithm to minimize the loss function. The final tree-based model fit was a gradient-boosted machine. This algorithm fits one tree at a time, first to all the outcomes in the training data and then to the residuals of the previous models, thus creating a combination of trees that increasingly weigh the *difficult to predict* events to a greater degree. The optimal number of splits for each individual tree, the total number of trees, and the learning rate were determined using 10-fold cross-validation in a similar method to that of the random forest model. In our final model, we had 10 splits for each tree in a total of 400 trees with a learning rate of 0.15. We stopped training if the validation AUC did not improve in 3 epochs. We used the xgboost package in R to construct the gradient-boosted decision tree models.

Deep Neural Networks Keras

Deep neural network models are composed of multiple processing layers. Neural networks are nonlinear models that involve creating a set of linear combinations of the original predictor variables and then using them as inputs into a hidden layer (or layers) of units, which then creates new combinations of these inputs to finally output the probability of the event of interest after a suitable transformation [39]. A feedforward multilayer perceptron neural network was used for this study. A penalty term, known as weight decay, and the number of hidden units in the model were determined using 10-fold cross-validation to maximize the training set AUC. We used a 4-layer feedforward model with an adaptive moment estimation optimizer, the binary cross-entropy loss function, and tuned hyperparameters using the R keras package. In the neural network model, continuous predictors are normalized using the mean and SDs. Binary variables encoding 0 and 1 are rescaled to encode −1 and 1. Finally, categorical variables use rescaled using effect encodings. The detailed architecture of the deep neural network in this study is shown in Figure 1.

Figure 1. The architecture of the 4-layered neural network to predict sepsis mortality. ReLu: Rectified Linear Unit.



Super Learner

Finally, we compared our results with the Super Learner algorithm, which uses cross-validation to estimate the performance of multiple ML models [33]. The Super Learner takes all weighted combinations from a set of candidate algorithms. After a set of algorithms is chosen, the meta-learning algorithm performs cross-validation to estimate the maximum likelihood of each selected algorithm on the data and selects the convex combination with the smallest squared prediction error on the test data set. In our case, we chose logistic regression as our meta-model. Overall, the generalization procedure learns the *n*-fold stratified predictions to maximize the likelihood function rather than minimize the mean squared error and to represent the meta-model in generating the best prediction. This approach has been proven to be as accurate as the best possible prediction algorithm. We used the 2 algorithms (random forest and xgboost) that we considered in this manuscript as candidate algorithms and compared the results with the models discussed in this paper. Our Super Learner scripts can be found in Multimedia Appendices 2-4.

Model Performance

In the test set (NIS 2014), we computed the prediction performance of each model that was derived above. First, we calculated the area under the receiver operating characteristic curve (AUROC) and confusion matrix results. The Delong test was used to compare the receiver operating characteristic curves between models. Second, the confusion matrix results were calculated. Third, given the imbalanced nature of our data set, we also calculated the area under the precision-recall curve (AUC-PR), recall, and precision of different ML models in predicting sepsis mortality. Fourth, calibration curves were constructed by plotting predicted probability versus actual probability from the ML models. The Brier scores of all the models considered were also calculated. The Brier score is a quadratic scoring rule where the squared differences between the actual binary outcomes and predicted probabilities are calculated. Therefore, lower values indicate better calibration. All analyses were performed using R (version 3.6.1).

Web Application

To increase the reproducibility and usability of this research on sepsis care and mortality, we generated a web-based application [40] for peer investigators to generate predictions of 30-day mortality for patients with sepsis and provide an introductory video (Multimedia Appendix 2). The web application is based on our Super Learner model and built using the Shiny package in R (version 4.0.5). The submission interface offers an example Microsoft Excel file with placeholder columns. Details of how to generate the variables are described in Multimedia Appendix 5.

Results

Baseline Characteristics

Figure 2 shows the flowchart of the cohort used in this study, and Table 3 provides descriptive statistics of survivors and nonsurvivors of sepsis from the cohort used. Table 4 shows the characteristics of patients with sepsis stratified by training and validation cohort.

Figure 2. Flowchart depicting the construction of the study cohort from the Nationwide Inpatient Sample (NIS) database. LASSO: least absolute shrinkage and selection operator.

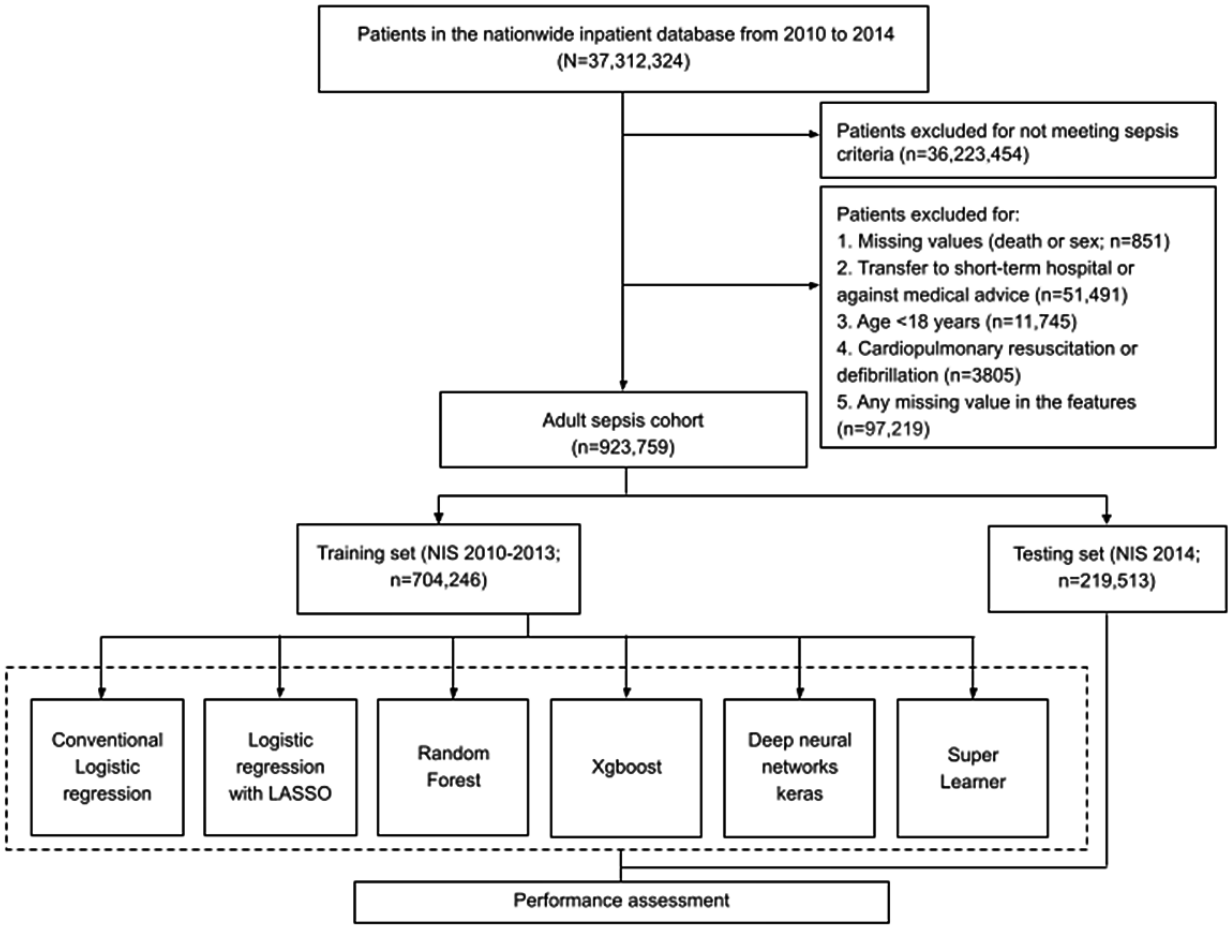


Table 3. Characteristics of patients with sepsis in the Nationwide Inpatient Sample stratified by in-hospital survival status (N=923,759).

Characteristics	Survivors of sepsis (n=726,918)	Nonsurvivors of sepsis (n=196,841)	Total
Age (years), mean (SE)	67.15 (16.44)	70.85 (14.88)	67.94 (16.19)
Women, n (%)	358,756 (49.4)	96,708 (49.1)	455,464 (49.3)
Race, n (%)			
White	511,579 (70.4)	137,807 (70)	649,386 (70.3)
Black	112,801 (15.5)	30,207 (15.3)	143,008 (15.5)
Hispanic	61,174 (8.4)	16,386 (8.3)	77,560 (8.4)
Others	41,364 (5.7)	12,441 (6.3)	53,805 (5.8)
Insurance, n (%)			
Medicare	221,228 (30.4)	60,933 (31)	282,161 (30.5)
Medicaid	185,758 (25.6)	48,838 (24.8)	234,596 (25.4)
Commercial	172,650 (23.8)	45,437 (23.1)	218,087 (23.6)
Other	147,282 (20.3)	41,633 (21.2)	188,915 (20.5)
Measures of acute illness severity, n (%)			
Early mechanical ventilation	118,939 (16.4)	76,773 (39)	195,712 (21.2)
Late mechanical ventilation	36,649 (5)	35,531 (18.1)	72,180 (7.8)
Shock	305,375 (42)	132,582 (67.4)	437,957 (47.4)
Hemodialysis	58,962 (8.1)	28,691 (14.6)	87,653 (9.5)
ICU[a] care (at least one day)	67,810 (9.3)	58,756 (29.8)	126,566 (13.7)
Underlying comorbidity, n (%)			
Anemia	265,364 (36.5)	55,632 (28.3)	320,996 (34.7)
Depression	81,827 (11.3)	14,612 (7.4)	96,439 (10.4)
Diabetes	256,947 (35.3)	57,294 (29.1)	314,241 (34)
Drug and substance abuse	25,311 (3.5)	4188 (2.1)	29,499 (3.2)
Chronic lung disease	188,546 (25.9)	50,749 (25.8)	239,295 (25.9)
Congestive heart failure	173,776 (23.9)	56,036 (28.5)	229,812 (24.9)
Hypertension	424,834 (58.4)	102,862 (52.3)	527,696 (57.1)
Hypothyroid disease	100,256 (13.8)	23,856 (12.1)	124,112 (13.4)
Liver disease	42,065 (5.8)	17,995 (9.1)	60,060 (6.5)
Renal failure, chronic	210,371 (28.9)	57,171 (29)	267,542 (29)
Lymphoma	13,691 (1.9)	5469 (2.8)	19,160 (2.1)
Metastatic carcinomas	30,789 (4.2)	17,109 (8.7)	47,898 (5.2)
Neurological conditions	117,134 (16.1)	27,791 (14.1)	144,925 (15.7)
Obesity	100,716 (13.9)	18,173 (9.2)	118,889 (12.9)
Malignant solid tumors	27,426 (3.8)	10,057 (5.1)	37,483 (4.1)
Rheumatoid arthritis or collagen vascular diseases	27,294 (3.8)	6324 (3.2)	33,618 (3.6)
Paraplegia	53,755 (7.4)	10,955 (5.6)	64,710 (7)
Perivascular conditions	68,641 (9.4)	22,853 (11.6)	91,494 (9.9)
Psychiatric diseases	44,282 (6.1)	6902 (3.5)	51,184 (5.5)
Pulmonary-circulatory	43,697 (6)	15,327 (7.8)	59,024 (6.4)
Weight loss	146,865 (20.2)	47,320 (24)	194,185 (21)
System dysfunction, n (%)			
Renal dysfunction	433,920 (59.7)	129,768 (65.9)	563,688 (61)

Characteristics	Survivors of sepsis (n=726,918)	Nonsurvivors of sepsis (n=196,841)	Total
Cardiovascular dysfunction or shock	281,647 (38.7)	132,079 (67.1)	413,726 (44.8)
Acute respiratory failure	161,921 (22.3)	116,406 (59.1)	278,327 (30.1)
CNS[b] dysfunction	162,716 (22.4)	51,146 (26)	213,862 (23.2)
Hepatic dysfunction	18,579 (2.6)	20,561 (10.4)	39,140 (4.2)
Lifestyle factors, n (%)			
Smoking	75,404 (10.4)	15,033 (7.6)	90,437 (9.8)
Alcoholism	32,879 (4.5)	10,674 (5.4)	43,553 (4.7)

[a]ICU: intensive care unit.

[b]CNS: central nervous system.

Table 4. Characteristics of patients with sepsis in the Nationwide Inpatient Sample stratified by training and validation cohort (N=923,759).

Characteristic	Training (2010-2013)		Testing (2014)	
	Survivors of sepsis (n=548,930)	Nonsurvivors of sepsis (n=155,316)	Survivors of sepsis (n=177,988)	Nonsurvivors of sepsis (n=41,525)
Age (years), mean (SE)	67.25 (16.46)	70.96 (14.93)	66.84 (16.37)	70.44 (14.68)
Women, n (%)	271,311 (49.4)	76,496 (49.3)	87,445 (49.1)	20,212 (48.7)
Race, n (%)				
White	385,330 (70.2)	108,405 (69.8)	126,249 (70.9)	29,402 (70.8)
Black	86,727 (15.8)	24,295 (15.6)	26,074 (14.6)	5912 (14.2)
Hispanic	45,887 (8.4)	12,954 (8.3)	15,287 (8.6)	3432 (8.3)
Others	30,986 (5.6)	9662 (6.2)	10,378 (5.8)	2779 (6.7)
Insurance, n (%)				
Medicare	166,023 (30.2)	47,814 (30.8)	55,205 (31)	13,119 (31.6)
Medicaid	136,607 (24.9)	37,627 (24.2)	49,151 (27.6)	11,211 (27)
Commercial	132,428 (24.1)	36,387 (23.4)	40,222 (22.6)	9050 (21.8)
Other	113,872 (20.7)	33,488 (21.6)	33,410 (18.8)	8145 (19.6)
Measures of acute illness severity, n (%)				
Early mechanical ventilation	92,718 (16.9)	60,822 (39.2)	26,221 (14.7)	15,951 (38.4)
Late mechanical ventilation	28,892 (5.3)	28,532 (18.4)	7757 (4.4)	6999 (16.9)
Shock	232,963 (42.4)	103,544 (66.7)	72,412 (40.7)	29,038 (69.9)
Hemodialysis	46,180 (8.4)	22,818 (14.7)	12,782 (7.2)	5873 (14.1)
ICU[a] care (at least one day)	53,146 (9.7)	46,914 (30.2)	14,664 (8.2)	11,842 (28.5)
Underlying comorbidity, n (%)				
Anemia	201,132 (36.6)	43,380 (27.9)	64,232 (36.1)	12,252 (29.5)
Depression	59,998 (10.9)	11,239 (7.2)	21,829 (12.3)	3373 (8.1)
Diabetes	191,296 (34.8)	44,598 (28.7)	65,651 (36.9)	12,696 (30.6)
Drug and substance abuse	17,689 (3.2)	3113 (2)	7622 (4.3)	1075 (2.6)
Chronic lung disease	140,276 (25.6)	39,550 (25.5)	48,270 (27.1)	11,199 (27)
Congestive heart failure	130,913 (23.8)	43,716 (28.1)	42,863 (24.1)	12,320 (29.7)
Hypertension	316,301 (57.6)	79,939 (51.5)	108,533 (61)	22,923 (55.2)
Hypothyroid disease	73,904 (13.5)	18,348 (11.8)	26,352 (14.8)	5508 (13.3)
Liver disease	30,753 (5.6)	13,796 (8.9)	11,312 (6.4)	4199 (10.1)
Renal failure, chronic	158,078 (28.8)	44,704 (28.8)	52,293 (29.4)	12,467 (30)
Lymphoma	10,371 (1.9)	4281 (2.8)	3320 (1.9)	1188 (2.9)
Metastatic carcinomas	23,087 (4.2)	13,352 (8.6)	7702 (4.3)	3757 (9)
Neurological conditions	87,994 (16)	21,699 (14)	29,140 (16.4)	6092 (14.7)
Obesity	71,693 (13.1)	13,392 (8.6)	29,023 (16.3)	4781 (11.5)
Malignant solid tumors	20,417 (3.7)	7814 (5)	7009 (3.9)	2243 (5.4)
Rheumatoid arthritis or collagen vascular diseases	20,368 (3.7)	4898 (3.2)	6926 (3.9)	1426 (3.4)
Paraplegia	40,811 (7.4)	8488 (5.5)	12,944 (7.3)	2467 (5.9)
Perivascular conditions	50,853 (9.3)	17,734 (11.4)	17,788 (10)	5119 (12.3)
Psychiatric diseases	32,698 (6)	5320 (3.4)	11,584 (6.5)	1582 (3.8)
Pulmonary-circulatory	32,214 (5.9)	11,625 (7.5)	11,483 (6.5)	3702 (8.9)
Weight loss	113,028 (20.6)	37,182 (23.9)	33,837 (19)	10,138 (24.4)

Characteristic	Training (2010-2013)		Testing (2014)	
	Survivors of sepsis (n=548,930)	Nonsurvivors of sepsis (n=155,316)	Survivors of sepsis (n=177,988)	Nonsurvivors of sepsis (n=41,525)
System dysfunction, n (%)				
Renal dysfunction	324,840 (59.2)	101,420 (65.3)	109,080 (61.3)	28,348 (68.3)
Cardiovascular dysfunction or shock	215,545 (39.3)	103,064 (66.4)	66,102 (37.1)	29,015 (69.9)
Acute respiratory failure	125,706 (22.9)	92,008 (59.2)	36,215 (20.3)	24,398 (58.8)
CNS[b] dysfunction	118,837 (21.6)	38,642 (24.9)	43,879 (24.7)	12,504 (30.1)
Hepatic dysfunction	14,091 (2.6)	15,752 (10.1)	4488 (2.5)	4809 (11.6)
Lifestyle factors, n (%)				
Smoking	54,038 (9.8)	11,205 (7.2)	21,366 (12)	3828 (9.2)
Alcoholism	24,025 (4.4)	8083 (5.2)	8854 (5)	2591 (6.2)

[a]ICU: intensive care unit.

[b]CNS: central nervous system.

Performance Comparison

Compared with the reference logistic regression model (0.786, 95% CI 0.783-0.788), all 4 ML methods showed superior discriminative ability (P<.001; Table 5). Of all 4 ML methods, the deep neural network showed the highest (P<.001) discriminative ability (0.893, 95% CI 0.891-0.895) followed by the gradient-boosting model (0.888, 95% CI 0.886-0.889). The AUC of the deep neural network (0.893, 95% CI 0.891-0.895) was higher than that of the Super Learner model (0.883, 95% CI 0.881-0.885). Both LASSO (0.878, 95% CI 0.876-0.879) and random forest (0.878, 95% CI 0.877-0.880) had an AUC that was slightly lower among the ML models but was nevertheless superior (P<.001) to the reference logistic model (Figures 3 and 4).

Of the ML models, the deep neural network also demonstrated higher specificity (0.794, 95% CI 0.793-0.796) and positive predictive value (0.484, 95% CI 0.480-0.488) while resulting in lower sensitivity (0.826, 95% CI 0.823-0.830) and negative predictive value (0.951, 95% CI 0.950-0.953) compared with the xgboost model, but these differences were statistically insignificant. The Super Learner showed similar results to our xgboost model, with statistically lower specificity (0.769, 95% CI 0.768-0.771) and positive predictive value (0.458, 95% CI 0.455-0.460) compared with the neural network model. However, the neural network model showed only marginally lower sensitivity (0.826, 95% CI 0.823-0.830) and negative predictive value (0.951, 95% CI 0.950-0.953) compared with the Super Learner.

The AUC-PR, recall, and precision of different ML models in predicting sepsis mortality are shown in Figure 4 and Table 6. The ML models showed superior AUC-PR measures (0.636-0.681) compared with the reference logistic regression model (0.442). In addition, being paralleled with our finding from the AUROC, the deep neural network model showed the highest AUC-PR (0.681) followed by the xgboost model (0.673).

Most of the models showed great calibration from a visual representation, which shows calibration plots characterized by visual inspection and reporting of the intercept and slope (Figure 5). The intercept's deviation from 0 indicates the extent to which predictions are underpredicting or overpredicting the probability of the event of interest—sepsis mortality. All of our models showed small departures (intercept <0.1) except for the random forest model, which overpredicted sepsis mortality (0.245). The random forest and neural network models slightly overpredicted sepsis mortality, whereas the reference logistic regression, LASSO, and xgboost models slightly underpredicted sepsis mortality. Compared with the reference logistic regression model, which had a slope of 1.048, LASSO (1.044), xgboost (1.087), and the neural network model (1.096) had similar slopes that were all close to 1. However, the random forest model showed the largest deviation from perfect calibration (1.458).

In addition, Table 7 shows the Brier scores of all the models. The deep neural network model exhibited the lowest Brier score of 0.954 followed by xgboost (0.102), which is in alignment with their high discriminatory ability. The ML models exhibited a good range of Brier scores (0.095-0.108), all of which were higher than those of the reference logistic regression model (0.129).

Table 5. Measures of model discrimination and accuracy in the validation data set (Nationwide Inpatient Sample 2014), including area under the curve (AUC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

Model	AUC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Reference logistic regression (Severe Sepsis Prediction score)	0.786 (0.783-0.788)	0.708 (0.704-0.713)	0.722 (0.720-0.774)	0.373 (0.370-0.376)	0.914 (0.912-0.915)
LASSO[a]	0.878 (0.876-0.879)	0.812 (0.808-0.816)	0.784 (0.782-0.786)	0.468 (0.464-0.471)	0.947 (0.946-0.948)
Random forest	0.878 (0.877-0.880)	0.818 (0.814-0.821)	0.771 (0.769-0.773)	0.454 (0.451-0.458)	0.948 (0.947-0.949)
Xgboost	0.888 (0.886-0.889)	0.829 (0.826-0.833)	0.781 (0.781-0.785)	0.472 (0.468-0.475)	0.952 (0.950-0.953)
Deep neural network	0.893 (0.891-0.895)	0.826 (0.823-0.830)	0.794 (0.793-0.796)	0.484 (0.480-0.488)	0.951 (0.950-0.953)
Super Learner	0.883 (0.881-0.885)	0.833 (0.829-0.837)	0.769 (0.768-0.771)	0.458 (0.455-0.460)	0.952 (0.951-0.953)

[a]LASSO: least absolute shrinkage and selection operator.

Figure 3. Receiver operating characteristic curves of different machine learning models in predicting sepsis mortality. AUC: area under the curve; LASSO: least absolute shrinkage and selection operator.

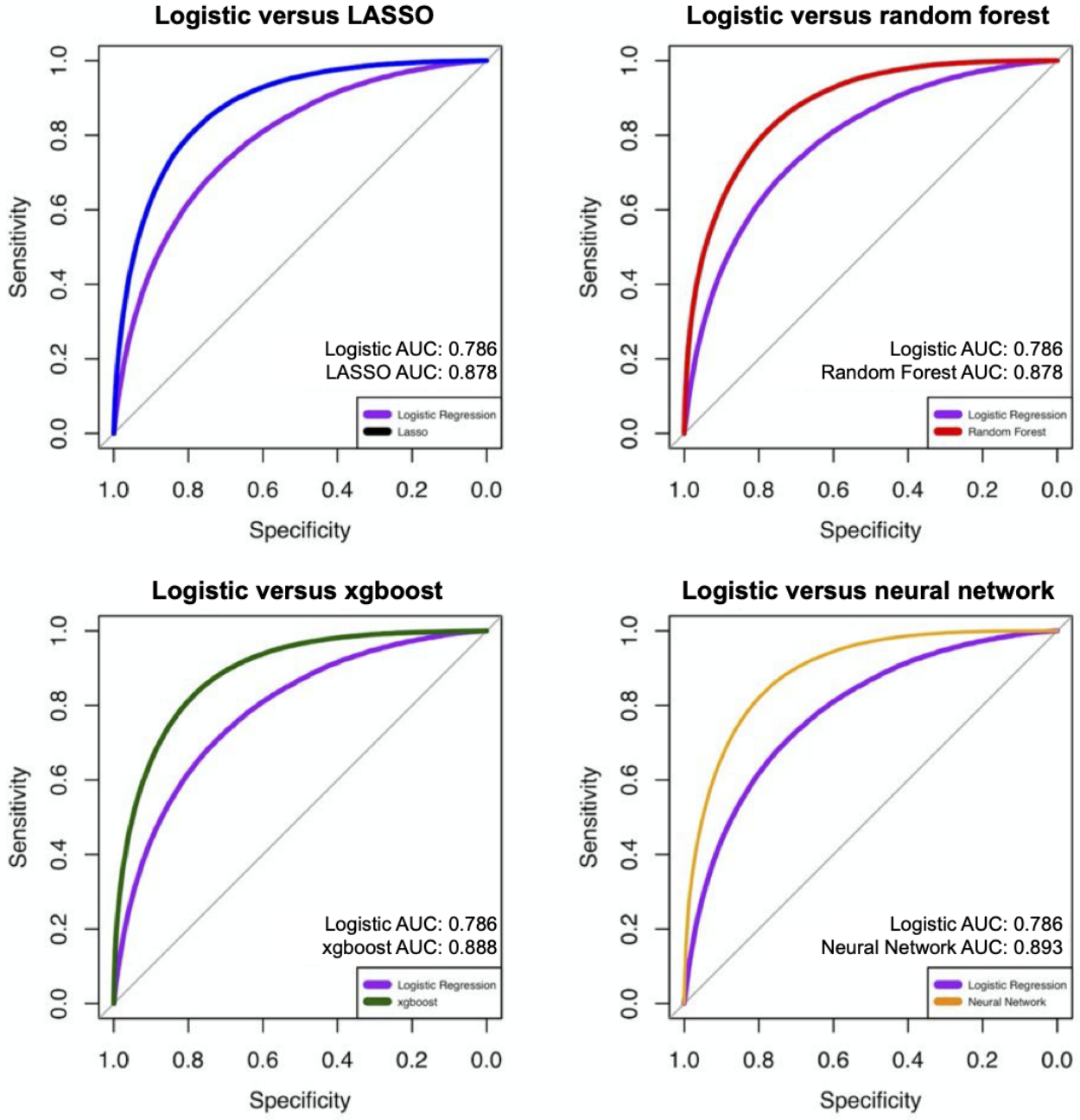


Figure 4. Precision-recall curves of different machine learning models in predicting sepsis mortality. AUC: area under the curve; LASSO: least absolute shrinkage and selection operator.

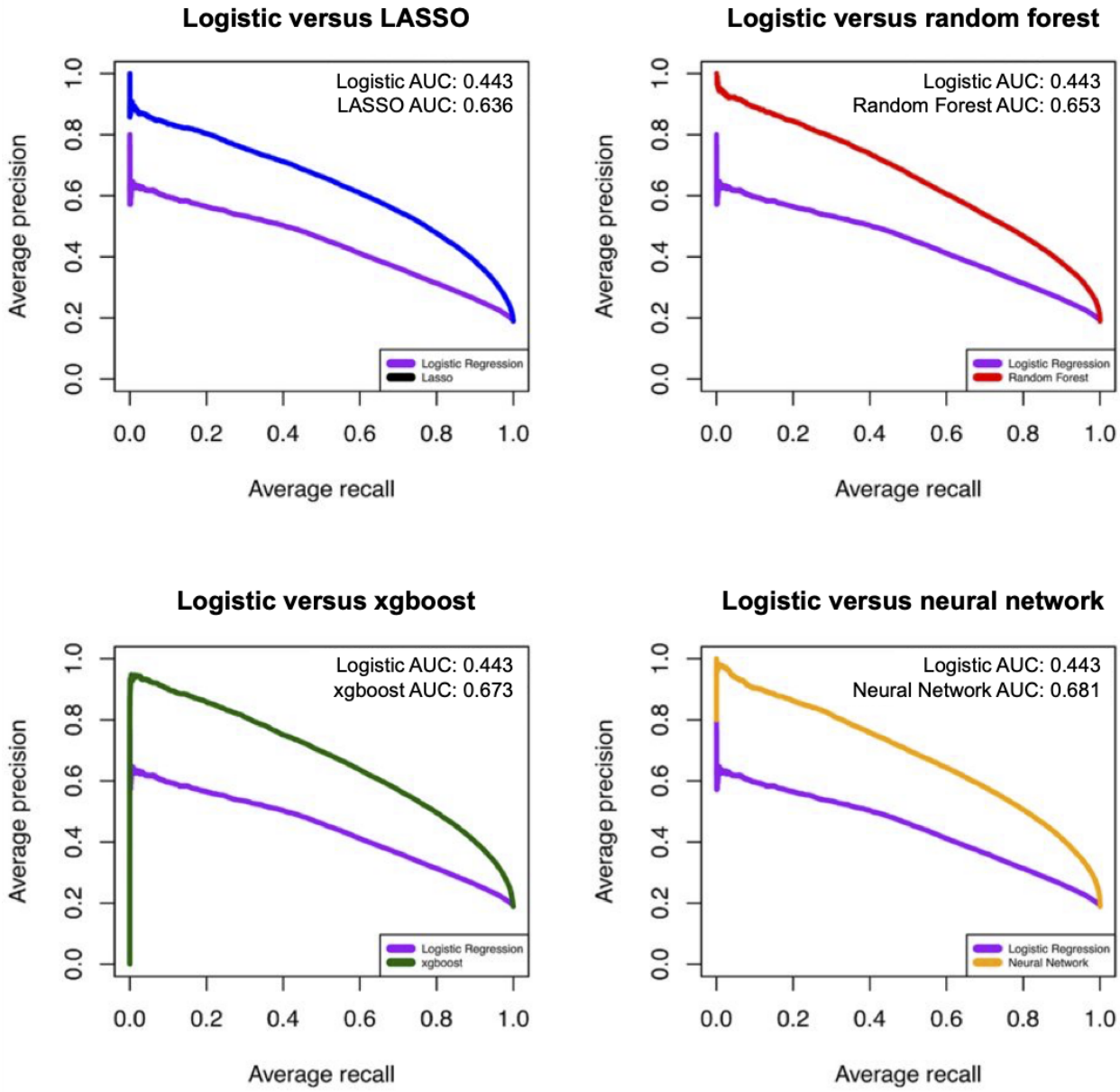


Table 6. The area under the precision-recall curve (AUC-PR), recall, and precision of different machine learning models in predicting sepsis mortality.

	AUC-PR, mean (SD)	Recall (95% CI)	Precision (95% CI)
Reference logistic regression	0.443 (0.003)	0.587 (0.583-0.591)	0.403 (0.401-0.405)
LASSO[a]	0.636 (0.001)	0.806 (0.805-0.807)	0.410 (0.410-0.411)
Random forest	0.653 (0.002)	0.806 (0.805-0.807)	0.415 (0.414-0.416)
Xgboost	0.673 (0.002)	0.814 (0.813-0.816)	0.420 (0.420-0.421)
Neural networks	0.681 (0.002)	0.815 (0.814-0.816)	0.427 (0.426-0.428)

[a]LASSO: least absolute shrinkage and selection operator.

Figure 5. Calibration plots of observed versus predicted hospital mortality and associated mortality ratios by risk deciles in the development and validation cohorts. LASSO: least absolute shrinkage and selection operator.

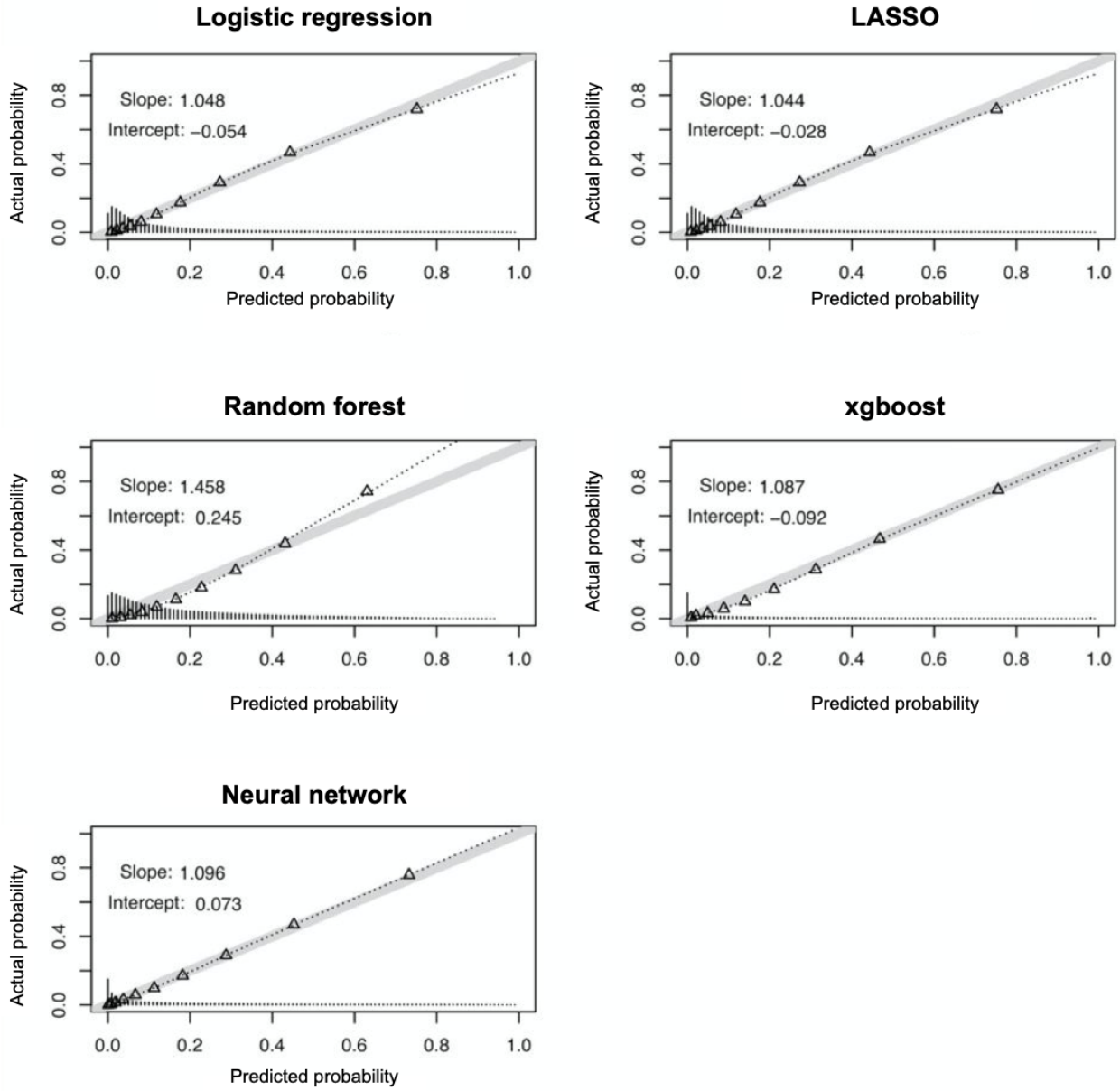


Table 7. Calibration measures of different machine learning models in predicting sepsis mortality.

	Brier score	Slope	Intercept
Reference logistic regression	0.129	1.048	−0.054
LASSO[a]	0.108	1.044	−0.028
Random forest	0.103	1.458	0.245
Xgboost	0.102	1.087	−0.092
Neural networks	0.0954	1.096	0.073

[a]LASSO: least absolute shrinkage and selection operator.

Variable Importance of the Random Forest by Gini Impurity and Xgboost Model by SHAP

The top 50 variables according to the variable importance of the random forest algorithm by the Gini Impurity are shown in Figure 6. The top 50 features with the highest mean SHAP values of the xgboost algorithm are shown in Figure 7. SHAP is a popular technique used to explain model predictions. SHAP is model-agnostic, with the ability to explain any given classifier. Lundberg and Lee [32] proposed SHAP as a united approach to explaining the output of any ML model. Acute respiratory failure and age were the 2 most important features from the random forest model as well as the xgboost model, and acute

respiratory failure was not a feature in the reference logistic regression model. In addition, we found many diagnosis (primary, secondary, and other) and procedure (primary and secondary) variables to be important predictors for sepsis mortality, which were not included in the reference logistic regression model. To assess collinearity, variance inflation factors of the final feature panel from SHAP were calculated from a total cohort combining both the training and validation cohorts. All 50 features showed variance inflation factor scores <5 except for early mechanical ventilation and late mechanical ventilation (Figure 8).

Figure 6. Variables of importance from random forest ranked by impurity-based variable importance. CNS: central nervous system; ICU: intensive care unit.

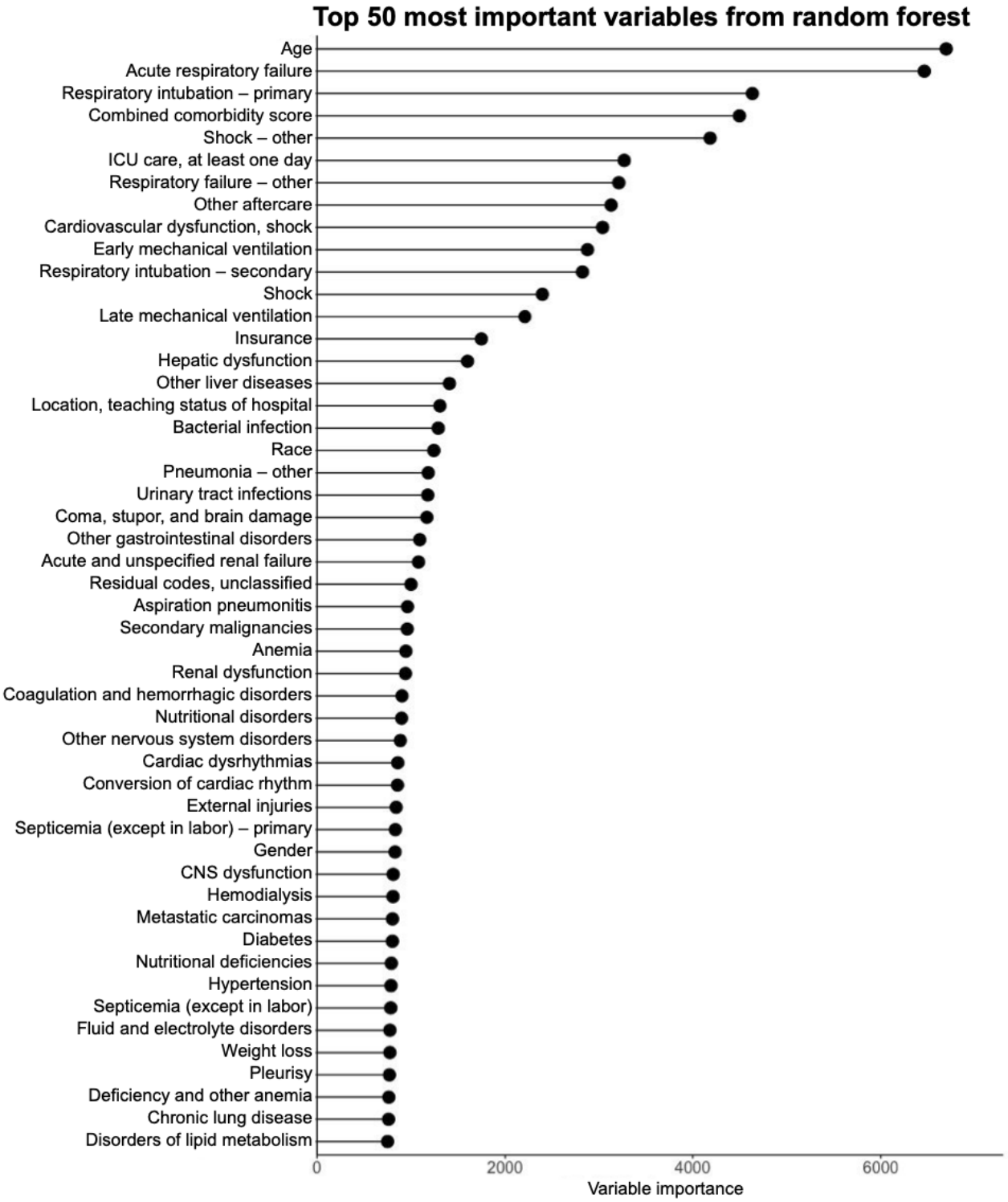


Figure 7. Variables of importance from xgboost ranked by mean Shapley Additive Explanations values (SHAP). ICU: intensive care unit.

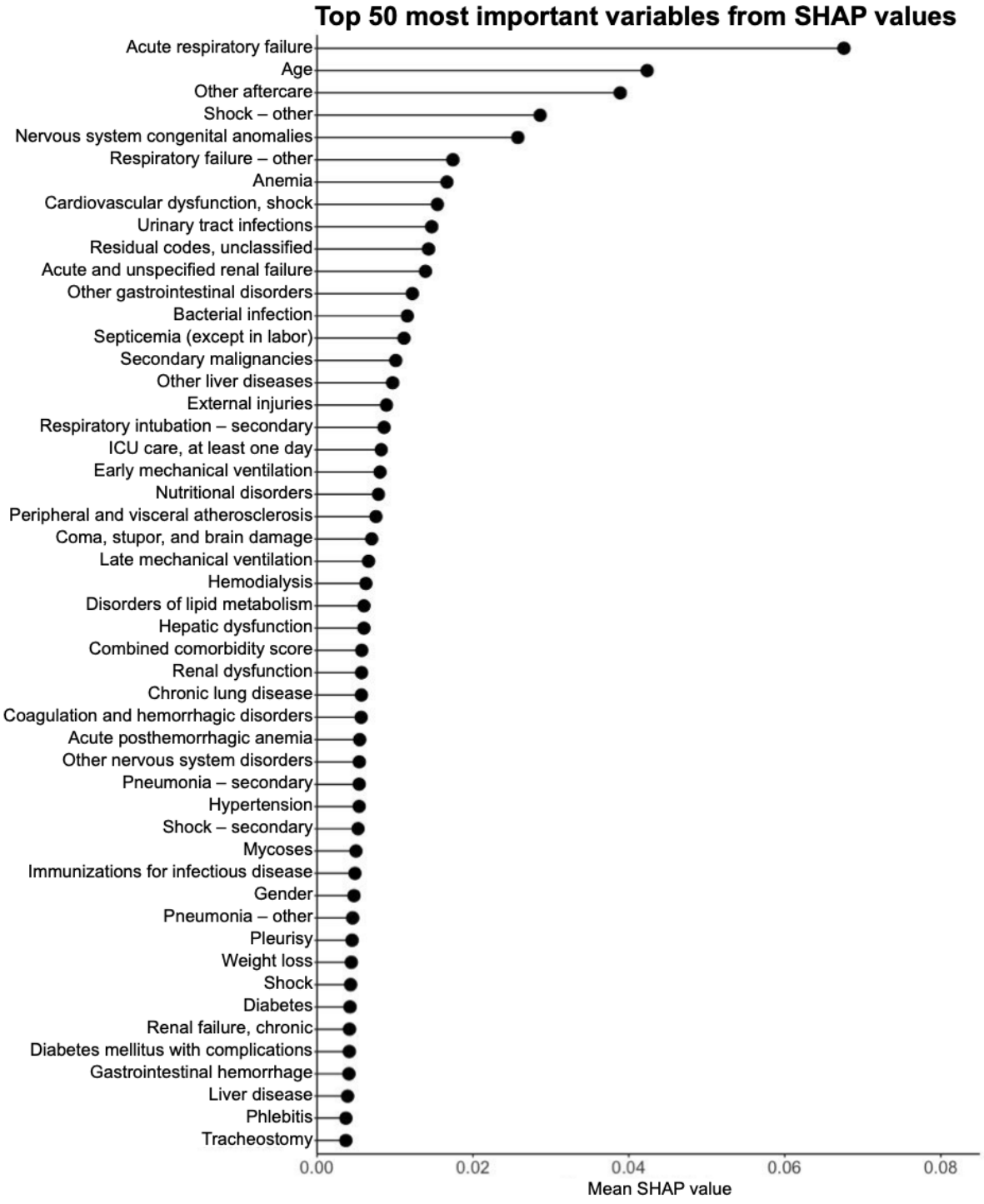
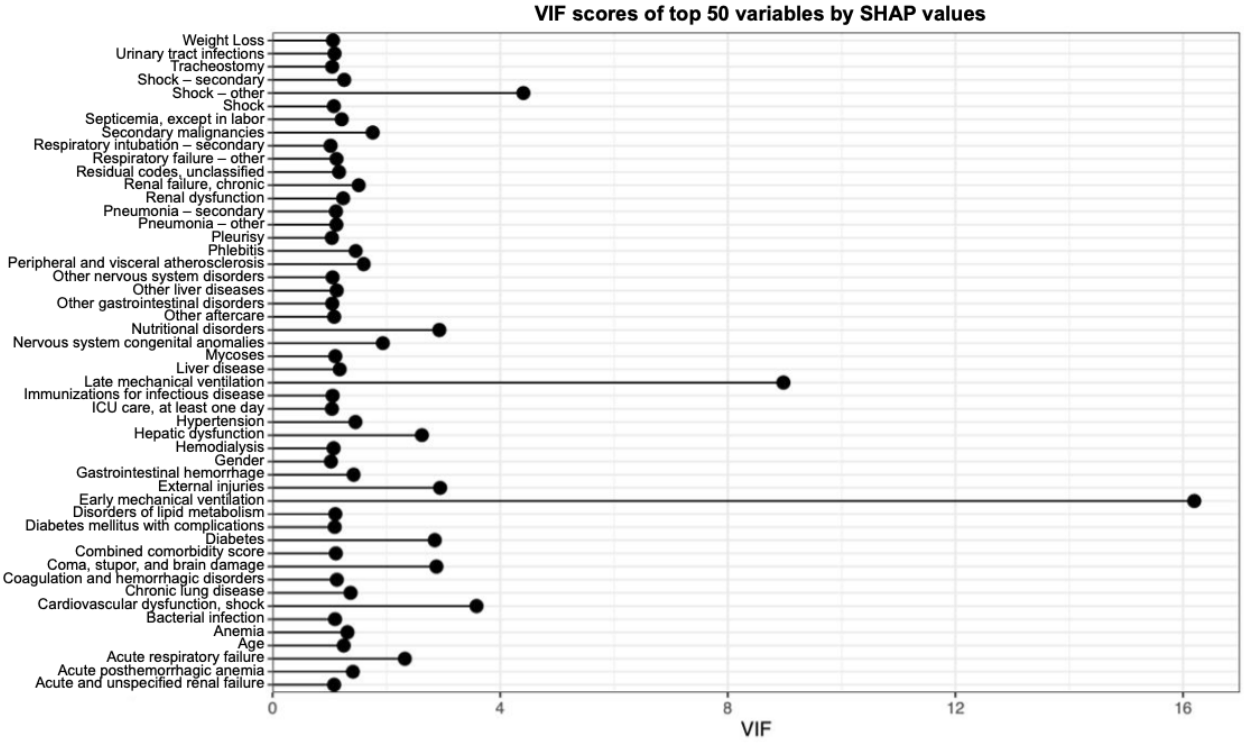


Figure 8. Variance inflation factor scores of top 50 variables by Shapley Additive Explanations (SHAP) values. ICU: intensive care unit; VIF: variance inflation factor.



Logistic Regression Model Using the Same Features as the ML Models

Moreover, Table 8 shows the performance of a logistic regression model that used the same features as the ML models. Overall, this logistic regression model with all features showed a very comparable calibration performance assessed by the Brier score (0.102) compared with those of the ML models (95% CI 0.0954-0.108). This model resulted in a slightly lower AUC-PR (0.634) compared with the ML models (95% CI 0.636-0.681) and a statistically lower (P<.001) AUROC of 0.857 (95% CI 0.855-0.859) compared with the ML models (95% CI 0.876-0.895).

Table 8. Performance comparison of the machine learning models with the logistic regression model with the same features.

	Brier score	AUC-PR[a], mean (SD)	AUC[b] (95% CI)	AUC *P* value
Logistic regression model—all features	0.102	0.634 (0.003)	0.857 (0.855-0.859)	N/A[c]
LASSO[d]	0.108	0.636 (0.001)	0.878 (0.876-0.879)	<.001[e]
Random forest	0.103	0.653 (0.002)	0.878 (0.877-0.880)	<.001[e]
Xgboost	0.102	0.673 (0.002)	0.888 (0.886-0.889)	<.001[e]
Neural networks	0.0954	0.681 (0.002)	0.893 (0.891-0.895)	<.001[e]

[a]AUPRC: area under the precision-recall curve.

[b]AUC: area under the curve.

[c]N/A: not applicable.

[d]LASSO: least absolute shrinkage and selection operator.

[e]Values are significant at P<.001.

Later, we trained a reference logistic regression model and a random forest model using this experiment and compared their performance to that of our original study design. I Instead of splitting our cohort by year, we split the training and testing cohorts randomly to see if our findings hold true. Of the patients from the entire data set, 75% (692,819) were assigned to the training set, and the remaining 25% (230,940) of the samples were assigned to the testing set. Using this approach, we obtained AUROC of 0.765 (95% CI 0.763-0.768) from the reference logistic regression model, whereas we observed superior performance from a random forest model with AUROC of 0.855 (95% CI 0.853-0.857). This finding is consistently with the results with our original approach of splitting our cohort by year where ML models showed superior discrimination performance compared with the reference logistic model.

Finally, we used the Gini Impurity to calculate the variable of importance in our random forest model. In Table 9, we present

the results of an analysis of the top 50 most important predictive features when a different train–test split method is used. When our cohort was split randomly, Table 9 shows the top 50 most important features from a random forest model. The third column shows whether these features were also in the top 50 in the previous random forest model using the train–test split-by-year approach (Figure 4). Of the 50 most important features, 44 (88%) were also top features identified by the previous random forest model using the train–test split-by-year approach (Figure 4). Although 6 features have changed, we note that 5 (83%) are low-ranking features with higher variability. As a result, despite having used 2 different train–test split approaches, the features identified and ranked in the top 50 most important features by both models had relatively consistent ranks (ie, the most important features were age and acute respiratory failure).

Table 9. Variables of importance from the random forest (random train–test split cohort).

Variable name	Importance rank	Top 50 from previous cohort
Acute respiratory failure	1	Yes
Age	2	Yes
Respiratory intubation and mechanical ventilation (primary procedure)	3	Yes
Combined comorbidity score	4	Yes
Shock (other diagnosis)	5	Yes
ICU[a] care (at least one day)	6	Yes
Cardiovascular dysfunction or shock	7	Yes
Other aftercare (other diagnosis)	8	Yes
Early mechanical ventilation	9	Yes
Respiratory intubation and mechanical ventilation (secondary procedure)	10	Yes
Shock	11	Yes
Late mechanical ventilation	12	Yes
Insurance	13	Yes
Hepatic dysfunction	14	Yes
Other liver diseases (other diagnosis)	15	Yes
Coma, stupor, and brain damage (other diagnosis)	16	Yes
Location or teaching status of hospital	*17*	*No*
Bacterial infection, unspecified site (other diagnosis)	18	Yes
Race	19	Yes
Urinary tract infections (other diagnosis)	20	Yes
Pneumonia (except that caused by tuberculosis or sexually transmitted disease; other diagnosis)	21	Yes
Other gastrointestinal disorders (other diagnosis)	22	Yes
Joint disorders and dislocations, trauma-related (secondary diagnosis)	23	Yes
Acute and unspecified renal failure (other diagnosis)	24	Yes
Residual codes, unclassified (other diagnosis)	25	Yes
Aspiration pneumonitis and food or vomitus (other diagnosis)	26	Yes
Secondary malignancies (other diagnosis)	27	Yes
Anemia	28	Yes
Renal dysfunction	29	Yes
Other nervous system disorders (other diagnosis)	30	Yes
Other nutritional, endocrine, and metabolic disorders (other diagnosis)	31	Yes
Coagulation and hemorrhagic disorders (other diagnosis)	32	Yes
Other injuries and conditions because of external causes (other diagnosis)	33	Yes
Cardiac dysrhythmias (other diagnosis)	34	Yes
Insertion, replacement, or removal of extracranial ventricular shunt (primary procedure)	35	Yes
CNS[b] dysfunction	36	Yes
Sex	37	Yes
Hemodialysis	38	Yes
Diabetes	39	Yes
Septicemia (except in labor; other diagnosis)	40	Yes
Nutritional deficiencies (other diagnosis)	41	Yes

Variable name	Importance rank	Top 50 from previous cohort
Hypertension	42	Yes
Administrative or social admission (other diagnosis)	43	Yes
Allergic reactions (other diagnosis)	44	No
Pleurisy, pneumothorax, and pulmonary collapse (other diagnosis)	45	No
Metastatic cancer	46	Yes
Weight loss	47	Yes
Deficiency and other anemia (other diagnosis)	48	No
Delirium, dementia, and amnestic and other cognitive disorders (other diagnosis)	49	No
Coronary atherosclerosis and other heart disease (other)	50	No

[a]ICU: intensive care unit.

[b]CNS: central nervous system.

Discussion

Principal Findings

In this study, we applied 5 ML algorithms (LASSO, random forest, xgboost, deep neural network, and Super Learner) using variables from a national administrative database to predict in-hospital mortality in a sepsis cohort identified using the previously validated Martin implementation. The AUROCs of the ML models were in the excellent range (95% CI 0.877-0.895), supporting our ML models' superior ability to discriminate mortality of patients with sepsis compared with the reference logistic regression model (95% CI 0.783-0.788). The ML models also showed superior AUC-PR measures (95% CI 0.636-0.681) compared with the reference logistic regression model (0.442). Among them, the models based on deep neural networks and xgboost outperformed the others in predicting sepsis mortality. To our knowledge, this is the first study to apply advanced ML models to predict sepsis mortality based on an administrative database.

It is important to distinguish between the 2 complementary uses of sepsis mortality risk prediction as they are distinct in their design and overall goals. EMR-integrated sepsis mortality prediction models are designed for use at the point of care to risk-stratify patients for clinical decision-making in the intensive care unit or emergency department [9-19]. However, unless proprietary systems are purchased, legal, technical, and financial barriers make it nearly impossible to extract the necessary clinical data from different EMR systems to assess performance across hospitals and states. In contrast, sepsis mortality prediction models based on administrative claims databases, which are available nationally, are designed to compare expected and actual sepsis mortality [20-24]. The latter was the focus of this study.

To date, traditional regression analyses have been applied to administrative data sources. Logistic regression of data ranging from single-center databases to regional and national databases has been used to predict sepsis mortality based on administrative data. Lagu et al [20] achieved an AUC of 0.78, Ford et al [21] achieved an AUC of 0.838, König et al [22] achieved an AUC >0.8, Schwarzkopf et al [24] achieved an AUC of 0.74, and Rhee et al [23] achieved an AUC of 0.776. In contrast to our approach, the aforementioned studies largely used traditional statistical models and did not use a validated Martin or Angus implementations approach to identify patients with sepsis. Moreover, they did not use the national inpatient database of the United States, which is the largest data set of US hospitalized patients. Recently, ML models have been applied to predicting sepsis mortality. Although some of them showed excellent performance, most of them were designed for point-of-care clinical application using the local EMR. In 2 previous studies using support vector machines, Ribas et al [12] achieved an AUC of 0.80, and Tsoukalas et al [13] obtained an AUC of 0.61. Taylor et al [14] used 500 clinical variables with a random forest model, which resulted in an AUC of 0.86. The study by Perng et al [15] used a support vector machine, k-nearest neighbor, random forest, and softmax with different extraction methods and achieved an AUC of 0.94. Kwon and Baek [16] used gradient boosting and random forests, achieving an AUC of 0.86. These ML algorithms were based on the local EMR database and may not be generalizable to other hospitals because of the case mix. By contrast, our ML models were based on a national administrative database with maximal generalizability [41].

As sepsis represents a major driver of cost and health care burden in the United States [4], improvement in sepsis care quality has been an important challenge. Considering the heterogeneous nature of sepsis, the calculation of sepsis risk-standardized mortality rates (RSMRs) is of great importance in measuring sepsis care quality across hospitals. Few relevant studies have been conducted based on a nationwide administrative database [23], and there remains much room for improved accuracy. Although hospital 30-day RSMRs for acute myocardial infarction, heart failure, and pneumonia have been reported by the Centers for Medicare and Medicaid Services [42], RSMRs for sepsis have not been well-characterized. Calculation of RSMR is important as identification of gaps between a facility's RSMR and those of the state or nation's highest-performing hospitals can lead hospital administrators, government policy makers, and other stakeholders to identify differences in practice and take action to improve sepsis care quality [43,44]. Disclosing discrepancies in RSMR also serves to reduce the asymmetry of information between consumers and health care providers and may spur market forces toward

a more efficient allocation and distribution of health care resources to improve care [45]. To calculate RSMR across hospitals, our ML models were the first step in developing accurate models.

We believe that our mortality prediction model is an important tool that can be applied in health care research, quality improvement, and health policy making. However, our results should be interpreted with several limitations. First, any variation in the quality of coding in administrative data might affect the reliability of our study, including payment-related incentives for coding, over- or undercoding of conditions or risk factors, inconsistencies in coding practices between hospitals, and new technologies applied in sepsis care [46-48]. Second, the Martin implementation with which we extracted the sepsis cohort has been criticized for the less stringent use of septicemia and the omission of immunologic and coagulopathic organ dysfunction [30]. Third, we used in-hospital mortality as an outcome in our study and excluded patients who were transferred between hospitals. Consequently, those transferred against medical advice or to short-term hospitals were not counted. Whether this focus on in-hospital mortality could be biased by the hospital discharge policy warrants further investigation [49]. Fourth, some sepsis-related local characteristics such as local disease prevalence cannot be captured in a nationwide claim-based database. Thus, these variables could not be modeled and might influence our comparison results. Fifth, despite the excellent performance of the ML models, they suffer from varying degrees of explainability issues, and the inferences about variables (especially those that are clinically modifiable) tend to be more challenging [50]. Sixth, our model cannot be continuously updated because of the recent policy change of the NIS to eliminate state and hospital identifiers. As there is a time lag of >6 years, further research is needed to refine and update our ML models. However, the results of our training and validating analyses suggest that the accuracy of our model may not be significantly affected by time. Seventh, despite strong discrimination and performance, the data set used in this study was highly imbalanced, consisting of many more surviving patients. For future studies, one should consider down-sampling the survivor group to have a balanced data set before training the models and comparing the model performance. Eighth, although the non-ML logistic regression model using the same set of features as the ML models suffered from a statistically lower (P<.001) AUROC of 0.857 (95% CI 0.855-0.859), as documented in Table 8, some clinicians may prefer to use a model with easier interpretability, a drawback of the multilayered deep neural networks [51]. Ninth, the training and testing cohorts were split by year in this study to better capture the cyclic seasonal change in infection. The randomness of the training and testing cohort splitting could be compromised.

Nevertheless, our study has multiple strengths. First, we demonstrated the strength of ML models in predicting sepsis mortality in an administrative database. Second, the data we used were from a sepsis cohort extracted using a validated approach from the NIS database, which is a large, standardized, nationwide database representative of US community hospitals. Third, the variables used in our models are easily accessible across different hospitals, thus having great generalizability. Fourth, our large sample size enabled our ML models to discover complex multi-way interactions and nonlinear relationships between the predictors and outcomes, prompting further investigations for other clinical researchers. Fifth, to increase the reproducibility and usability of this research on sepsis care and mortality, we also generated a web-based application that will allow peer investigators to obtain predicted 30-day sepsis mortality calculations.

Conclusions

In conclusion, our study demonstrates the value of ML models in predicting sepsis mortality in an administrative database as they are able to achieve higher discrimination and calibration. Knowledge of these ML models paves the way for the development of more accurate models to compare RSMRs across hospitals and geographic regions. This represents the first study to use an ML approach to improve the prediction of sepsis mortality in the NIS.

Acknowledgments

This work was supported by a research grant from the Taiwan Ministry of Science and Technology (MOST110-2314-B-002-053-MY3). No funding bodies had any role in the study design, data collection and analysis, the decision to publish, or the preparation of the manuscript.

Authors' Contributions

CCL has 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, concept and design, critical revision of the manuscript for important intellectual content, and obtaining funding and supervision. JYP contributed to the study idea and design, undertook the data analysis, developed the algorithm, wrote the first draft, and contributed to the subsequent drafts. TCH analyzed the data, constructed the code lists, developed the algorithms, and critically revised the manuscript. JRH, CYC, and WTH selected the health conditions, interpreted the data, and critically revised the manuscript for important intellectual content. JH analyzed the data and developed the algorithms. ML was responsible for the interpretation of the data and critical revision of the manuscript. All authors reviewed and interpreted the results, commented on the report, contributed to revisions, and read and approved the final version.

Conflicts of Interest

None declared. All authors have completed the International Committee of Medical Journal Editors uniform disclosure form and declare no support from any organization for the submitted work (or describe if any), no financial relationships with any organizations that might have an interest in the submitted work in the previous 3 years (or describe if any), and no other relationships or activities that could appear to have influenced the submitted work.

Multimedia Appendix 1
NIS_Sepsis_Main_Function.
[PDF File (Adobe PDF File), 136 KB - jmir_v24i4e29982_app1.pdf]

Multimedia Appendix 2
Introductory video of our sepsis mortality prediction web application.
[MP4 File (MP4 Video), 11623 KB - jmir_v24i4e29982_app2.mp4]

Multimedia Appendix 3
H2O ensemble Linux operating system.
[PDF File (Adobe PDF File), 81 KB - jmir_v24i4e29982_app3.pdf]

Multimedia Appendix 4
H2O ensemble Windows operating system.
[PDF File (Adobe PDF File), 85 KB - jmir_v24i4e29982_app4.pdf]

Multimedia Appendix 5
Coding book.
[XLSX File (Microsoft Excel File), 77 KB - jmir_v24i4e29982_app5.xlsx]

References

1. Rhee C, Jones TM, Hamad Y, Pande A, Varon J, O'Brien C, Centers for Disease Control and Prevention (CDC) Prevention Epicenters Program. Prevalence, underlying causes, and preventability of sepsis-associated mortality in US acute care hospitals. JAMA Netw Open 2019;2(2):e187571 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.7571] [Medline: 30768188]
2. Wang HE, Jones AR, Donnelly JP. Revised national estimates of emergency department visits for sepsis in the United States. Crit Care Med 2017;45(9):1443-1449 [FREE Full text] [doi: 10.1097/CCM.0000000000002538] [Medline: 28817480]
3. Moore JX, Donnelly JP, Griffin R, Howard G, Safford MM, Wang HE. Defining sepsis mortality clusters in the United States. Crit Care Med 2016;44(7):1380-1387 [FREE Full text] [doi: 10.1097/CCM.0000000000001665] [Medline: 27105174]
4. Paoli CJ, Reynolds MA, Sinha M, Gitlin M, Crouser E. Epidemiology and costs of sepsis in the United States-an analysis based on timing of diagnosis and severity level. Crit Care Med 2018;46(12):1889-1897 [FREE Full text] [doi: 10.1097/CCM.0000000000003342] [Medline: 30048332]
5. Bergmann S, Tran M, Robison K, Fanning C, Sedani S, Ready J, et al. Standardising hospitalist practice in sepsis and COPD care. BMJ Qual Saf 2019;28(10):800-808. [doi: 10.1136/bmjqs-2018-008829] [Medline: 30894422]
6. Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J. Mortality Probability Models (MPM II) based on an international cohort of intensive care unit patients. JAMA 1993;270(20):2478-2486. [Medline: 8230626]
7. Ferreira FL, Bota DP, Bross A, Mélot C, Vincent JL. Serial evaluation of the SOFA score to predict outcome in critically ill patients. JAMA 2001;286(14):1754-1758. [doi: 10.1001/jama.286.14.1754] [Medline: 11594901]
8. Knaus WA. APACHE 1978-2001: the development of a quality assurance system based on prognosis: milestones and personal reflections. Arch Surg 2002;137(1):37-41. [doi: 10.1001/archsurg.137.1.37] [Medline: 11772212]
9. Sivayoham N, Rhodes A, Cecconi M. The MISSED score, a new scoring system to predict mortality in severe sepsis in the emergency department: a derivation and validation study. Eur J Emerg Med 2014;21(1):30-36. [doi: 10.1097/MEJ.0b013e328364a8d4] [Medline: 23883776]
10. Granja C, Póvoa P, Lobo C, Teixeira-Pinto A, Carneiro A, Costa-Pereira A. The predisposition, infection, response and organ failure (Piro) sepsis classification system: results of hospital mortality using a novel concept and methodological approach. PLoS One 2013;8(1):e53885 [FREE Full text] [doi: 10.1371/journal.pone.0053885] [Medline: 23349756]
11. Osborn TM, Phillips G, Lemeshow S, Townsend S, Schorr CA, Levy MM, et al. Sepsis severity score: an internationally derived scoring system from the surviving sepsis campaign database*. Crit Care Med 2014;42(9):1969-1976. [doi: 10.1097/CCM.0000000000000416] [Medline: 24919160]

12. Ribas V, López JC, Ruiz-Sanmartin A, Ruiz-Rodríguez JC, Rello J, Wojdel A, et al. Severe sepsis mortality prediction with relevance vector machines. Annu Int Conf IEEE Eng Med Biol Soc 2011;2011:100-103. [doi: 10.1109/IEMBS.2011.6089906] [Medline: 22254260]
13. Tsoukalas A, Albertson T, Tagkopoulos I. From data to optimal decision making: a data-driven, probabilistic machine learning approach to decision support for patients with sepsis. JMIR Med Inform 2015;3(1):e11 [FREE Full text] [doi: 10.2196/medinform.3445] [Medline: 25710907]
14. Taylor RA, Pare JR, Venkatesh AK, Mowafi H, Melnick ER, Fleischman W, et al. Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data-driven, machine learning approach. Acad Emerg Med 2016;23(3):269-278 [FREE Full text] [doi: 10.1111/acem.12876] [Medline: 26679719]
15. Perng JW, Kao IH, Kung CT, Hung SC, Lai YH, Su CM. Mortality prediction of septic patients in the emergency department based on machine learning. J Clin Med 2019;8(11):1906 [FREE Full text] [doi: 10.3390/jcm8111906] [Medline: 31703390]
16. Kwon YS, Baek MS. Development and validation of a quick sepsis-related organ failure assessment-based machine-learning model for mortality prediction in patients with suspected infection in the emergency department. J Clin Med 2020;9(3):875 [FREE Full text] [doi: 10.3390/jcm9030875] [Medline: 32210033]
17. Joshi M, Ashrafian H, Arora S, Khan S, Cooke G, Darzi A. Digital alerting and outcomes in patients with sepsis: systematic review and meta-analysis. J Med Internet Res 2019;21(12):e15166 [FREE Full text] [doi: 10.2196/15166] [Medline: 31859672]
18. Sandhu S, Lin AL, Brajer N, Sperling J, Ratliff W, Bedoya AD, et al. Integrating a machine learning system into clinical workflows: qualitative study. J Med Internet Res 2020;22(11):e22421 [FREE Full text] [doi: 10.2196/22421] [Medline: 33211015]
19. Muralitharan S, Nelson W, Di S, McGillion M, Devereaux P, Barr NG, et al. Machine learning-based early warning systems for clinical deterioration: systematic scoping review. J Med Internet Res 2021;23(2):e25187 [FREE Full text] [doi: 10.2196/25187] [Medline: 33538696]
20. Lagu T, Lindenauer PK, Rothberg MB, Nathanson BH, Pekow PS, Steingrub JS, et al. Development and validation of a model that uses enhanced administrative data to predict mortality in patients with sepsis. Crit Care Med 2011;39(11):2425-2430. [doi: 10.1097/CCM.0b013e31822572e3] [Medline: 22005222]
21. Ford DW, Goodwin AJ, Simpson AN, Johnson E, Nadig N, Simpson KN. A severe sepsis mortality prediction model and score for use with administrative data. Crit Care Med 2016;44(2):319-327 [FREE Full text] [doi: 10.1097/CCM.0000000000001392] [Medline: 26496452]
22. König V, Kolzter O, Albuszies G, Thölen F. Einflussgrößen auf die krankenhaussterblichkeit bei patienten mit sepsis – entwicklung eines risikoadjustierten modells auf basis der leistungsdaten deutscher krankenhäuser. Z Evid Fortbild Qual Gesundhwes 2018;133:30-39. [doi: 10.1016/j.zefq.2018.03.001] [Medline: 29610028]
23. Rhee C, Wang R, Song Y, Zhang Z, Kadri SS, Septimus EJ, et al. Risk adjustment for sepsis mortality to facilitate hospital comparisons using centers for disease control and prevention's adult sepsis event criteria and routine electronic clinical data. Crit Care Explor 2019;1(10):e0049 [FREE Full text] [doi: 10.1097/CCE.0000000000000049] [Medline: 32166230]
24. Schwarzkopf D, Fleischmann-Struzek C, Rüddel H, Reinhart K, Thomas-Rüddel DO. A risk-model for hospital mortality among patients with severe sepsis or septic shock based on German national administrative claims data. PLoS One 2018;13(3):e0194371 [FREE Full text] [doi: 10.1371/journal.pone.0194371] [Medline: 29558486]
25. Doupe P, Faghmous J, Basu S. Machine learning for health services researchers. Value Health 2019;22(7):808-815 [FREE Full text] [doi: 10.1016/j.jval.2019.02.012] [Medline: 31277828]
26. Song X, Mitnitski A, Cox J, Rockwood K. Comparison of machine learning techniques with classical statistical models in predicting health outcomes. Stud Health Technol Inform 2004;107(Pt 1):736-740. [Medline: 15360910]
27. Introduction to the HCUP nationwide inpatient sample (NIS). Healthcare Cost and Utilization Project (HCUP). Rockville, MD: Agency for Healthcare Research and Quality; 2011. URL: https://www.hcup-us.ahrq.gov/db/nation/nis/NIS_Introduction_2011.jsp [accessed 2022-03-28]
28. Introduction to the HCUP national inpatient sample (NIS). Healthcare Cost and Utilization Project (HCUP). 2012. URL: https://www.hcup-us.ahrq.gov/db/nation/nis/NIS_Introduction_2012.jsp [accessed 2022-03-28]
29. Martin GS, Mannino DM, Eaton S, Moss M. The epidemiology of sepsis in the United States from 1979 through 2000. N Engl J Med 2003;348(16):1546-1554. [doi: 10.1056/NEJMoa022139] [Medline: 12700374]
30. Iwashyna TJ, Odden A, Rohde J, Bonham C, Kuhn L, Malani P, et al. Identifying patients with severe sepsis using administrative claims: patient-level validation of the angus implementation of the international consensus conference definition of severe sepsis. Med Care 2014;52(6):e39-e43 [FREE Full text] [doi: 10.1097/MLR.0b013e318268ac86] [Medline: 23001437]
31. Breiman L. Random forests. Mach Learn 2001;45(1):5-32. [doi: 10.1023/A:1010933404324]
32. Lundberg S, Lee SI. A unified approach to interpreting model predictions. arXiv. Preprint posted online November 25, 2017. [FREE Full text]
33. van der Laan MJ, Polley EC, Hubbard AE. Super learner. Stat Appl Genet Mol Biol 2007;6:Article25. [doi: 10.2202/1544-6115.1309] [Medline: 17910531]

34. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33(1):1-22 [FREE Full text] [Medline: 20808728]
35. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. Crit Care Med 2016;44(2):368-374 [FREE Full text] [doi: 10.1097/CCM.0000000000001571] [Medline: 26771782]
36. Winterstein AG, Choi Y, Meissner HC. Association of age with risk of hospitalization for respiratory syncytial virus in preterm infants with chronic lung disease. JAMA Pediatr 2018;172(2):154-160 [FREE Full text] [doi: 10.1001/jamapediatrics.2017.3792] [Medline: 29204660]
37. Breiman L. Bagging predictors. Mach Learn 1996;24(2):123-140. [doi: 10.1007/bf00058655]
38. Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Statist 2001;29(5):1189-1232. [doi: 10.1214/aos/1013203451]
39. Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA, Arshad H. State-of-the-art in artificial neural network applications: a survey. Heliyon 2018;4(11):e00938 [FREE Full text] [doi: 10.1016/j.heliyon.2018.e00938] [Medline: 30519653]
40. Sepsis Mortality Prediction. URL: https://hit3lab.shinyapps.io/SepsisMortalityPrediction/ [accessed 2022-03-28]
41. Gandhi S, Salmon JW, Kong SX, Zhao SZ. Administrative databases and outcomes assessment: an overview of issues and potential utility. J Manag Care Pharm 1999;5(3):215-222. [doi: 10.18553/jmcp.1999.5.3.215]
42. Outcome measures. Centers for Medicare & Medicaid Services. 2021. URL: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/OutcomeMeasures [accessed 2022-03-28]
43. Berta P, Martini G, Moscone F, Vittadini G. The association between asymmetric information, hospital competition and quality of healthcare: evidence from Italy. J R Stat Soc Ser A Stat Soc 2016;179(4):907-926. [doi: 10.1111/rssa.12214]
44. Desai NR, Ott LS, George EJ, Xu X, Kim N, Zhou S, et al. Variation in and hospital characteristics associated with the value of care for Medicare beneficiaries with acute myocardial infarction, heart failure, and pneumonia. JAMA Netw Open 2018;1(6):e183519 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.3519] [Medline: 30646247]
45. Angst C, Agarwal R, Gao GG, Khuntia J, McCullough JS. Information technology and voluntary quality disclosure by hospitals. Decis Support Syst 2014;57:367-375. [doi: 10.1016/j.dss.2012.10.042]
46. Iezzoni LI. Assessing quality using administrative data. Ann Intern Med 1997;127(8 Pt 2):666-674. [doi: 10.7326/0003-4819-127-8_part_2-199710151-00048] [Medline: 9382378]
47. Farmer SA, Black B, Bonow RO. Tension between quality measurement, public quality reporting, and pay for performance. JAMA 2013;309(4):349-350. [doi: 10.1001/jama.2012.191276] [Medline: 23340634]
48. Hashimoto RE, Brodt ED, Skelly AC, Dettori JR. Administrative database studies: goldmine or goose chase? Evid Based Spine Care J 2014;5(2):74-76 [FREE Full text] [doi: 10.1055/s-0034-1390027] [Medline: 25278880]
49. Vasilevskis EE, Kuzniewicz MW, Dean ML, Clay T, Vittinghoff E, Rennie DJ, et al. Relationship between discharge practices and intensive care unit in-hospital mortality performance: evidence of a discharge bias. Med Care 2009;47(7):803-812. [doi: 10.1097/MLR.0b013e3181a39454] [Medline: 19536006]
50. London AJ. Artificial intelligence and black-box medical decisions: accuracy versus explainability. Hastings Cent Rep 2019;49(1):15-21. [doi: 10.1002/hast.973] [Medline: 30790315]
51. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019;110:12-22. [doi: 10.1016/j.jclinepi.2019.02.004] [Medline: 30763612]

Abbreviations

AUC: area under the curve
AUC-PR: area under the precision-recall curve
AUROC: area under the receiver operating characteristic curve
EMR: electronic medical record
LASSO: least absolute shrinkage and selection operator
ML: machine learning
NIS: National Inpatient Sample
RSMR: risk-standardized mortality rate
SHAP: Shapley Additive Explanations

Edited by G Eysenbach; submitted 28.04.21; peer-reviewed by X Cheng, R Bajpai; comments to author 22.06.21; revised version received 07.09.21; accepted 06.03.22; published 13.04.22.

Please cite as:
Park JY, Hsu TC, Hu JR, Chen CY, Hsu WT, Lee M, Ho J, Lee CC
Predicting Sepsis Mortality in a Population-Based National Database: Machine Learning Approach
J Med Internet Res 2022;24(4):e29982
URL: https://www.jmir.org/2022/4/e29982
doi:10.2196/29982
PMID:35416785

Original Paper

Modeling Data Journeys to Inform the Digital Transformation of Kidney Transplant Services: Observational Study

Videha Sharma[1,2], BMBS, BMedSci; Iliada Eleftheriou[1], PhD; Sabine N van der Veer[1], PhD; Andrew Brass[1], PhD; Titus Augustine[2,3], MBBS, MS; John Ainsworth[1], PhD

[1]Centre for Health Informatics, Division of Informatics, Imaging and Data Science, The University of Manchester, Manchester, United Kingdom
[2]Department of Renal and Pancreatic Transplantation, Manchester University National Health Service Foundation Trust, Manchester, United Kingdom
[3]Division of Diabetes, Endocrinology and Gastroenterology, The University of Manchester, Manchester, United Kingdom

Corresponding Author:
Videha Sharma, BMBS, BMedSci
Centre for Health Informatics
Division of Informatics, Imaging and Data Science
The University of Manchester
Vaughan House
Portsmouth Street
Manchester, M13 9GV
United Kingdom
Phone: 44 7735360958
Email: videha.sharma@postgrad.manchester.ac.uk

Abstract

Background: Data journey modeling is a methodology used to establish a high-level overview of information technology (IT) infrastructure in health care systems. It allows a better understanding of sociotechnical barriers and thus informs meaningful digital transformation. Kidney transplantation is a complex clinical service involving multiple specialists and providers. The referral pathway for a transplant requires the centralization of patient data across multiple IT solutions and health care organizations. At present, there is a poor understanding of the role of IT in this process, specifically regarding the management of patient data, clinical communication, and workflow support.

Objective: To apply data journey modeling to better understand interoperability, data access, and workflow requirements of a regional multicenter kidney transplant service.

Methods: An incremental methodology was used to develop the data journey model. This included review of service documents, domain expert interviews, and iterative modeling sessions. Results were analyzed based on the LOAD (landscape, organizations, actors, and data) framework to provide a meaningful assessment of current data management challenges and inform ways for IT to overcome these challenges.

Results: Results were presented as a diagram of the organizations (n=4), IT systems (n>9), actors (n>4), and data journeys (n=0) involved in the transplant referral pathway. The diagram revealed that all movement of data was dependent on actor interaction with IT systems and manual transcription of data into Microsoft Word (Microsoft, Inc) documents. Each actor had between 2 and 5 interactions with IT systems to capture all relevant data, a process that was reported to be time consuming and error prone. There was no interoperability within or across organizations, which led to delays as clinical teams manually transferred data, such as medical history and test results, via post or email.

Conclusions: Overall, data journey modeling demonstrated that human actors, rather than IT systems, formed the central focus of data movement. The IT landscape did not complement this workflow and exerted a significant administrative burden on clinical teams. Based on this study, future solutions must consider regional interoperability and specialty-specific views of data to support multi-organizational clinical services such as transplantation.

(J Med Internet Res 2022;24(4):e31825) doi:10.2196/31825

KEYWORDS

digital transformation; health information exchange; interoperability; medical informatics; data journey modelling; kidney transplantation

Introduction

Data journey modeling is an emerging methodology developed to help understand the sociotechnical challenges and boundaries of data movement as part of digital transformation [1,2]. It has been used successfully to identify risks and costs of information technology (IT) projects within health care systems, such as the United Kingdom National Health Service (NHS) [3]. Specifically, data journey modeling provides a high-level overview of data entities, IT systems, manual processes, and organizations associated with a clinical service. It is a cross-collaborative methodology bridging health informaticians and clinical domain experts with the aim of producing a conceptual overview of the IT infrastructure pertinent to a clinical service. This allows a better understanding of how services are delivered from a data-centric perspective and helps inform meaningful solutions. As such, data journey modeling has been shown to identify opportunities for improving operational efficiency, data management, and patient safety, among other potential benefits [3]. The purpose of this study was to apply data journey modeling to a specific clinical use case, kidney transplantation, that was planning to undergo digital transformation.

Kidney transplantation is a regional, multi-organizational clinical service [4]. It is delivered at large university hospitals (in transplant centers), which receive patients from neighboring renal referral units. This hub-and-spoke model allows a wide geographical area to be covered and is similar to other specialist services, such as cancer, genetics, and vascular services. The patient journey in transplantation is complex and requires the capture of large volumes of heterogeneous clinical data. Multiple clinical teams are involved, and patients naturally cross organizational boundaries as they transition from declining kidney function to kidney failure and ultimately to kidney transplantation. Data capture during this patient journey requires meticulous administration to prevent delays and bottlenecks [5]. However, managing high-volume, complex clinical data across organizations is time consuming and error prone and incurs significant administrative costs. The 2014 United Kingdom Transplant First initiative recognized this, singling out “inefficient use of technology and administrative support” as one of the key barriers to timely transplantation [6]. The *American Journal of Transplantation* further highlighted the impact of the lack of integration of hospital-wide electronic patient records (EPRs) on kidney transplant care [7].

Owing to the aforementioned reasons, transplantation is a clinical area that will benefit from digital solutions to improve the management and flow of data. Health IT has been shown to successfully achieve these intended benefits; however, novel interventions are often marred by non-adoption and failure. [8]. A lack of understanding of the technical and organizational context for change is one of the key factors limiting success [9,10]. Further barriers exist due to a lack of consideration of the social aspect of interventions, which rely on human input and are therefore affected by resistance to change and failure to share perceived benefits with end users [11]. New interventions are often developed without including end users in the requirement-gathering process, and as a result, solutions are unsuccessful at achieving their intended benefits [1,12]. In an effort to successfully overcome these challenges, data journey modeling was identified as a methodology to understand the current IT infrastructure and involve domain experts in developing potential solutions.

The transplant referral process is an integral part of the overall transplant patient journey. It depends on the capture of data from various internal and external sources at the transplant center, concluding with the patient being registered on the national organ waiting list. This study aims to understand this process from a data journey perspective. Specific objectives were to (1) map the data management processes, including the role of IT support in a regional transplant network, (2) identify challenges and categorize them based on established frameworks, and (3) use the resulting findings to suggest potential solutions.

Methods

Overview

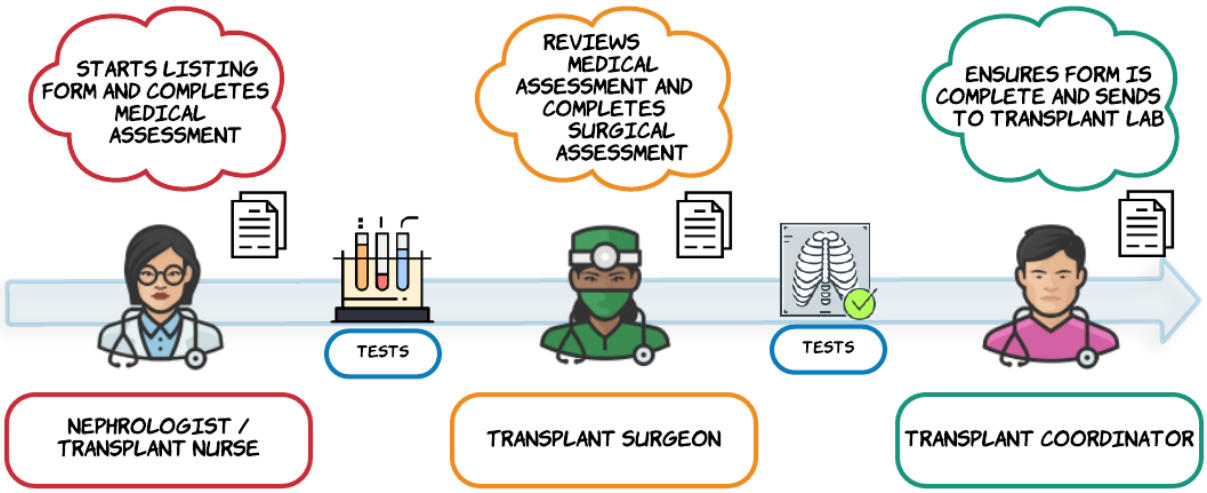
We followed an iterative and incremental approach to build the data journey model with input from clinical and administrative domain experts. We used the modeling process to identify potential challenges to data management and validated the final version of the model with domain experts who were not involved in the original modeling.

Context

The context for our study was the transplant center at the Manchester University NHS Foundation Trust (Manchester, United Kingdom). It is the largest kidney transplant center in the United Kingdom [13], receiving patients from 2 further regional renal referral units (Royal Salford NHS Foundation Trust and Lancashire Teaching Hospitals NHS Foundation Trust). The transplant center registers around 300 new patients on the national transplant waiting list every year. Patients are also under the care of a local general practice, which maintains long-term well-being through community-based medical care.

Each referral includes several hospital visits, medical tests, and clinical assessments. Multiple health care professionals are involved at different stages of the pathway. Data capture along the pathway is undertaken on a Microsoft Word (Microsoft Inc) document called the “listing form.” Various sections of the listing form are populated with patient data by members of the clinical team at multiple clinical time points. The captured data include routine health care data, such as medical history, test results, and examination findings. A complete and accurate listing form is required to assess the fitness of patients for transplantation and to permit registration on the national waiting list. Once the form is completed and the patient is deemed suitable for transplantation, the form is sent to the transplantation laboratory for registration (Figure 1).

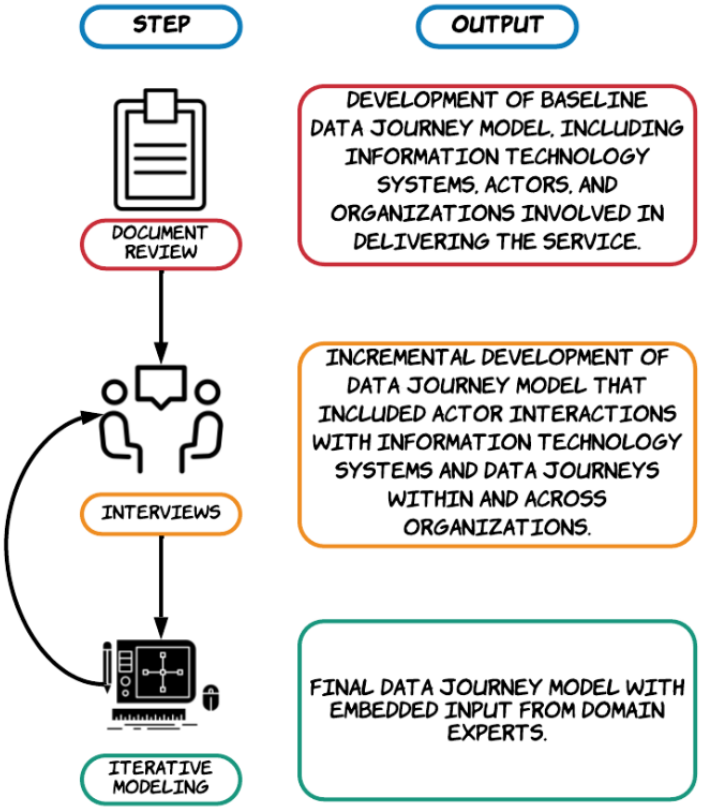
Figure 1. Data management in the transplant referral pathway is based on the transplant listing form.



Data Journey Modeling

Data journey modeling had 3 steps, summarized in Figure 2. The aim was to establish which IT systems contained transplant-related data, which organizations were involved in delivering the service, and which individuals delivered direct care or administration (ie, which individuals were actors), and to understand the interactions of the actors with the systems. This would provide a comprehensive overview of the IT infrastructure, the processes undertaken to extract and store data, and the data journeys, as part of the referral pathway. We then analyzed the results using an established framework, which was developed alongside data journey modeling, to help characterize our findings and draw meaningful conclusions [1]. Finally, we evaluated the final version of the model and our findings from the modeling process with domain experts who were not originally involved in developing the methodology.

Figure 2. Summary of data journey modeling steps with associated output of each step.



Document Review

We reviewed local written protocols pertaining to deceased donors, living donors, and transplant recipient pathways at the transplant center. We extracted all data entities routinely expected to be captured on the listing form and cross-referenced which IT systems these items were stored in. We identified which other health care organizations were involved in delivering the service and drew their boundaries. Finally, we established which actors played a role in the referral pathway within the transplant center. With this information, we designed a baseline iteration of the data journey model that demonstrated

the technical and organizational infrastructure but was still missing the actors and data journeys. We used Lucidchart software (Lucid Software, Inc), a web-based diagram and visual design application, to draw our model iterations.

Domain Expert Interviews

We conducted informal interviews and held small group meetings with domain experts working at the transplant center to gather information needed to further develop the model. We defined a domain expert as any member of the clinical or administrative team that was involved in direct patient care or back-office management of transplant-related data. We ensured this covered all the necessary actors identified through document review and the baseline iteration of the model. We spoke with 4 transplant coordinators, 2 nephrologists, 2 surgeons, 1 transplant assessment nurse, 2 secretaries, and 1 laboratory administrator. Domain expert interviews provided information on the processes used to extract and store data and the data journeys between IT systems and across organizational boundaries. Meetings lasted between 15 and 60 minutes; we kept written records of these meetings to increase accuracy and recall.

Iterative Modeling

We followed an Agile-inspired method to develop the model, based on an incremental approach. Agile is an adaptive project methodology that relies on continuous collaboration with stakeholders to change the output based on feedback and repeated cycles of review [14,15]. It has been shown to successfully accomplish goals in health care projects and is well suited to the development of a model that depends on embedding feedback from domain experts to iterate a final version [16].

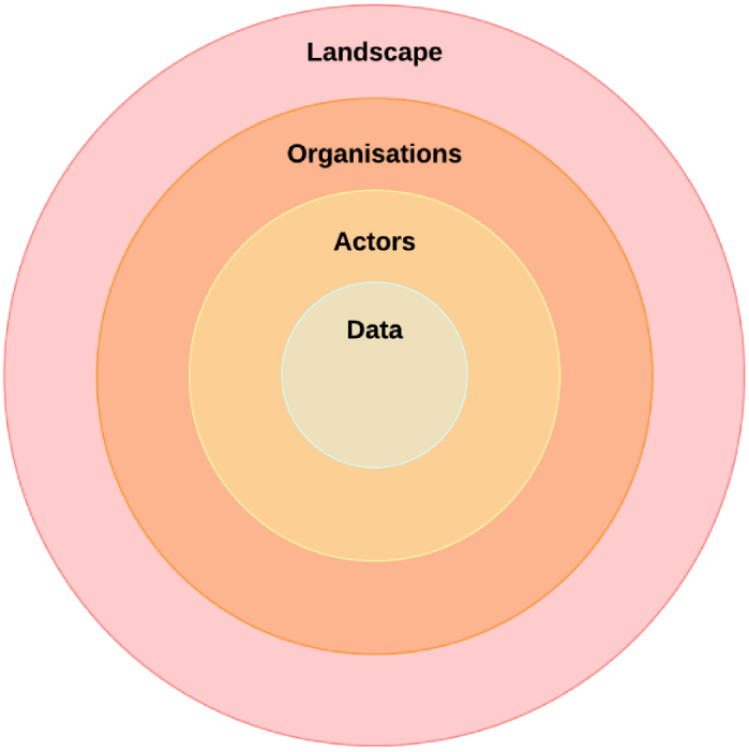
We modeled the processes that the various actors undertook in their work to deal with the key data entities and either capture and store or move data from one system to another. A total of 5 iteration sessions were held with the data journey modeler and domain experts to create the final model for analysis.

Analysis and External Evaluation

We used the LOAD framework to analyze the final version of the data journey model and categorize our findings. LOAD stands for “landscape, organization, actors, and data,” each denoting a dimension of IT as part of a clinical service (Figure 3) [1]. Using the LOAD framework ensured we comprehensively analyzed the model and associated data journeys, allowing us to identify technical barriers, such as lack of systems interoperability, and social challenges, such as manual work-arounds.

We then externally evaluated the final model by conducting semistructured interviews with domain experts who were not directly involved with model development. Interviewees included 2 transplant coordinators, 1 transplant surgeon, and 1 nephrologist. We presented them with the model and asked them if it accurately reflected the clinical workflow and data management processes at the transplant center. We prompted them to consider elements of the LOAD framework and think about how time spent on data management impacted delivery of the service and patient experience. The meetings typically lasted 30 minutes and were recorded in the form of research notes.

Figure 3. LOAD framework. LOAD: landscape, organizations, actors, data.



Results

Baseline Iteration of the Data Journey Model

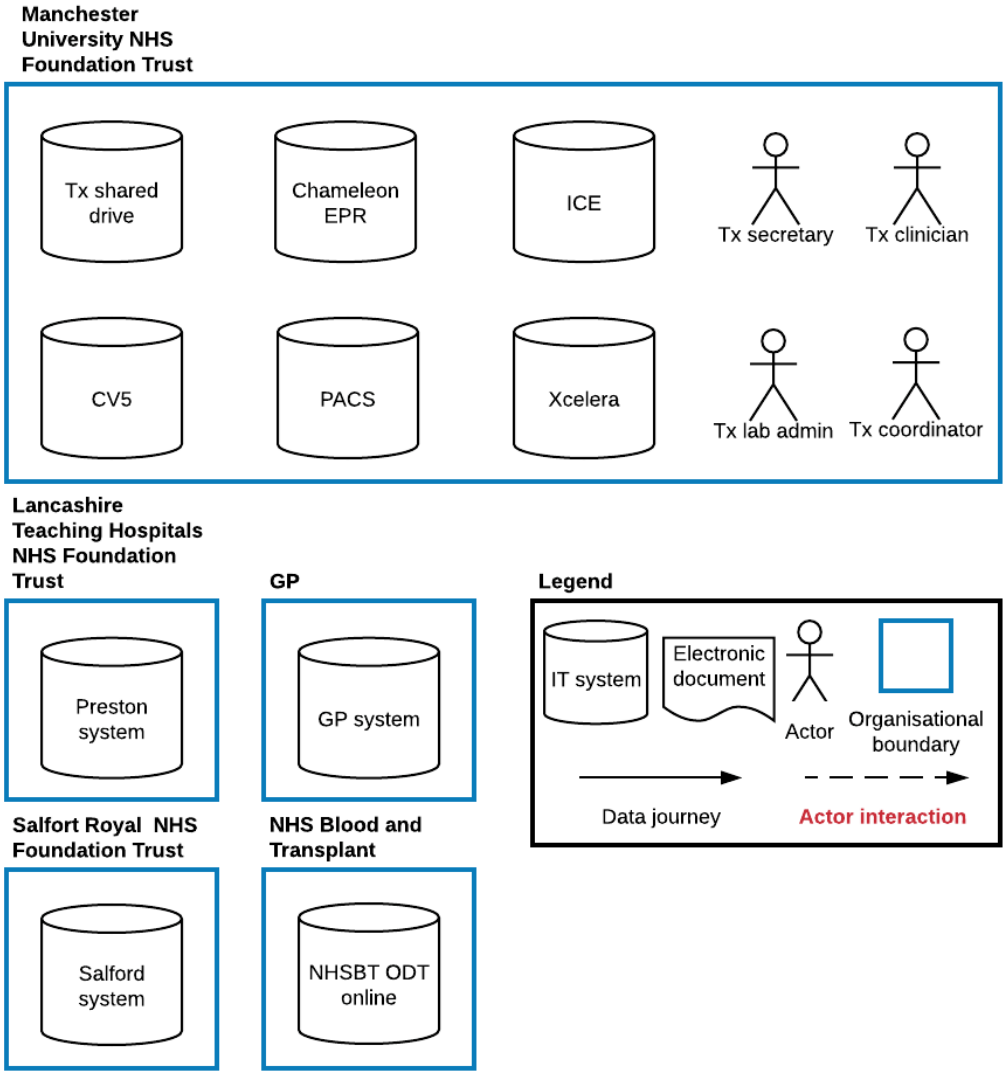
Based on the document review, we established the basic elements of our model. There were four organizations contributing patient data pertinent to delivering the service: 1 transplant center, 2 referring centers, and 1 general practice clinic. Within the transplant center, we identified 6 IT systems that held data related to the transplant referral pathway (Table 1). There were also several external IT systems outside of the organizational boundary of the transplant center that contained pertinent data. These were systems at general practice clinics containing medical history and medication data and systems at other trusts containing local medical history and results. As we did not map IT systems at other organizations in detail, we denoted them as a single IT system, although each organization may have had multiple systems in use. Finally, once data collection along the clinical pathway was complete, the data were transferred through a web-based system called Organ Donation and Transplantation Online (developed in-house by NHS Blood and Transplant) to register the patient on the national waiting list.

We identified a total of 4 actors that played a role in managing clinical data: clinicians, transplant coordinators, secretaries, and administrators. The term "clinician" referred to multiple specialists, including nephrologists, surgeons, and transplant assessment nurses. However, as their roles were similar from a data perspective, we denoted them as "transplant clinicians" for the purposes of our model. Figure 4 demonstrates the output of document review and the first iteration of the data journey model.

Table 1. Summary of all information technology systems at the Manchester University NHS Foundation Trust, their suppliers, and their clinical data management purposes. NHS: National Health Service; EPR: electronic patient record.

System	Supplier	Purpose
Chameleon EPR	In-house	Correspondence/results
Integrated Clinical Environment	CliniSys Group	Ordering tests
Picture Archiving and Communication System	General Electric Co.	General radiology
ClinicalVision 5	Constellation Kidney Group	Renal history/dialysis details
xCELERA	Philips NV	Cardiovascular imaging
Shared drive	Microsoft Inc.	Transplant listing form

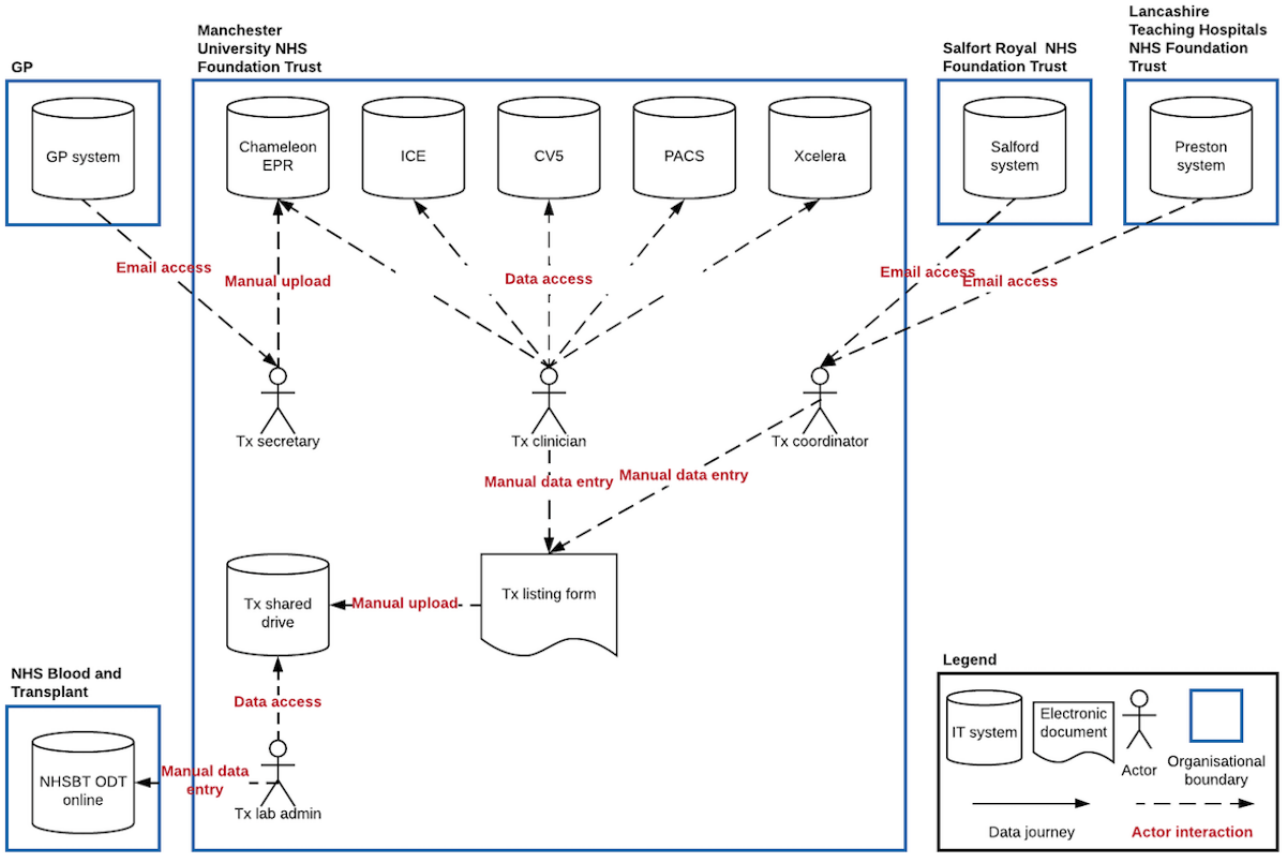
Figure 4. Baseline iteration of data journey model demonstrating information technology systems, organizational boundaries, and actors. CV5: Clinical Vision 5; EPR: electronic patient record; GP: general practice; ICE: integrated clinical environment; IT: information technology; NHS: National Health Service; NHSBT ODT: National Health Service Organ Donation and Transplantation; PACS: picture archiving and communication system; Tx: transplant.



Final Data Journey Model

The baseline model and domain expert interviews iteratively informed actor interactions and data journeys, which were added to the model to create the final version. The organizations were rearranged, placing the transplant center at the center of the model and the other organizations around it. There were no direct data journeys between IT systems within the transplant center or between systems across organizational boundaries. It became clear that the shared drive was the central focus of data management. This was an in-house solution resulting from the need to centrally capture and view clinical data; this need was not being met by existing systems. To complete the workflow, a minimum of 12 separate actor interactions with IT systems were necessary. Actors had the following minimal number of interactions with the IT systems: clinicians, 5 interactions; coordinators, 3 interactions; secretaries, 2 interactions; and administrators, 2 interactions. The final data journey model is shown in Figure 5.

Figure 5. Final data journey model, demonstrating the information technology landscape and data journeys in kidney transplant referral. CV5: Clinical Vision 5; EPR: electronic patient record; GP: general practice; ICE: integrated clinical environment; IT: information technology; NHS: National Health Service; NHSBT ODT: National Health Service Organ Donation and Transplantation; PACS: picture archiving and communication system; Tx: transplant.



LOAD Analysis

The final data journey model and feedback from external evaluation by domain experts allowed us to analyze findings based on the LOAD dimensions.

Landscape

The overall landscape demonstrated the complexity of the transplant referral pathway from a data perspective. The IT systems were not developed for the needs of the transplant service and have not been updated as the requirements have changed over time. A lack of interoperability across organizational boundaries raised data governance issues, and it was unclear whether data sharing between the organizations were the result of formal agreements. There was no IT system that provided a unified view of transplant data, which resulted in a work-around solution in the form of Microsoft Word documents and shared drives. This has led to a landscape where human actors, rather than IT systems, form the central focus of data movement.

Organizations

Key data were mainly stored internally within the transplant center's organizational boundary. Patient data, such as results of investigations not undertaken at the center, were stored externally at referral units and general practices. There were no direct data journeys from IT systems at external organizations to the transplant center; this data was typically transferred via post, email, or fax to the transplant coordinators. They then manually scanned paper-based data and saved it to the shared drive alongside other electronic data. We found that two-thirds of the patients going through the pathway were from external referral units. This meant that for the majority of patients registered on the waiting list, there were no up-to-date clinical data at the transplant center. All interviewees reported that this posed a significant challenge to clinical workflow. Time was spent chasing down data from referral units and there were frequent delays due to the need for repeated requests. An additional social challenge was the lack of accountability, with clinical staff being unclear who was responsible for data being updated and accurate: the transplant center or the referral units.

Actors

Data journeys were wholly dependent on actor interaction with IT systems and manual transcription of data. Key data was stored across multiple IT systems, which led to loss of efficiency as clinicians had to log in multiple times to view and extract data. Only 2 actor groups were able to interact with the shared drive, which meant that in their absence a patient would not be able to progress along the listing pathway. Domain experts reported that this created a bottleneck for the overall data journey and resulted in patient delays. Due to the impracticality of switching between multiple applications to access and transcribe data, actors reported using heuristic work-arounds, such as the use of 2 devices (eg, laptop and desktop). However, from interviews it emerged that there was variation in digital aptitude, and actors reported a range of experiences of interacting with the systems.

Data

We found that the listing form included a total of 247 data fields that needed to be populated. All required data were stored in the 5 IT systems of the transplant center and in the systems of the general practice or referral centers. There were no data journeys between IT systems or from systems to the transplant shared drive. To move data to the shared drive, clinicians had to access the different systems and transcribe (ie, type) clinical data into the relevant fields and save the form in the designated shared folder. The file name was saved as the patient's first and last name. All data required to populate the form were in electronic format. Data were directly transcribed without any clinical expertise required for transformation or manipulation. Domain expert interviews revealed that transcription errors and incomplete data fields were a source of both patient risk and delays in the listing pathway. There was also no current way of confirming data accuracy or obliging data completion. Interviewees further expressed their frustration at the time-consuming nature of the tasks, which detracted from time spent with patients.

Risk Mitigation Strategies

The above findings suggest that a regional solution with an agreed data sharing and governance contract would help mitigate the risks of the current fragmented landscape. A need has emerged for a central clinical data repository with a user interface accessible at the transplant center and referral units. Considering the range of multi-disciplinary actors involved in the transplant referral pathway, the user interface will have to be adaptable and easy to operate in order to lower barriers to adoption. Technically, such a solution would benefit from being web based and from using cloud storage to provide security and safe access across organizational boundaries. Interoperability and open data standards would underpin this integration of data across IT systems. Critically, a deep understanding of needs and requirements, as provided through the results of this study, should drive the development of solutions to achieve the intended benefits. This also holds true for health IT projects in other clinical domains, demonstrating the value of this methodology.

Discussion

Principal Findings

This study applied data journey modeling to evaluate the kidney transplant referral pathway and successfully identified the data, IT systems, actors, and organizations involved, as well as the relationship between them. This has provided an overview of the data landscape and highlighted the complexity of data administration, as well as the lack of data flow. We found that clinical staff must undertake cumbersome manual processes to summarize and visualize data from multiple IT systems. Work-arounds have been created in the absence of a meaningful solution to address the needs and requirements of the clinical workflow. The lack of interoperability and central access to relevant data increases the effort and time required to complete transplant referral, which can delay patients' registration on the transplant list.

Relation to Other Studies

This is the first study to apply data journey modeling to transplant services. Previous studies have highlighted the complexity of kidney transplantation from a clinical management perspective. These recommended the use of IT solutions, such as business process management technology, to lower management costs [17]. Our study has established the dependence on manual processes to administer data, which is likely to incur management costs. The current data landscape strictly serves a documentation process, and does not provide any process support. Experience across the European Union shows that contemporary IT systems and EPR systems must provide functionality beyond data capture to better support the needs of clinical services [18].

This study found that data journeys in the transplant pathway naturally crossed specialty and organizational boundaries. However, with the absence of interoperability there was a dependence on actor interaction to share data. In other clinical areas, access to data across organizational boundaries continues to be a significant challenge [19]. The introduction of a national EPR system in Finland has facilitated implementation of digital pathways across nephrology and transplantation [20]. However, larger nations with more heterogeneous populations and geographical variations face challenges in harmonizing fragmented health care data [21]. Data journey modeling, such as that performed in this study, confirms that interoperability remains one of the key barriers to meaningful digital transformation.

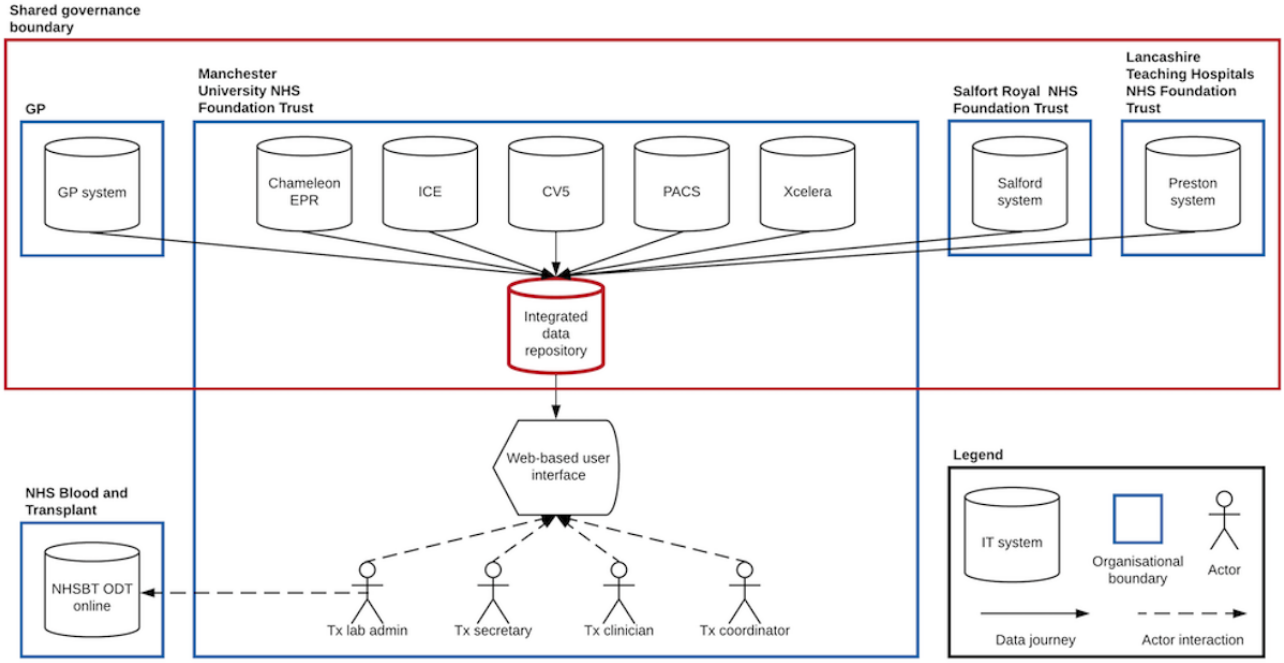
Implications for Practice and Future Concepts

Data journey modeling showed that during the referral pathway, clinicians are not required to transform or manipulate any data in order to complete the form—thus the IT challenge is one of summarizing and viewing relevant information in a format that allows seamless and enhanced clinical decision-making. In the United Kingdom, general practices recognize the value of early, customized viewing of clinical data, and primary-care IT systems are more intuitive for clinicians' use [22]. However, in the hospital setting, a paradox exists in which IT systems commonly detract from patient contact due to dependence on user interaction to view data [23,24]. An early study by Zeng et al [25] evaluated concept-oriented views of clinical data versus traditional chronological presentation of data, such as in current EPRs. They demonstrated that visualizing data in their clinical context, such as the disease or organ system, reduced information overload and increased the accuracy of data retrieval. Based on the data fields identified in this study, for kidney transplantation this would include presenting a single-screen summary of relevant demographics, as well as medical and social history, with details pertinent to dialysis and previous surgeries. This would allow clinicians to focus on the patient at the time of an encounter, and add relevant clinical details not previously recorded in any IT system to the patients' record, such as residual urine output, exercise tolerance, or examination findings [26].

Findings from this study highlight the technical requirements for a transplant-specific solution with a regional, integrated data store that spans the relevant organizations with an application

processing interface that meets the needs and requirements of the clinical workflow (Figure 6). Separate data and application layers for health care IT may help overcome current interoperability barriers and enable development of modular service-specific solutions [27]. Centralized clinical data repositories may facilitate application of model-view-controller software development, giving individual clinical areas the opportunity to design views to suit their context [28]. Semantic interoperability across systems allows data to be readily exchanged, analyzed, and interpreted, and is a prerequisite for meaningful digital transformation. In contrast, digital data stored in isolated databases not only slows down medical progress, but also limits technological innovations such as real time analytics and the reuse of data for research. [19,26]. Solutions such as Fast Healthcare Interoperability Resources protocols and OpenEHR archetypes may address these challenges going forward, but development is still required before widespread adoption [29,30].

Figure 6. Conceptual overview of a proposed solution including a regional integrated data repository with a web-based clinical user interface. CV5: Clinical Vision 5; EPR: electronic patient record; GP: general practice; ICE: integrated clinical environment; IT: information technology; NHS: National Health Service; NHSBT ODT: National Health Service Organ Donation and Transplantation; PACS: picture archiving and communication system; Tx: transplant.



The Healthcare Information and Management Systems Society has defined digital maturity of individual health care providers based on capabilities, interoperability, and governance [31]. However, due to the multicenter nature of transplant services, we found that digital maturity was limited by the least mature organization that formed part of delivering the service. Thus, even if the transplant center had an advanced and unified EPR system, the fact that patients were referred by other organizations unable to share data implied that clinical processes could not be adequately supported. Evaluating the potential impact of any novel solution should therefore be undertaken using interoperability frameworks [32]. In addition, capturing quantitative data, such as the time taken to be added to the transplant waiting list, could provide a measure of impact.

Across health care, clinical data remains constrained to organizational boundaries, and new EPR procurement does not actively consider regional workflow or data sharing, reinforcing vendor lock-in [33]. In response to this, NHS England launched the "Local Health and Care Record Exemplars," a project tasked with increasing clinical information sharing across primary, secondary, and social care within a region [34,35]. Transplantation may be an excellent use case for such interoperability initiatives to demonstrate value to clinicians, policy-makers and, crucially, patients. Linked data will provide the basis for learning health systems that are intuitive to their populations' needs and inform timely interventions to improve long-term health and social care outcomes. [36,37].

Limitations of This Study

A number of other models to evaluate health IT infrastructure exist. The data journey model and LOAD framework have been developed based on the UK health care context and were chosen as the most appropriate tools to use [38]. However, they have not been widely applied in other clinical areas, potentially because they rely heavily on domain expertise to provide input during the modeling process. In our case, the study was led by a clinical research fellow who was able to help bridge the gap between the clinical and academic stakeholders. Finally, this study looked at only a single regional transplant center. This leaves it unknown to what extent our findings would translate to other regions, warranting further investigation.

Conclusion

Complex clinical care pathways must be fully understood to allow meaningful solutions to be presented as part of digital transformation initiatives. Data journey modeling successfully provided valuable sociotechnical factors for health IT in kidney

transplantation. It highlighted how a lack of interoperability led to time-consuming manual interaction with multiple systems to summarize data for transplant referral. Data crossed multiple organizational boundaries, and all movement of data depended on actor interaction, even though no data were transformed or manipulated. Future solutions must consider regional interoperability, bespoke systems that meet clinical requirements, and automated processes that free clinical staff from administrative burdens.

Conflicts of Interest

None declared.

References

1. Eleftheriou I, Embury S, Brass A. Data journey modelling: predicting risk for IT developments. In: The Practice of Enterprise Modeling. PoEM 2016. Lecture Notes in Business Information Processing. 2016 Oct 28 Presented at: 9th IFIP Working Conference on The Practice of Enterprise Modeling; Skovde, Sweden. Springer; 8-10 Nov 2016; Skovde, Sweden p. 72-86. [doi: 10.1007/978-3-319-48393-1_6]
2. Eleftheriou I. Data journey modelling: identifying cost and risk in large, complex, socio-technical systems. Manchester, United Kingdom: University of Manchester; 2019 Jul 26. URL: https://www.research.manchester.ac.uk/portal/files/122867819/FULL_TEXT.PDF [accessed 2022-04-14]
3. Eleftheriou I, Embury S, Moden R, Dobinson P, Brass A. Data journeys: Identifying social and technical barriers to data movement in large, complex organisations. J Biomed Inform 2018 Feb;78:102-122 [FREE Full text] [doi: 10.1016/j.jbi.2017.12.001] [Medline: 29223464]
4. Dudley C, Harden P. Renal Association Clinical Practice Guideline on the assessment of the potential kidney transplant recipient. Nephron Clin Pract 2011;118 Suppl 1:c209-c224 [FREE Full text] [doi: 10.1159/000328070] [Medline: 21555897]
5. Staes C, Evans R, Narus S, Huff S, Sorensen J. System analysis and improvement in the process of transplant patient care. 2007 Presented at: MEDINFO 2007; 20-24 Aug, 2007; Brisbane, Australia p. 915-919.
6. Barcley M, Burnapp L. Transplant First: Timely Listing for Kidney Transplantation. NHS Kidney Care. 2013. URL: https://www.england.nhs.uk/improvement-hub/wp-content/uploads/sites/44/2017/11/Transplant-First-Timely-Listing-for-Kidney-Transplantation.pdf [accessed 2021-07-25]
7. Pondrom S. The AJT report: news and issues that affect organ and tissue transplantation. Missing the mark: hospital-wide electronic medical record systems don't always accommodate transplantation's specific needs. Am J Transplant 2012 Jan;12(1):1-2 [FREE Full text] [doi: 10.1111/j.1600-6143.2011.03936.x] [Medline: 22244120]
8. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
9. Buntin MB, Burke MF, Hoaglin MC, Blumenthal D. The benefits of health information technology: a review of the recent literature shows predominantly positive results. Health Aff (Millwood) 2011 Mar;30(3):464-471 [FREE Full text] [doi: 10.1377/hlthaff.2011.0178] [Medline: 21383365]
10. Brender J, Ammenwerth E, Nykänen P, Talmon J. Factors influencing success and failure of health informatics systems--a pilot Delphi study. Methods Inf Med 2006;45(1):125-136. [Medline: 16482383]
11. 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]
12. Zheng L, Yang T, Pardo T, Jiang Y. Understanding the "boundary" in information sharing and integration. 2015 Presented at: 42nd Hawaii International Conference on System Sciences. IEEE; 2009; Hawaii, USA. [doi: 10.1109/hicss.2009.481]
13. Organ Donation and Transplantation Activity Report 2018/19. NHS Blood and Transplant. URL: https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/19220/activity-report-2019-2020.pdf [accessed 2021-07-26]
14. Kitzmiller R, Hunt E, Sproat SB. Adopting best practices: "Agility" moves from software development to healthcare project management. Comput Inform Nurs 2006;24(2):75-82; quiz 83. [doi: 10.1097/00024665-200603000-00005] [Medline: 16554690]
15. Špundak M. Mixed Agile/Traditional Project Management Methodology – Reality or Illusion? Procedia - Social and Behavioral Sciences 2014 Mar;119:939-948. [doi: 10.1016/j.sbspro.2014.03.105]
16. Ap Dafydd D, Williamson R, Blunt P, Blunt DM. Development of training-related health care software by a team of clinical educators: their experience, from conception to piloting. Adv Med Educ Pract 2016;7:635-640 [FREE Full text] [doi: 10.2147/AMEP.S108426] [Medline: 27853396]
17. Andellini M, Fernandez Riesgo S, Morolli F, Ritrovato M, Cosoli P, Petruzzellis S, et al. Experimental application of Business Process Management technology to manage clinical pathways: a pediatric kidney transplantation follow up case.

BMC Med Inform Decis Mak 2017 Nov 03;17(1):151 [FREE Full text] [doi: 10.1186/s12911-017-0546-x] [Medline: 29100512]
18. Stroetmann K, Artmann J, Stroetmann V. Developing national eHealth infrastructures: results and lessons from Europe. 2011 Presented at: AMIA Annual Symposium Proceedings; 22-26 Oct 2011; Washington, USA.
19. Lehne M, Sass J, Essenwanger A, Schepers J, Thun S. Why digital medicine depends on interoperability. NPJ Digit Med 2019;2:79 [FREE Full text] [doi: 10.1038/s41746-019-0158-1] [Medline: 31453374]
20. Savikko J, Rauta V. Implementing eHealth in Kidney Transplantation in Finland. Transplant Proc 2019 Mar;51(2):464-465. [doi: 10.1016/j.transproceed.2019.01.064] [Medline: 30879568]
21. Iroju O, Soriyan A, Gambo I, Olaleke J. Interoperability in healthcare: benefits, challenges and resolutions. International Journal of Innovation and Applied Studies 2013;3(1):262-270.
22. Benson T. Why general practitioners use computers and hospital doctors do not—Part 1: incentives. BMJ 2002 Nov 09;325(7372):1086-1089 [FREE Full text] [Medline: 12424171]
23. Westbrook JI, Ampt A, Kearney L, Rob MI. All in a day's work: an observational study to quantify how and with whom doctors on hospital wards spend their time. Med J Aust 2008 May 05;188(9):506-509. [Medline: 18459920]
24. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: a path to validation. NPJ Digit Med 2019;2:38 [FREE Full text] [doi: 10.1038/s41746-019-0111-3] [Medline: 31304384]
25. Zeng Q, Cimino JJ, Zou KH. Providing concept-oriented views for clinical data using a knowledge-based system: an evaluation. J Am Med Inform Assoc 2002;9(3):294-305 [FREE Full text] [doi: 10.1197/jamia.m1008] [Medline: 11971890]
26. Topol E. The Topol review: preparing the healthcare workforce to deliver the digital future. Health Education England. 2019. URL: https://topol.hee.nhs.uk/wp-content/uploads/HEE-Topol-Review-2019.pdf [accessed 2021-07-26]
27. Garde S, Knaup P, Hovenga E, Heard S. Towards semantic interoperability for electronic health records. Methods Inf Med 2007;46(3):332-343. [doi: 10.1160/ME5001] [Medline: 17492120]
28. Johnson CM, McIlwain S, Gray O, Willson B, Vorderstrasse A. Creating a sustainable collaborative consumer health application for chronic disease self-management. J Biomed Inform 2017 Jul;71:198-206 [FREE Full text] [doi: 10.1016/j.jbi.2017.06.004] [Medline: 28600025]
29. Saripalle R, Runyan C, Russell M. Using HL7 FHIR to achieve interoperability in patient health record. J Biomed Inform 2019 Jun;94:103188 [FREE Full text] [doi: 10.1016/j.jbi.2019.103188] [Medline: 31063828]
30. Christensen B, Ellingsen G. Evaluating Model-Driven Development for large-scale EHRs through the openEHR approach. Int J Med Inform 2016 May;89:43-54. [doi: 10.1016/j.ijmedinf.2016.02.004] [Medline: 26980358]
31. Cresswell K, Sheikh A, Krasuska M, Heeney C, Franklin BD, Lane W, et al. Reconceptualising the digital maturity of health systems. Lancet Digit Health 2019 Sep;1(5):e200-e201 [FREE Full text] [doi: 10.1016/S2589-7500(19)30083-4] [Medline: 33323267]
32. Carvalho JV, Rocha, Abreu A. Maturity Models of Healthcare Information Systems and Technologies: a Literature Review. J Med Syst 2016 Apr 15;40(6):131. [doi: 10.1007/s10916-016-0486-5]
33. Panch T, Mattie H, Celi LA. The "inconvenient truth" about AI in healthcare. NPJ Digit Med 2019;2:77 [FREE Full text] [doi: 10.1038/s41746-019-0155-4] [Medline: 31453372]
34. Scott P, Dunscombe R, Evans D, Mukherjee M, Wyatt J. Learning health systems need to bridge the 'two cultures' of clinical informatics and data science. J Innov Health Inform 2018 Jun 15;25(2):126-131 [FREE Full text] [doi: 10.14236/jhi.v25i2.1062] [Medline: 30398454]
35. Castle-Clarke S, Hutchings R. Achieving a digital NHS. Nuffield Trust. URL: https://www.nuffieldtrust.org.uk/research/achieving-a-digital-nhs-lessons-for-national-policy-from-the-acute-sector [accessed 2021-07-26]
36. Ainsworth J, Buchan I. Combining Health Data Uses to Ignite Health System Learning. Methods Inf Med 2015;54(6):479-487. [doi: 10.3414/ME15-01-0064] [Medline: 26395036]
37. Benson T, Grieve G. Principles of Health Interoperability. In: Principles of Health Interoperability SNOMED CT, HL7 and FHIR. Switzerland: Springer; 2016.
38. Cresswell K, Williams R, Sheikh A. Developing and Applying a Formative Evaluation Framework for Health Information Technology Implementations: Qualitative Investigation. J Med Internet Res 2020 Jun 10;22(6):e15068 [FREE Full text] [doi: 10.2196/15068] [Medline: 32519968]

Abbreviations

EHR: electronic patient record
GP: general practice
IT: information technology
LOAD: landscape, organizations, actors, data
NHS: National Health Service

Edited by G Eysenbach; submitted 07.07.21; peer-reviewed by O Sverdlov, S Hume; comments to author 25.11.21; revised version received 27.12.21; accepted 09.02.22; published 21.04.22.

Please cite as:
Sharma V, Eleftheriou I, van der Veer SN, Brass A, Augustine T, Ainsworth J
Modeling Data Journeys to Inform the Digital Transformation of Kidney Transplant Services: Observational Study
J Med Internet Res 2022;24(4):e31825
URL: https://www.jmir.org/2022/4/e31825
doi:10.2196/31825
PMID:35451983

Original Paper

Physician Gender, Patient Risk, and Web-Based Reviews: Longitudinal Study of the Relationship Between Physicians' Gender and Their Web-Based Reviews

Danish Hasnain Saifee[1], PhD; Matthew Hudnall[1], PhD; Uzma Raja[2], PhD

[1]Department of Information Systems, Statistics, and Management Science, The University of Alabama, Tuscaloosa, AL, United States
[2]Department of Systems and Technology, Auburn University, Auburn, AL, United States

Corresponding Author:
Matthew Hudnall, PhD
Department of Information Systems, Statistics, and Management Science
The University of Alabama
801 University Blvd
Tuscaloosa, AL, 35487-0290
United States
Phone: 1 205 348 0856
Email: matthew.hudnall@ua.edu

Abstract

Background: Web-based reviews of physicians have become exceedingly popular among health care consumers since the early 2010s. A factor that can potentially influence these reviews is the gender of the physician, because the physician's gender has been found to influence patient-physician communication. Our study is among the first to conduct a rigorous longitudinal analysis to study the effects of the gender of physicians on their reviews, after accounting for several important clinical factors, including patient risk, physician specialty, and temporal factors, using time fixed effects. In addition, this study is among the first to study the possible gender bias in web-based reviews using statewide data from Alabama, a predominantly rural state with high Medicaid and Medicare use.

Objective: This study conducts a longitudinal empirical investigation of the relationship between physician gender and their web-based reviews using data across the state of Alabama, after accounting for patient risk and temporal effects.

Methods: We created a unique data set by combining data from web-based physician reviews from the popular physician review website, RateMDs, and clinical data from the Center for Medicare and Medicaid Services for the state of Alabama. We used longitudinal econometric specifications to conduct an econometric analysis, while controlling for several important clinical and review characteristics across four rating dimensions (helpfulness, knowledge, staff, and punctuality). The overall rating and these four rating dimensions from RateMDs were used as the dependent variables, and physician gender was the key explanatory variable in our panel regression models.

Results: The panel used to conduct the main econometric analysis included 1093 physicians. After controlling for several clinical and review factors, the physician random effects specifications showed that male physicians receive better web-based ratings than female physicians. Coefficients and corresponding SEs and *P* values of the binary variable *GenderFemale* (1 for female physicians and 0 otherwise) with different rating variables as outcomes were as follows: *OverallRating* (coefficient –0.194, SE 0.060; *P*=.001), *HelpfulnessRating* (coefficient –0.221, SE 0.069; *P*=.001), *KnowledgeRating* (coefficient –0.230, SE 0.065; *P*<.001), *StaffRating* (coefficient –0.123, SE 0.062; *P*=.049), and *PunctualityRating* (coefficient –0.200, SE 0.067; *P*=.003). The negative coefficients indicate a bias toward male physicians versus female physicians for aforementioned rating variables.

Conclusions: This study found that female physicians receive lower web-based ratings than male physicians even after accounting for several clinical characteristics associated with the physicians and temporal effects. Although the magnitude of the coefficients of *GenderFemale* was relatively small, they were statistically significant. This study provides support to the findings on gender bias in the existing health care literature. We contribute to the existing literature by conducting a study using data across the state of Alabama and using a longitudinal econometric analysis, along with incorporating important clinical and review controls associated with the physicians.

(J Med Internet Res 2022;24(4):e31659) doi:10.2196/31659

KEYWORDS

web-based physician reviews; gender; gender bias; patient perception; Alabama; patient risk

Introduction

Background

Web-based reviews of physicians have been gaining significant popularity among health care consumers or patients over the past 2 decades. Some examples of popular websites for web-based physician reviews are RateMDs [1], Vitals [2], and HealthGrades [3]. The prominence of these reviews is enhanced as the health care landscape in the United States becomes more patient-centric. Patients are becoming more involved in the management of their own health care. Although the review websites were initially popular among certain demographics [4], over time, they have gained significant popularity across a substantial portion of patient population. In fact, a recent survey of web-based physician reviews found that approximately 95% of the respondents viewed web-based reviews to be *somewhat reliable* or *very reliable*, and approximately 70% of respondents said that their choice of a physician was affected by the ratings or reviews on web-based physician review websites [5].

The literature on web-based reviews of physicians has been growing in the past 10 years. Using data from the United States and other countries, numerous studies have examined the content and valence of web-based physician or hospital reviews and the factors that could explain their variance [6-15]. A substream of this literature examined the relationship between the clinical outcomes or performance of physicians and their web-based reviews. The results were quite mixed [13]. Some studies have found a statistically significant association between physicians' clinical performance and their web-based reviews [16-18]. On the other hand, some studies have found that physicians with better clinical practices or outcomes do not receive better web-based reviews [19-21].

Another substream has investigated the influence of web-based physician reviews on patients' choices. There has been a significant interest among health care researchers and practitioners in the health care consumers' awareness of web-based physician reviews [22]. Several studies have investigated whether web-based physician reviews impact patients' choices and whether there are certain characteristics of these reviews that impact the choice. These studies found that high number of reviews and high valence of reviews were associated with a more positive attitude toward the rated physicians and their selection by patients [23-25].

The increasing reliance on web-based physician reviews is indicated by other surveys also [26]. These surveys of web-based reviews also reveal that a significant portion of patients checks the web-based reviews of physicians, even if they were referred to these physicians by their health care providers. Collectively, these findings reveal the extent to which web-based reviews of physicians have become prominent among patients or health care consumers.

As web-based health care information, including physician reviews, is publicly available and easily accessible, there has been a long-standing concern among the health care providers and research communities about the quality and clinical relevance of web-based health care information [27]. The interaction between health care providers and their patients can affect the patients' opinions of them. In turn, these opinions can become web-based reviews that are accessible to anyone searching for their physicians' information on the web.

There has been a long-established interest among researchers in the impact of physician gender on patient communication and patients' choice of physicians. Extant literature has found that female physicians tend to engage in patient-centered communication [28-30] and do not receive ratings as high as their male counterparts [31,32]. It has also been proposed that the relationship between physicians and their patients might be affected by the physician's gender and different expectations of patients from male and female physicians [33-35]. The dynamics of patients' communication and relationship with physicians of different genders have received significant attention in the extant literature [36,37].

Questions about whether patients have a preference for male physicians over female physicians, and vice versa, and whether their opinions of physicians are affected by the physicians' gender have also received substantial attention from health care professionals and researchers. For instance, in a survey of 185 patients, Fennema et al [38] found that 43% of women and 12% of men preferred a female physician, whereas 31% of men and 9% of women preferred a male physician and that patients who preferred male physicians reported technical competence to be a more prominent characteristic of male physicians. In a different survey, Kerssens et al [39] did not find a preference for surgeons or anesthesiologists of a particular gender, but found preferences for female physicians as gynecologists in 8.5:1 ratio and general practitioners in 2.32:1 ratio among female respondents. In another survey of 125 women, Plunkett et al [40] found that the gender of a physician was not of primary importance when selecting an obstetrician or gynecologist. Some of these studies have also attempted to identify the mechanisms that may have led to their findings. There have also been calls for suggestions on making health care workplaces more equitable for female physicians [41].

With the proliferation of web-based physician reviews among patients or health care consumers, a natural and important question is, "Whether and to what extent is a physician's gender related to their online reviews after accounting for patient risk and time shocks (time fixed effects)?"

After a careful review of the existing literature, we found that the potential effect of physician gender on web-based reviews of physicians has not received sufficient attention. In the few studies that have examined the relationship between physicians' gender and their web-based reviews, the findings have been mixed. For example, Dunivin et al [32] and Thawani et al [42] found that female physicians receive lower ratings than male physicians. On the other hand, Emmert and Meier [43] found that female physicians receive better ratings than their male

counterparts. Marrero et al [44] found that female surgeons receive more positive ratings for social interaction, whereas male surgeons receive better ratings for technical aspects. Clearly, the possible effect of physicians' gender on their web-based reviews, or lack thereof, requires more thorough examination.

In the examination of the aforementioned relationship, it is important to account for the characteristics of patients, such as patient risk, in some form. It is also important to account for the variation in the reviews over time to determine the direct relationship between physicians' gender and their web-based reviews. Including patient risk allows us to account for the health characteristics of a significant patient population under the care of physicians. Not controlling for such characteristics can potentially bias the results because a physician's interaction can be affected by the existing health condition of their patients. Therefore, we examine the effect of physician gender on web-based patient reviews, while controlling for patient health risks over time.

Objective

To the best of our knowledge, our study is among the first to examine the effect of physicians' gender on their web-based reviews over time and after accounting for patient risk. Furthermore, our study is the first to conduct such an investigation using physician data across Alabama, a state that has received very little attention in the literature on web-based physician reviews. We accomplish our analysis by using a unique data set that we created by combining data from web-based physician reviews from a popular physician review website, RateMDs, and clinical data from the Center for Medicare and Medicaid Services (CMS) for the state of Alabama.

Methods

Ethics Approval

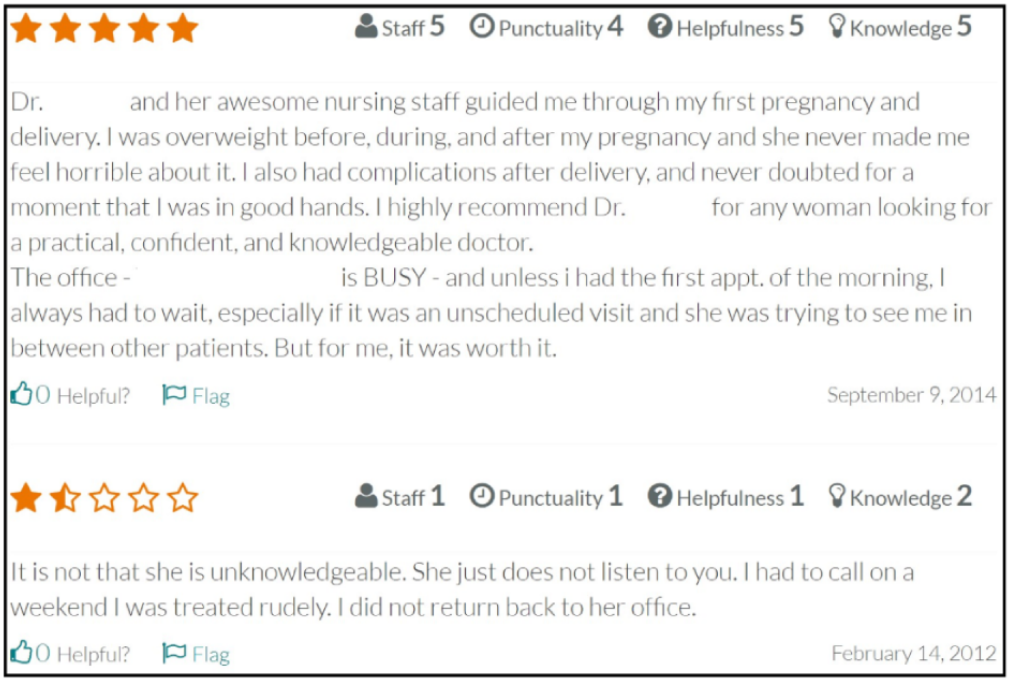
No ethics board review or approval was required for this study. All the raw data that were collected for this study are publicly available on the web.

Data

To study whether web-based reviews of physicians are more favorable toward male or female physicians, we constructed a panel data set of physicians in Alabama using data from 2 sources. The unit of analysis in our study was a physician, and the time periods in the panel were years. We collected data on web-based reviews and the gender of physicians from RateMDs to construct our web-based review data set spanning from 2012 to 2018. We used Python (Python Software Foundation) to collect data from RateMDs. We also obtained clinical data on physicians from Medicare Provider Utilization and Payment Data: Physician and Other Supplier [45], which traversed the same time frame of 2012 to 2018. We combined the data from these sources using a combination of physicians' first names, last names, specialty, and years. Our final unbalanced panel data set had 1093 matching physicians over a 7-year time span (2012 to 2018) that matched both data sets. There were 5912 physicians in the RateMDs data set who had at least one review, and there were a total of approximately 26,600 reviews across these physicians. Among these 5912 physicians, 2673 (45.21%) physicians had reviews in at least two years. We were able to match 40.89% (1093/2673) of these physicians with our data from CMS, and this 40.89% (1093/2673) of the physicians constituted the panel used to conduct the panel analysis in this study.

Each physician in our final panel has a unique national provider identification number that was collected from CMS. This ensured that all the physicians in our final panel were unique. Figure 1 shows an anonymized selection of reviews from RateMDs for a physician in our data set. As shown in Figure 1, a physician can receive numeric ratings on four different dimensions (staff, punctuality, helpfulness, and knowledge). Along with these numeric ratings, a physician can also receive textual comments. The dates on which the reviews were provided on RateMDs is also shown in Figure 1. Patient reviews on RateMDs and optional responses by the physicians are free of charge. Paid tiers for physicians exist on RateMDs, but they do not allow for the alteration of reviews. The paid tiers allow for physicians to be notified of new ratings, the ability to feature a rating, appointment requests, photos, and other features, but no paid feature inhibits the ability of a person to post a review on the site.

Figure 1. Example screenshot of RateMDs reviews for a physician.



Measures

As we were examining whether the web-based reviews of physicians are favorable to male or female physicians, we constructed our dependent variables using the numeric physician ratings from RateMDs. Physicians on RateMDs can be rated on four dimensions: helpfulness, knowledge, staff, and punctuality. The ratings for each of these dimensions are on a scale of 1 to 5, with 5 being the best possible score and 1 being the lowest score. To capture the information in each of these four dimensions, we constructed the following four dependent variables: *HelpfulnessRating*, *KnowledgeRating*, *StaffRating*, and *PunctualityRating*. *HelpfulnessRating* was the average of the ratings received by a physician on the helpfulness dimension in a year. Similarly, *KnowledgeRating*, *StaffRating*, and *PunctualityRating* were the averages of the ratings received by a physician on the knowledge, staff, and punctuality dimensions, respectively. To capture the combined information across these dimensions, we constructed a panel variable, *OverallRating*. For this purpose, initially, we constructed a variable *NetRating* using the average of the ratings received on the four aforementioned dimensions. Then, we constructed *OverallRating* by calculating the average of *NetRating* in each year, similar to how we constructed *HelpfulnessRating*, *KnowledgeRating*, *StaffRating*, and *PunctualityRating*.

Our key explanatory variable was a time-invariant variable, *GenderFemale*, which equals 1 for female physicians and 0 for male physicians. We obtained data on the gender of the physicians from RateMDs. We also used several control variables to account for the clinical aspects associated with the physician and with the textual comments that go alongside numeric *RateMDs* ratings. Our control variables included *RiskScore*, *TopicCare*, *TopicSurgery*, *TopicStaff*, and *Specialty*.

RiskScore was the average yearly hierarchical condition category (HCC) risk score calculated by CMS using data on Medicare beneficiaries [45]. HCC coding can provide information about patient complexity and a description of the medical complications a patient is experiencing. HCC relies on the International Classification of Diseases–10th Edition coding to assign risk scores to patients [46]. A physician with high *RiskScore* would have Medicare beneficiaries with high risk scores (above-average spending). This variable allowed us to control for the patient risk score of the Medicare patients under the care of a physician. As Medicare is one of the largest health care insurers or payers in the United States, *RiskScore* helped us to account for the patient risk of a significant proportion of the patient population under the care of physicians.

TopicCare was the proportion of textual reviews received by a physician each year, in which the dominant underlying theme was care provided by the physician. *TopicStaff* was the proportion of textual reviews in which the dominant underlying theme was the office or staff of the physician. *TopicSurgery* was the proportion of textual reviews in which the dominant underlying theme was the surgical proficiency of the physician.

To construct these topics (latent topics), we used topic modeling techniques based on Latent Dirichlet Allocation (LDA) [47,48]. LDA has been used extensively for topic modeling in the extant literature on web-based reviews of products and services, including several studies involving web-based physician reviews [9,49-53]. The following sections provide a brief description of the main steps through which we used topic modeling to construct the aforementioned topic variables. We used R (R Foundation for Statistical Computing) for topic modeling.

We created a corpus of all the reviews using an R text-mining package(TM) within RStudio, after which we converted the corpus to lower case [54-56]. We also replaced punctuation, numbers, and stop words. We *stemmed* the corpus to allow us to reduce words with a common root to the root word, such as *nurse* and *nursing* to the word fragment *nurs*. Next, we created

the *document-term matrix*, which stored the frequencies of stemmed words in our textual comment corpus by each textual comment. Then, we leveraged the LDA algorithm and used an R package (topicmodels) to extract topics from our textual comments [57-59]. These R packages have been widely used in the literature mentioned previously to construct latent topics or themes from textual data. For each comment, a probability was assigned to each of the identified latent themes or topics, and the probabilities summed up to 1 for each comment. We classified each comment based on the topic that had the highest probability. We identified the most common words within each of the 3 target latent topics, as shown in Textbox 1. We chose these 3 topics because it was the minimum number of topics that we could use to clearly categorize the experiences with the physicians and their staff reported in textual reviews [20,21]. Textbox 1 shows the stemmed words most closely (probabilistically) associated with each of the 3 review comment topics.

Textbox 1. Most prominent words (after stemming) by topic.

TopicCare

- care, doctor, staff, recommend, patient, time, knowledg, help, friend, love, wonder, high, listen, excel, and feel

TopicStaff

- time, office, doctor, wait, staff, patient, appoint, call, nurs, rude, visit, day, question, hour, and talk

TopicSurgery

- doctor, surgeri, pain, care, medic, life, patient, treat, recommend, time, day, surgeon, procedur, treatment, and feel

We had physicians from across 34 specialties in our final panel data set. The 15 specialties with most physicians (in descending order of the number of physicians) were as follows: general (family) practice, obstetrics and gynecology, internal medicine, orthopedic surgery, neurology, otolaryngology, cardiology, ophthalmology and optometry, psychiatry, dermatology, general surgery, podiatry, urology, endocrinology, and rheumatology. Physicians in these 15 specialties accounted for approximately 85.73% (937/1093) of all the physicians in our panel data set. Table S1 in Multimedia Appendix 1 lists the number of male and female physicians across specialties in the panel data set. The physician specialties were time-invariant binary variables. Controlling for the specialties allowed us to compare the effect of the physicians' gender on their reviews after accounting for the numerous unobservable time-invariant clinical aspects that could influence physicians of both genders within each specialty. We also conducted further robustness checks by including additional clinical review control variables. These control measures helped us significantly distinguish our research from previous studies.

Analysis

We used physician random effects panel regression, along with year fixed effects to account for time shocks. A time shock in the context of this paper can be considered as an event or collection of events that can impact physicians across the board in the duration of a year. For example, a statewide or nationwide health care policy change would likely have an impact on physicians across different specialties. As the analysis used panel data, it was important to account for such time shocks. We did so by including year fixed effects in our regression specifications. We used Stata (StataCorp) for conducting our econometric analysis.

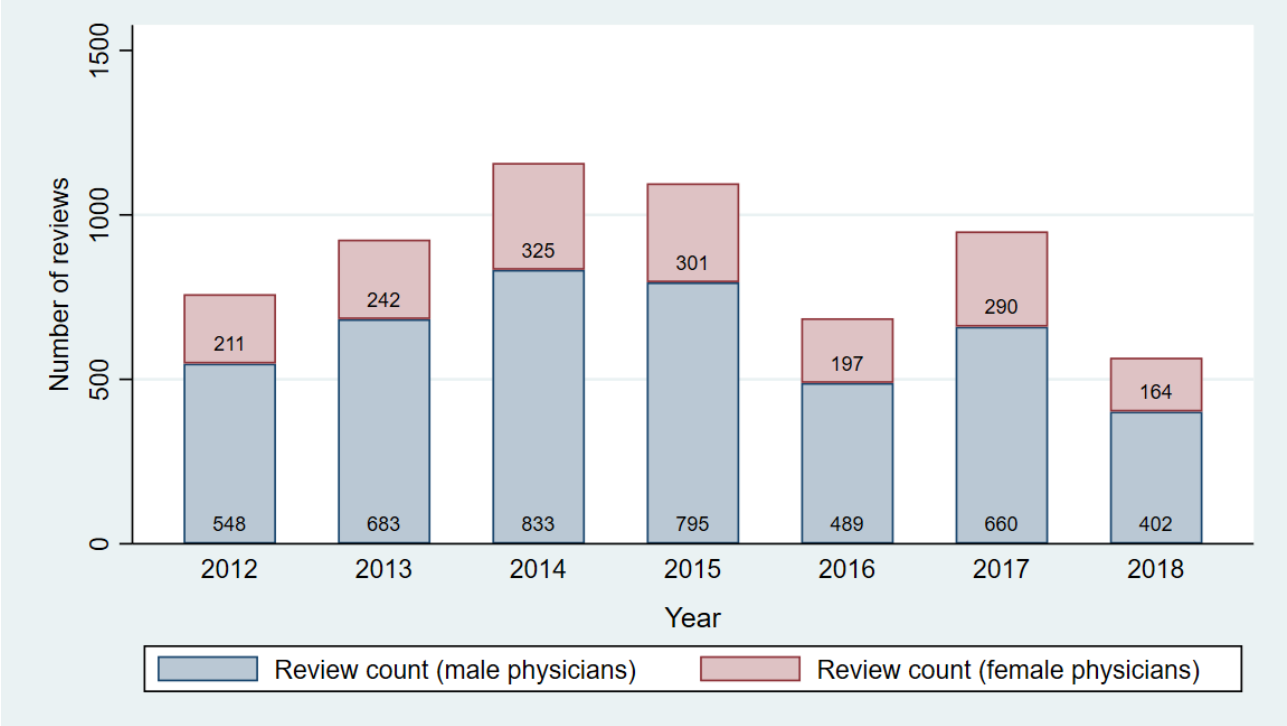
We leveraged the physician random effects model instead of the physician fixed effects model to estimate the effect of physician gender because of the following reasons: (1) our main explanatory variable, *PhysicianGender*, was time-invariant, and physician fixed effects would have subsumed the *PhysicianGender* variable and (2) a physician's gender can be safely assumed to be randomly assigned in the context of our study, and thus, it was very unlikely that there were unobserved variables that could simultaneously drive or influence both the physician gender and their web-based reviews. The year fixed effects allowed us to account for the time shocks in the health care industry or web-based physician review websites that can influence physicians across the state of Alabama. The SEs shown in all the panel regression specifications were robust. For brevity, we do not report the coefficients, SEs, and *P* values of the different specialties and year fixed effects. The sum of *TopicCare*, *TopicSurgery*, and *TopicStaff* was equal to 1. In our specifications, *TopicStaff* was the base topic variable, and thus, not included in the regressions. One of the specialties and one of the years acted as the base specialty and base year, respectively, and thus, were not included in the regression specifications.

Results

Descriptive Statistics

Figure 2 shows the distribution of the number of reviews for male and female physicians across the years from the original RateMDs data set. This chart and the subsequent figures were created using the 1093 physicians who were present in our panels across CMS and RateMDs data used for panel regressions. Our panel consisted of a 7-year period spanning from 2012 to 2018 to include a broad set of historical data that were also relatively current. As shown in Figure 2, the year 2014 had the highest number of reviews, whereas 2018 had the lowest number of reviews across the physicians in our panel, and there were ample number of physician reviews across all years in our panel.

Figure 2. Distribution of total number of physician reviews across years.



Figures 3-7 show plots of the average annual values of *OverallRating*, *HelpfulnessRating*, *KnowledgeRating*, *StaffRating*, and *PunctualityRating* for male and female physicians for the 1093 physicians from RateMDs who were in the panel. As shown in Figure 3, the average *OverallRating* of male physicians was consistently higher than that of female physicians. The average annual ratings on all 4 dimensions were more favorable for male physicians across most years. The time trends depicted in these figures revealed reviews more favorable toward male physicians than female physicians. The variation in the difference in the average values is visible in these figures and warrants a thorough longitudinal investigation of the effect of physician gender on the web-based ratings. Accordingly, we conducted a longitudinal or panel empirical investigation of the effect of a physician's gender on their web-based ratings. As stated previously, we controlled for several clinical and review characteristics associated with physicians, and by doing so, we isolated the direct effect of a physician's gender on their web-based ratings.

Figure 3. Comparison of average overall ratings for female and male physicians across years.

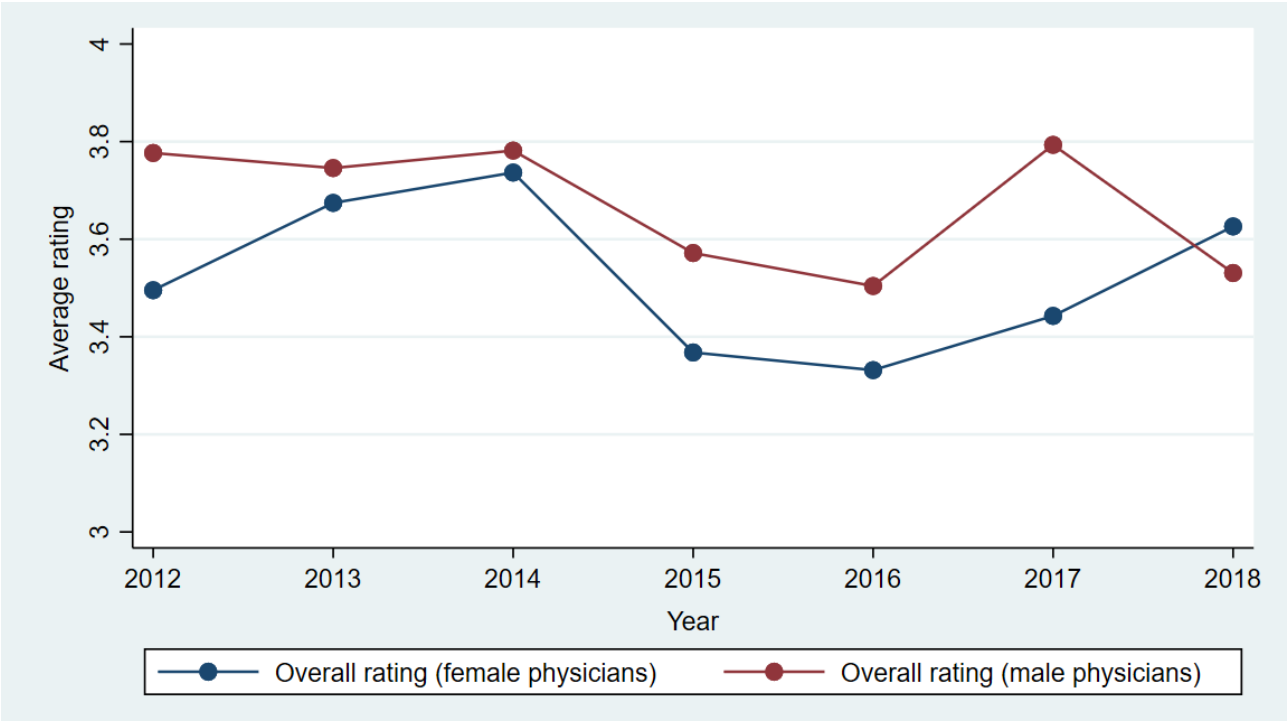


Figure 4. Comparison of average helpfulness ratings for female and male physicians across years.

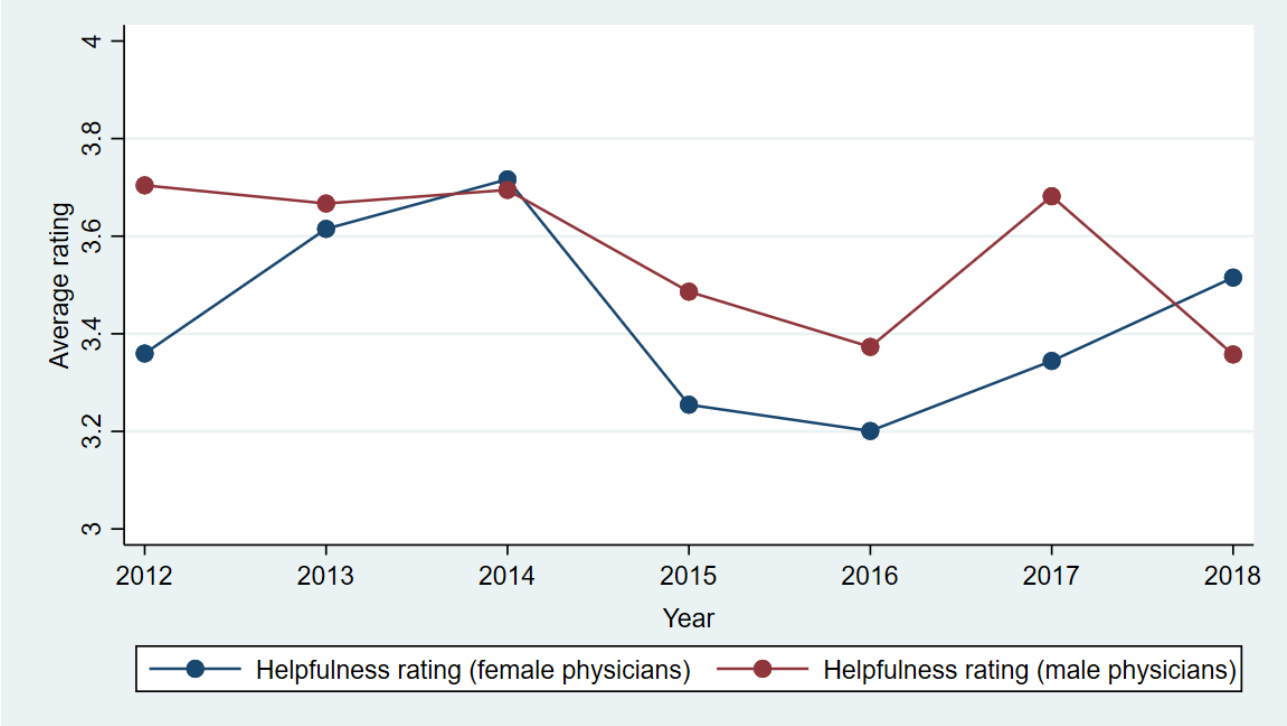


Figure 5. Comparison of average knowledge ratings for female and male physicians across years.

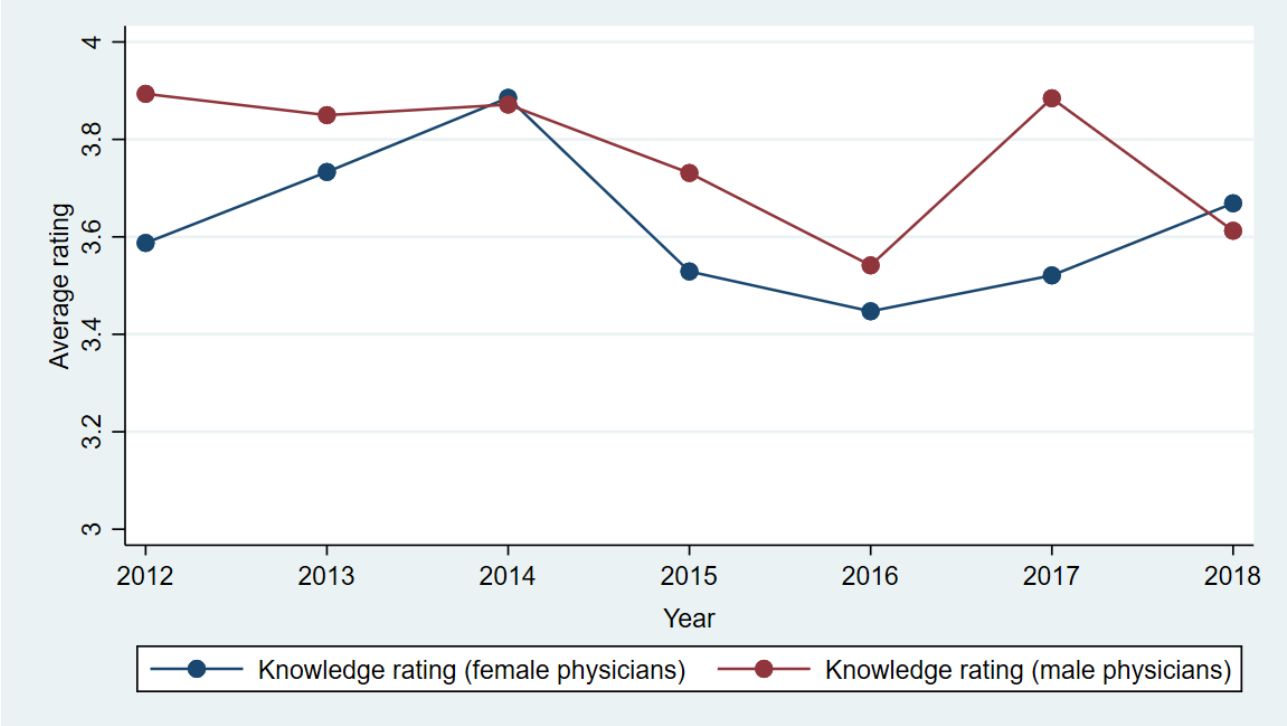


Figure 6. Comparison of average staff ratings for female and male physicians across years.

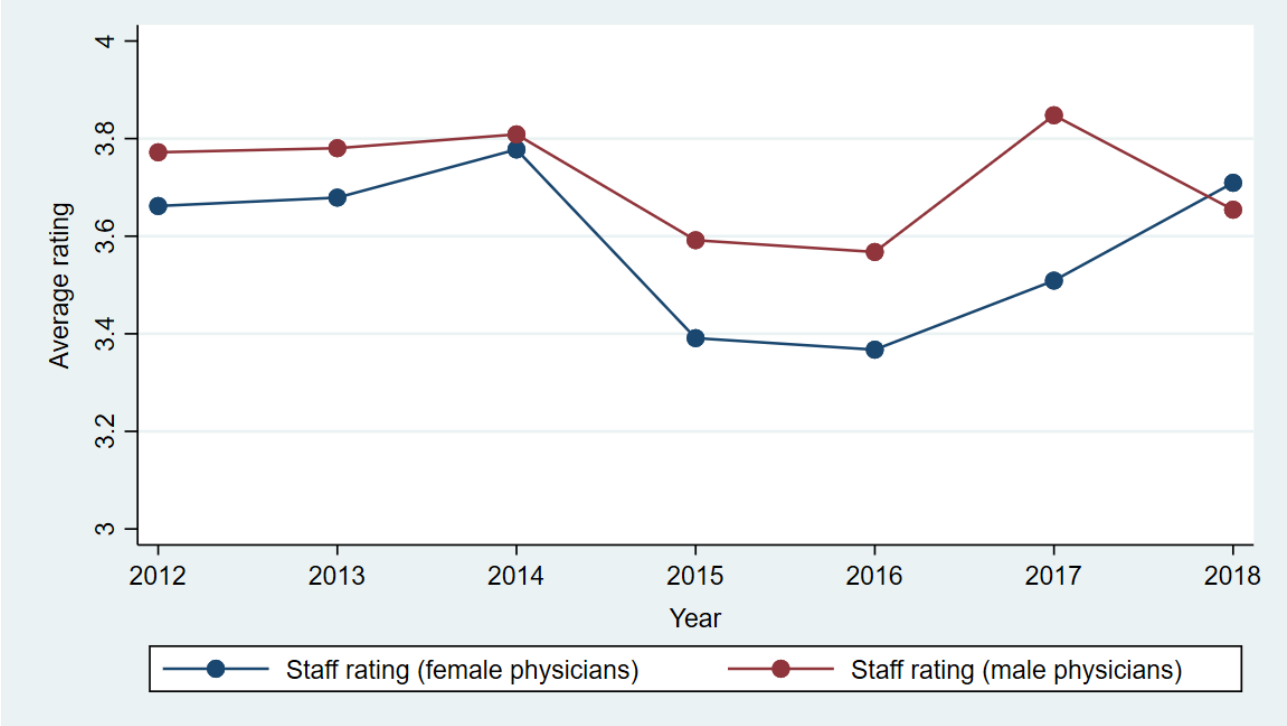


Figure 7. Comparison of average punctuality ratings for female and male physicians across years.

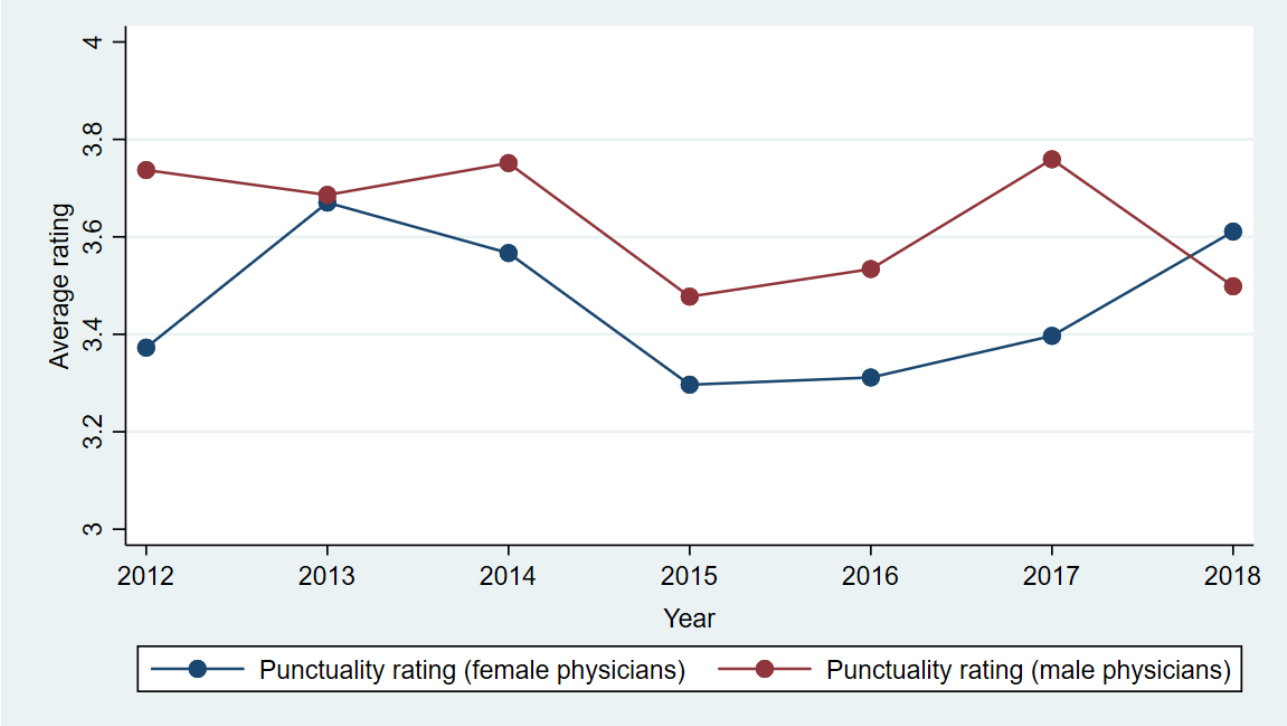


Table 1 shows the descriptive statistics of the various dependent variables, topic controls, and *RiskScore* control. The average values of the rating variables were between 3.5 and 3.6. In our panel, approximately 25.34% (277/1093) of the physicians were women. In Alabama, female physicians account for approximately 28.5% (3025/10,614) of the overall physician population [60]. This suggests that the overall distribution of physician gender in our panel was fairly representative of that in Alabama.

Table 1. Descriptive statistics (number of observations=3446).

Variable	Values, mean (SD)	Values, median	Values, minimum	Values, maximum
OverallRating	3.64 (1.43)	4.25	1	5
HelpfulnessRating	3.54 (1.65)	4.37	1	5
KnowledgeRating	3.74 (1.54)	5	1	5
StaffRating	3.69 (1.48)	4	1	5
PunctualityRating	3.60 (1.49)	4	1	5
TopicCare	0.41 (0.45)	0	0	1
TopicSurgery	0.27 (0.40)	0	0	1
TopicStaff	0.32 (0.42)	0	0	1
RiskScore	1.23 (0.41)	1.14	0.53	5.62

Effect of Gender

Table 2 provides the results of random effects panel regression, with *OverallRating* as the dependent variable. We included physician specialties as controls and year fixed effects in each of the regression specifications. The SEs of each specification were robust. As shown in Table 2, the coefficient of *GenderFemale* was negative and statistically significant, implying that female physicians tend to receive worse overall web-based ratings than their male counterparts. The coefficient of *RiskScore* was statistically insignificant in all the specifications, implying that physicians who treat Medicare patients of high risk tend to not receive better or worse overall ratings than their counterparts who treat Medicare patients of low risk. The coefficients of *TopicCare* and *TopicSurgery* were positive and statistically significant, implying that the physicians who receive a high proportion of review comments with an underlying theme of physician care and surgical aspects tend to have better overall ratings than those who receive a high proportion of review comments with an underlying theme of their office and staff. In Tables 3 and 4, the coefficient of *GenderFemale* was negative and statistically significant for *HelpfulnessRating*, *KnowledgeRating*, and *PunctualityRating*, but not for *StaffRating*.

The coefficient of *RiskScore* was statistically insignificant for each of the four rating dimensions, whereas that of *TopicCare* and *TopicSurgery* were positive and statistically significant. The magnitude of the coefficient of *GenderFemale* was close to 0.2. This means that, on average, female physicians receive ratings lower by 0.2 points than their male counterparts. For example, on average, if male physicians receive a rating of 4 out of 5, their female counterparts would receive a rating of 3.8 out of 5.

Table 2. Estimation for OverallRating (N=1093)[a].

Variable	Coefficient (SE)	*P* value
GenderFemale	−0.162 (0.060)	.007
RiskScore	−0.056 (0.086)	.52
TopicCare	1.557 (0.058)	<.001
TopicSurgery	0.739 (0.071)	<.001

[a]Specialty controls=yes; year fixed effects=yes; robust SE=yes; overall R-squared=0.267; within R-squared=0.168; between R-squared =0.339.

Table 3. Estimation for HelpfulnessRating and KnowledgeRating (N=1093).

Variable	*HelpfulnessRating* [a]		*KnowledgeRating* [b]	
	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value
GenderFemale	−0.185 (0.069)	.008	−0.198 (0.065)	.002
RiskScore	0.003 (0.098)	.97	−0.057 (0.094)	.54
TopicCare	1.702 (0.069)	<.001	1.492 (0.064)	<.001
TopicSurgery	0.688 (0.084)	<.001	0.513 (0.080)	<.001

[a]Specialty controls=yes; year fixed effects=yes; robust SE=yes; overall R-squared=0.239; within R-squared=0.153; between R-squared=0.310.

[b]Specialty controls=yes; year fixed effects=yes; robust SE=yes; overall R-squared=0.220; within R-squared=0.137; between R-squared=0.282.

Table 4. Random effects panel regression (StaffRating and PunctualityRating; N=1093).

Variable	*StaffRating* [a]		*PunctualityRating* [b]	
	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value
GenderFemale	−0.095 (0.062)	.13	−0.172 (0.067)	.01
RiskScore	−0.045 (0.087)	.61	−0.127 (0.105)	.23
TopicCare	1.547 (0.063)	<.001	1.488 (0.063)	<.001
TopicSurgery	0.923 (0.076)	<.001	0.832 (0.074)	<.001

[a]Specialty controls=yes; year fixed effects=yes; robust SE=yes; overall R-squared=0.247; within R-squared=0.155; between R-squared=0.315.

[b]Specialty controls=yes; year fixed effects=yes; robust SE=yes; overall R-squared=0.234; within R-squared=0.130; between R-squared=0.318.

Robustness Checks

We added additional control variables to check whether our findings would change. The three additional variables were *BeneficiaryCount*, *ServicesCount*, and *WordCount*. *BeneficiaryCount* was the number of Medicare beneficiaries under the care of a physician in a year. *ServicesCount* was the number of services provided by a physician in a year. *WordCoun*t was the average number of words in the review comments received by a physician in a year. Tables S2-S4 in Multimedia Appendix 1 provide the results of panel specifications with additional control variables. Table S2 in Multimedia Appendix 1 provides the results with *OverallRating* as the dependent variable. Table S3 in Multimedia Appendix 1 provides the results with *Helpfulness* and *KnowledgeRating* as the dependent variables, and Table S4 in Multimedia Appendix 1 provides the results with *StaffRating* and *PunctualityRating* as the dependent variables. As can be observed in Tables S2-S4 in Multimedia Appendix 1, the coefficients *of GenderFemale* were negative and statistically significant for *OverallRating* and each of the four rating dimensions, including *StaffRating*. The magnitude of coefficient of *GenderFemale* was close but slightly higher than those in Tables 2-4.

We conducted further robustness checks by removing the specialties in our panel in which both genders were not represented. This helped us mitigate the concern that a possible bias may arise owing to the absence of physicians of one of the genders in any of the specialties in our panel. The results displayed in Tables S5-S7 in Multimedia Appendix 1 are consistent with our original findings that female physicians receive lower ratings than their male counterparts.

In our next robustness check, we conducted our main regression analysis without topic controls. This test was conducted to examine whether the topic variables may have introduced a systemic bias in the specifications owing to the manner in which they were constructed and whether the negative coefficient of *GenderFemale* variable may have been an artifact. As can be observed from the results in Tables S8-S10 in Multimedia Appendix 1, the coefficient of *GenderFemale* was negative and statistically significant across the specifications, even after topic controls were excluded. This further supports our main finding that female physicians tend to receive worse web-based reviews than their male counterparts. The topic controls play an important role in our specifications because they help to explain part of the variance in the web-based ratings. This can be further understood by comparing the overall R-squared, within R-squared, and between R-squared values in Tables 2, 3, and 4 with those in Tables S8, S9, and S10 in Multimedia Appendix 1, respectively. The 3 R-squared values were substantially higher in Tables 2-4, which means that the topic controls explained a considerable part of the variance in the web-based rating variables.

In summary, we conducted three additional robustness checks as explained above: (1) included additional control variables, (2) removed the specialties that did not include physicians of both genders, and (3) removed the topic controls. After conducting these robustness checks, we can conclude that female physicians tend to receive worse web-based reviews than their male counterparts. This finding is consistent across the regression specifications used in this study.

A concern could be about how representative the data in our panel are of the original data collected from RateMDs and Medicare (CMS). To address this concern, we calculated the descriptive statistics of the variables shown in Table 1 using the original longitudinal data collected from RateMDs and Medicare. The descriptive statistics are shown in Table S11 in Multimedia Appendix 1. A comparison of the statistical values in Table S11 in Multimedia Appendix 1 shows that the panel data used for the econometric analysis in our study are fairly representative of the original data collected from the 2 aforementioned sources.

Discussion

Overview

Our study provides an important contribution to the growing literature on web-based physician reviews and physician gender. A possible concern could be that the differences observed in the reviews between physicians of different genders could be driven by the differences in the quality of care or outcomes delivered by physicians of different genders. To address this concern, we performed a substantial search of the existing literature examining the differences between the quality of clinical care or outcomes delivered by male and female physicians. We found several research papers in this context [61-65], but we could not find significant evidence from extant research that male physicians deliver better care than female physicians.

Principal Findings

We found that male physicians receive better web-based reviews than female physicians after controlling for their clinical characteristics such as specialty and patient risk. Although the difference between the web-based ratings for male and female physicians was statistically significant, the average magnitude of the difference was not substantial. Our findings support that of Dunivin et al [32] and Thawani et al [42], but do not support the findings of Emmert and Meier [43], who found that during the examined time frame, female physicians had better reviews than male physicians. Their results indicated a slight but statistically significant preference for female physicians (2% differential in the percentage of reviews below the mean for each gender) compared with our results that found a 0.2 differential on a 5-point scale in favor of male physicians (4% difference). Possible reasons for these differences could be attributed to cultural variations between the patient populations in Alabama and Germany and that the reviews collected by Emmert and Meier [43] included more female respondents than male respondents. It is also possible that the relationship between patients and their physicians were not in favor of male physicians in Germany, and temporal shifts in patient-physician relationships over the time frames examined could also impact the result differences (2012 vs 2012-2018 in our data).

Implications

Our findings have important implications for health care researchers, professionals, and policy makers. First, the empirical evidence of web-based reviews is less favorable toward female physicians, after accounting or controlling for several clinical aspects (including specialty and Medicare patient risk), and temporal effects should inform health care professionals and policy makers that patients' opinions are consistently more favorable toward male physicians than toward female physicians. This cannot be overlooked even though the magnitude of the effect of gender on web-based reviews is not sizable.

Policy and Design Suggestions

Gender bias in reviews has been reported across multiple domains, including academia. Murray et al [66] found that male faculty tended to receive higher ratings for overall teaching quality than female faculty, and Turrentine et al [67] and Rojek et al [68] found implicit bias in the narrative evaluations, with a bias toward men receiving more superlative praise. Studies have shown that measures can be taken to help reduce gender bias in reviews. Peterson et al [69] found that simply informing students of potential gender biases can have significant effects on the evaluation of female instructors, and Rivera and Tilcsik [70] found that by changing the rating scale from a 10-point to a 6-point rating system, gender bias can be reduced.

Large societal-level aspects may also be in effect; however, that would seemingly be very hard to account for within a single portal. Sprague and Massoni [71] found that male teachers are more likely to be held to an entertainer standard, whereas female teachers are held to a nurturer standard. These biases are formed throughout an individual's life, and therefore, are harder to adjust for, even when directly informing users of the potential for bias. By leveraging the lessons learned from gender bias studies, web-based physician review sites could help to mitigate, but not eliminate, gender bias within their systems.

Concentrated efforts to educate and inform patients about female physicians' competence are needed. This can help to reduce implicit bias among patients toward the competence of female physicians compared with their male counterparts. These websites serve as an important resource for both reviewers and readers of the reviews, and the information needs to flow well. At the same time, readers of the reviews may be served better if the reviewers are asked to provide opinions about physicians of different genders before they provide a review for a physician. To solicit reviewers' predisposed opinions about physicians of different genders, the questions can be framed in a manner that does not make the reviewers feel that they are being investigated for their opinions. After collecting their opinions on this issue, the websites may consider filtering the reviews provided by reviewers with an overt bias against physicians of one gender. The question of how to design the website to reduce the possible gender bias is complex and requires serious thought and consideration from both researchers and website designers. By leveraging previous research efforts targeted at informing users of bias potential, review portals can better collect and present information about physicians.

Limitations

Our study has a few limitations. First, we constructed our patient risk scores using the HCC risk score from Medicare data. Although Medicare is among the largest health care payers or insurers in the United States, further studies can attempt to validate the findings of our study using clinical data from other insurers. For instance, a significant proportion of the patient population in the United States has insurance from private insurers. Future studies can attempt to validate our findings by constructing clinical variables, such as risk scores, using clinical data from one or more private insurers. Second, we focused on the physician data from Alabama. Although it is 1 state, it provides a good mix of rural and urban counties. Future studies could extend this work to other states and compare the findings across a broader set of patients and health care providers.

Future Studies

The findings of this study suggest that gender bias in web-based reviews needs to be examined more closely. Additional studies that identify factors impacting this gender bias could help us develop strategies to mitigate gender bias in web-based reviews. Given the shortage of health care providers and the need for a robust and diverse health care workforce, such studies can help not only the service providers but also policy makers, educators, and administrators. If the administrators of hospitals and clinics are made aware of this bias and acknowledge it accordingly, institutional changes can be implemented to support and empower women to take up more leadership roles in clinical settings. As Sandberg [72] points out in her New York Times best seller, as fewer women are in leadership roles than men, it can be challenging for junior women to have mentorship opportunities. A possible solution to this problem could be the performance evaluations of male leadership personnel to include the number of women mentored and focused initiatives and

incentive opportunities for women to take on pathways to leadership roles.

These focused efforts can provide a strong signal to patients about the competence of female physicians and, in turn, increase their confidence in the care provided by female physicians. This can further help to improve the overall care delivered to patients, as the increase in patients' confidence can improve their communication with physicians, irrespective of the physicians' gender. However, an open research question is whether the bias observed in web-based physician reviews is also observable in offline physician surveys. To examine this question, studies that compare reviews of male and female physicians in web-based and offline media need to be conducted.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Tables depicting the results of additional analysis including robustness checks.
[DOCX File , 39 KB - jmir_v24i4e31659_app1.docx]

References

1. RateMDs. URL: https://www.ratemds.com [accessed 2021-06-28]
2. Vitals. URL: https://www.vitals.com/ [accessed 2021-06-28]
3. Healthgrades. URL: https://www.healthgrades.com/ [accessed 2021-06-28]
4. Terlutter R, Bidmon S, Röttl J. Who uses physician-rating websites? Differences in sociodemographic variables, psychographic variables, and health status of users and nonusers of physician-rating websites. J Med Internet Res 2014 Mar 31;16(3):e97. [doi: 10.2196/jmir.3145] [Medline: 24686918]
5. Binary Fountain. 2018. URL: https://www.binaryfountain.com/wp-content/uploads/2018/09/Binary-Fountain-2018-Consumerism-Survey-Infographic.pdf [accessed 2021-06-24]
6. Detz A, López A, Sarkar U. Long-term doctor-patient relationships: patient perspective from online reviews. J Med Internet Res 2013 Jul 02;15(7):e131 [FREE Full text] [doi: 10.2196/jmir.2552] [Medline: 23819959]
7. Emmert M, Halling F, Meier F. Evaluations of dentists on a German physician rating website: an analysis of the ratings. J Med Internet Res 2015 Jan 12;17(1):e15 [FREE Full text] [doi: 10.2196/jmir.3830] [Medline: 25582914]
8. Gao GG, McCullough JS, Agarwal R, Jha AK. A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. J Med Internet Res 2012 Feb 24;14(1):e38 [FREE Full text] [doi: 10.2196/jmir.2003] [Medline: 22366336]
9. Hao H, Zhang K. The voice of Chinese health consumers: a text mining approach to web-based physician reviews. J Med Internet Res 2016 May 10;18(5):e108 [FREE Full text] [doi: 10.2196/jmir.4430] [Medline: 27165558]
10. Hong YA, Liang C, Radcliff TA, Wigfall LT, Street RL. What do patients say about doctors online? A systematic review of studies on patient online reviews. J Med Internet Res 2019 Apr 08;21(4):e12521 [FREE Full text] [doi: 10.2196/12521] [Medline: 30958276]
11. Kadry B, Chu LF, Kadry B, Gammas D, Macario A. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. J Med Internet Res 2011 Nov 16;13(4):e95 [FREE Full text] [doi: 10.2196/jmir.1960] [Medline: 22088924]
12. Liu JJ, Matelski JJ, Bell CM. Scope, breadth, and differences in online physician ratings related to geography, specialty, and year: observational retrospective study. J Med Internet Res 2018 Mar 07;20(3):e76 [FREE Full text] [doi: 10.2196/jmir.7475] [Medline: 29514775]
13. Segal J, Sacopulos M, Sheets V, Thurston I, Brooks K, Puccia R. Online doctor reviews: do they track surgeon volume, a proxy for quality of care? J Med Internet Res 2012 Apr 10;14(2):e50 [FREE Full text] [doi: 10.2196/jmir.2005] [Medline: 22491423]
14. Huerta TR, Hefner JL, Ford EW, McAlearney AS, Menachemi N. Hospital website rankings in the United States: expanding benchmarks and standards for effective consumer engagement. J Med Internet Res 2014 Feb 25;16(2):e64 [FREE Full text] [doi: 10.2196/jmir.3054] [Medline: 24568892]
15. Emmert M, Sander U, Pisch F. Eight questions about physician-rating websites: a systematic review. J Med Internet Res 2013 Feb 01;15(2):e24 [FREE Full text] [doi: 10.2196/jmir.2360] [Medline: 23372115]
16. Emmert M, Meszmer N, Sander U. Do health care providers use online patient ratings to improve the quality of care? Results from an online-based cross-sectional study. J Med Internet Res 2016 Sep 19;18(9):e254 [FREE Full text] [doi: 10.2196/jmir.5889] [Medline: 27644135]
17. Murphy GP, Awad MA, Osterberg EC, Gaither TW, Chumnarnsongkhroh T, Washington SL, et al. Web-based physician ratings for California physicians on probation. J Med Internet Res 2017 Aug 22;19(8):e254 [FREE Full text] [doi: 10.2196/jmir.7488] [Medline: 28830852]

18. Lu SF, Rui H. Can we trust online physician ratings? Evidence from cardiac surgeons in Florida. Manag Sci 2018 Jun;64(6):2557-2573. [doi: 10.1287/mnsc.2017.2741]
19. Okike K, Peter-Bibb TK, Xie KC, Okike ON. Association between physician online rating and quality of care. J Med Internet Res 2016 Dec 13;18(12):e324 [FREE Full text] [doi: 10.2196/jmir.6612] [Medline: 27965191]
20. Saifee DH, Bardhan IR, Lahiri A, Zheng ZE. Adherence to clinical guidelines, electronic health record use, and online reviews. J Manag Inf Syst 2019 Oct 09;36(4):1071-1104. [doi: 10.1080/07421222.2019.1661093]
21. Saifee DH, Zheng ZE, Bardhan IR, Lahiri A. Are online reviews of physicians reliable indicators of clinical outcomes? A focus on chronic disease management. Inf Syst Res 2020 Dec;31(4):1282-1300. [doi: 10.1287/isre.2020.0945]
22. Emmert M, Meier F, Pisch F, Sander U. Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. J Med Internet Res 2013 Aug 28;15(8):e187 [FREE Full text] [doi: 10.2196/jmir.2702] [Medline: 23985220]
23. Grabner-Kräuter S, Waiguny MK. Insights into the impact of online physician reviews on patients' decision making: randomized experiment. J Med Internet Res 2015 Apr 09;17(4):e93 [FREE Full text] [doi: 10.2196/jmir.3991] [Medline: 25862516]
24. Han X, Qu J, Zhang T. Exploring the impact of review valence, disease risk, and trust on patient choice based on online physician reviews. Telemat Inform 2019 Dec;45:101276. [doi: 10.1016/j.tele.2019.101276]
25. Li S, Lee-Won RJ, McKnight J. Effects of online physician reviews and physician gender on perceptions of physician skills and primary care physician (PCP) selection. Health Commun 2019 Oct;34(11):1250-1258. [doi: 10.1080/10410236.2018.1475192] [Medline: 29792519]
26. Brohan M. Even with a referral, 80% of patients still check out doctors online. Digital Commerce 360. 2018. URL: https://www.digitalcommerce360.com/2018/05/22/even-with-a-referral-80-of-patients-still-check-out-doctors-online/ [accessed 2022-04-04]
27. Murray E, Lo B, Pollack L, Donelan K, Catania J, Lee K, et al. The impact of health information on the internet on health care and the physician-patient relationship: national U.S. survey among 1.050 U.S. physicians. J Med Internet Res 2003;5(3):e17 [FREE Full text] [doi: 10.2196/jmir.5.3.e17] [Medline: 14517108]
28. Roter DL, Hall JA, Aoki Y. Physician gender effects in medical communication: a meta-analytic review. JAMA 2002 Aug 14;288(6):756-764. [doi: 10.1001/jama.288.6.756] [Medline: 12169083]
29. Shin DW, Roter DL, Roh YK, Hahm SK, Cho B, Park HK, Board Certification Committee of The Korean Academy of Family Medicine. Physician gender and patient centered communication: the moderating effect of psychosocial and biomedical case characteristics. Patient Educ Couns 2015 Jan;98(1):55-60. [doi: 10.1016/j.pec.2014.10.008] [Medline: 25457177]
30. Nicolai J, Demmel R. The impact of gender stereotypes on the evaluation of general practitioners' communication skills: an experimental study using transcripts of physician-patient encounters. Patient Educ Couns 2007 Dec;69(1-3):200-205. [doi: 10.1016/j.pec.2007.08.013] [Medline: 17942269]
31. Hall JA, Roter DL, Blanch-Hartigan D, Mast MS, Pitegoff CA. How patient-centered do female physicians need to be? Analogue patients' satisfaction with male and female physicians' identical behaviors. Health Commun 2015;30(9):894-900. [doi: 10.1080/10410236.2014.900892] [Medline: 25175277]
32. Dunivin Z, Zadunayski L, Baskota U, Siek K, Mankoff J. Gender, soft skills, and patient experience in online physician reviews: a large-scale text analysis. J Med Internet Res 2020 Jul 30;22(7):e14455 [FREE Full text] [doi: 10.2196/14455] [Medline: 32729844]
33. Weisman CS, Teitelbaum MA. Physician gender and the physician-patient relationship: recent evidence and relevant questions. Soc Sci Med 1985;20(11):1119-1127. [doi: 10.1016/0277-9536(85)90189-3] [Medline: 3895448]
34. Roter DL, Hall JA. Physician gender and patient-centered communication: a critical review of empirical research. Annu Rev Public Health 2004;25:497-519. [doi: 10.1146/annurev.publhealth.25.101802.123134] [Medline: 15015932]
35. Hall JA, Blanch-Hartigan D, Roter DL. Patients' satisfaction with male versus female physicians: a meta-analysis. Med Care 2011 Jul;49(7):611-617. [doi: 10.1097/MLR.0b013e318213c03f] [Medline: 21478782]
36. Hall JA, Roter DL. Do patients talk differently to male and female physicians? A meta-analytic review. Patient Educ Couns 2002 Dec;48(3):217-224. [doi: 10.1016/s0738-3991(02)00174-x] [Medline: 12477606]
37. Sandhu H, Adams A, Singleton L, Clark-Carter D, Kidd J. The impact of gender dyads on doctor-patient communication: a systematic review. Patient Educ Couns 2009 Sep;76(3):348-355. [doi: 10.1016/j.pec.2009.07.010] [Medline: 19647969]
38. Fennema K, Meyer DL, Owen N. Sex of physician: patients' preferences and stereotypes. J Fam Pract 1990 Apr;30(4):441-446. [Medline: 2324696]
39. Kerssens JJ, Bensing JM, Andela MG. Patient preference for genders of health professionals. Soc Sci Med 1997 May;44(10):1531-1540. [doi: 10.1016/s0277-9536(96)00272-9] [Medline: 9160442]
40. Plunkett BA, Kohli P, Milad MP. The importance of physician gender in the selection of an obstetrician or a gynecologist. Am J Obstet Gynecol 2002 May;186(5):926-928. [doi: 10.1067/mob.2002.123401] [Medline: 12015516]
41. Hasebrook J, Hahnenkamp K, Buhre WF, de Korte-de Boer D, Hamaekers AE, Metelmann B, et al. Medicine goes female: protocol for improving career options of females and working conditions for researching physicians in clinical medical

research by organizational transformation and participatory design. JMIR Res Protoc 2017 Aug 02;6(8):e152 [FREE Full text] [doi: 10.2196/resprot.7632] [Medline: 28768613]
42. Thawani A, Paul MJ, Sarkar U, Wallace BC. Are online reviews of physicians biased against female providers? In: Proceedings of the 4th Machine Learning for Healthcare Conference. 2019 Presented at: PMLR '19; August 8-10, 2019; Ann Arbor, MI, USA p. 406-423.
43. Emmert M, Meier F. An analysis of online evaluations on a physician rating website: evidence from a German public reporting instrument. J Med Internet Res 2013 Aug 06;15(8):e157 [FREE Full text] [doi: 10.2196/jmir.2655] [Medline: 23919987]
44. Marrero K, King E, Fingeret AL. Impact of surgeon gender on online physician reviews. J Surg Res 2020 Jan;245:510-515. [doi: 10.1016/j.jss.2019.07.047] [Medline: 31446193]
45. Centers for Medicare & Medicaid Services. 2020. URL: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Downloads/Medicare-Physician-and-Other-Supplier-PUF-Methodology.pdf [accessed 2022-04-04]
46. Hierarchical condition category coding. American Academy of Family Physicians. 2017. URL: https://www.aafp.org/family-physician/practice-and-career/getting-paid/coding/hierarchical-condition-category.html [accessed 2021-06-24]
47. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res 2003;3:993-1022.
48. Schweinberger M. Topic modeling with R. Language Technology and Data Analysis Laboratory. 2021. URL: https://slcladal.github.io/topicmodels.html [accessed 2021-06-05]
49. Guo Y, Barnes SJ, Jia Q. Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent dirichlet allocation. Tourism Management 2017 Apr;59(C):467-483. [doi: 10.1016/j.tourman.2016.09.009]
50. Hao H, Zhang K, Wang W, Gao G. A tale of two countries: international comparison of online doctor reviews between China and the United States. Int J Med Inform 2017 Mar;99:37-44. [doi: 10.1016/j.ijmedinf.2016.12.007] [Medline: 28118920]
51. Tirunillai S, Tellis GJ. Mining marketing meaning from online chatter: strategic brand analysis of big data using latent dirichlet allocation. J Mark Res 2014 Aug 01;51(4):463-479. [doi: 10.1509/jmr.12.0106]
52. Wallace BC, Paul MJ, Sarkar U, Trikalinos TA, Dredze M. A large-scale quantitative analysis of latent factors and sentiment in online doctor reviews. J Am Med Inform Assoc 2014;21(6):1098-1103 [FREE Full text] [doi: 10.1136/amiajnl-2014-002711] [Medline: 24918109]
53. Zhang Y, Chen M, Huang D, Wu D, Li Y. iDoctor: personalized and professionalized medical recommendations based on hybrid matrix factorization. Future Gener Comput Syst 2017 Jan;66:30-35. [doi: 10.1016/j.future.2015.12.001]
54. Feinerer I. Introduction to the tm package text mining in R. Microsoft R Application Network. 2013. URL: https://mran.microsoft.com/snapshot/2017-12-04/web/packages/tm/vignettes/tm.pdf [accessed 2022-04-04]
55. Feinerer I, Hornik K. Package 'tm'. The Comprehensive R Archive Network. 2020. URL: http://cran.rapporter.net/web/packages/tm/tm.pdf [accessed 2022-04-04]
56. Grün B, Hornik K. topicmodels: an R package for fitting topic models. J Stat Soft 2011;40(13):1-30. [doi: 10.18637/jss.v040.i13]
57. Amado A, Cortez P, Rita P, Moro S. Research trends on big data in marketing: a text mining and topic modeling based literature analysis. Eur Res Manag Bus Econ 2018 Jan;24(1):1-7. [doi: 10.1016/j.iedeen.2017.06.002]
58. Calheiros AC, Moro S, Rita P. Sentiment classification of consumer-generated online reviews using topic modeling. J Hosp Mark Manag 2017 Apr 27;26(7):675-693. [doi: 10.1080/19368623.2017.1310075]
59. Vacher JP, Chave J, Ficetola FG, Sommeria‐Klein G, Tao S, Thébaud C, et al. Large‐scale DNA‐based survey of frogs in Amazonia suggests a vast underestimation of species richness and endemism. J Biogeogr 2020 May 18;47(8):1781-1791. [doi: 10.1111/jbi.13847]
60. Alabama Physician Workforce Profile. Association of American Medical Colleges. 2019. URL: https://www.aamc.org/media/37726/download [accessed 2021-06-24]
61. Berthold HK, Gouni-Berthold I, Bestehorn KP, Böhm M, Krone W. Physician gender is associated with the quality of type 2 diabetes care. J Intern Med 2008 Oct;264(4):340-350 [FREE Full text] [doi: 10.1111/j.1365-2796.2008.01967.x] [Medline: 18397244]
62. Dahrouge S, Seale E, Hogg W, Russell G, Younger J, Muggah E, et al. A comprehensive assessment of family physician gender and quality of care: a cross-sectional analysis in Ontario, Canada. Med Care 2016 Mar;54(3):277-286. [doi: 10.1097/MLR.0000000000000480] [Medline: 26765146]
63. Meier A, Yang J, Liu J, Beitler JR, Tu XM, Owens RL, et al. Female physician leadership during cardiopulmonary resuscitation is associated with improved patient outcomes. Crit Care Med 2019 Jan;47(1):e8-13 [FREE Full text] [doi: 10.1097/CCM.0000000000003464] [Medline: 30303843]
64. Tsugawa Y, Jena AB, Figueroa JF, Orav EJ, Blumenthal DM, Jha AK. Comparison of hospital mortality and readmission rates for Medicare patients treated by male vs female physicians. JAMA Intern Med 2017 Feb 01;177(2):206-213 [FREE Full text] [doi: 10.1001/jamainternmed.2016.7875] [Medline: 27992617]

65. Wallis CJ, Ravi B, Coburn N, Nam RK, Detsky AS, Satkunasivam R. Comparison of postoperative outcomes among patients treated by male and female surgeons: a population based matched cohort study. BMJ 2017 Oct 10;359:j4366 [FREE Full text] [doi: 10.1136/bmj.j4366] [Medline: 29018008]
66. Murray D, Boothby C, Zhao H, Minik V, Bérubé N, Larivière V, et al. Exploring the personal and professional factors associated with student evaluations of tenure-track faculty. PLoS One 2020;15(6):e0233515 [FREE Full text] [doi: 10.1371/journal.pone.0233515] [Medline: 32492028]
67. Turrentine FE, Dreisbach CN, St Ivany AR, Hanks JB, Schroen AT. Influence of gender on surgical residency applicants' recommendation letters. J Am Coll Surg 2019 Apr;228(4):356-65.e3. [doi: 10.1016/j.jamcollsurg.2018.12.020] [Medline: 30630084]
68. Rojek AE, Khanna R, Yim JW, Gardner R, Lisker S, Hauer KE, et al. Differences in narrative language in evaluations of medical students by gender and under-represented minority status. J Gen Intern Med 2019 May;34(5):684-691 [FREE Full text] [doi: 10.1007/s11606-019-04889-9] [Medline: 30993609]
69. Peterson DA, Biederman LA, Andersen D, Ditonto TM, Roe K. Mitigating gender bias in student evaluations of teaching. PLoS One 2019;14(5):e0216241 [FREE Full text] [doi: 10.1371/journal.pone.0216241] [Medline: 31091292]
70. Rivera LA, Tilcsik A. Scaling down inequality: rating scales, gender bias, and the architecture of evaluation. Am Sociol Rev 2019 Mar 12;84(2):248-274. [doi: 10.1177/0003122419833601]
71. Sprague J, Massoni K. Student evaluations and gendered expectations: what we can't count can hurt us. Sex Roles 2005 Dec;53(11-12):779-793. [doi: 10.1007/s11199-005-8292-4]
72. Sandberg S, Scovell N. Lean in: women, work, and the will to lead. New York, NY, USA: Random House; 2013.

Abbreviations

CMS: Center for Medicare and Medicaid Services
HCC: hierarchical condition category
LDA: Latent Dirichlet Allocation

Edited by A Mavragani; submitted 29.06.21; peer-reviewed by B Metelmann, R Krukowski, D Verran, M Salimi; comments to author 18.07.21; revised version received 02.09.21; accepted 16.03.22; published 08.04.22.

Please cite as:
Saifee DH, Hudnall M, Raja U
Physician Gender, Patient Risk, and Web-Based Reviews: Longitudinal Study of the Relationship Between Physicians' Gender and Their Web-Based Reviews
J Med Internet Res 2022;24(4):e31659
URL: https://www.jmir.org/2022/4/e31659
doi:10.2196/31659
PMID:35394435

Original Paper

Cost-Effectiveness of Mobile Health–Based Integrated Care for Atrial Fibrillation: Model Development and Data Analysis

Xueyan Luo[1,2,3], MSc; Wei Xu[4], PhD; Wai-Kit Ming[5], MD, PhD; Xinchan Jiang[6], PhD; Quan Yuan[7], PhD; Han Lai[1], MSc; Chunji Huang[8], PhD; Xiaoni Zhong[1,3], MSc

[1]School of Public Health and Management, Chongqing Medical University, Chongqing, China
[2]School of Biological and Chemical Engineering, Chongqing University of Education, Chongqing, China
[3]Research Center for Medicine and Social Development, Chongqing Medical University, Chongqing, China
[4]School of International Pharmaceutical Business, China Pharmaceutical University, Nanjin, China
[5]Department of Infectious Diseases and Public Health, City University of Hong Kong, Hong Kong, Hong Kong
[6]School of Pharmacy, The Chinese University of Hong Kong, Hong Kong, Hong Kong
[7]Chong Qing Pharmaceutical Group Co Ltd, Chongqing, China
[8]School of Basic Medical Science, Army Medical University, Chongqing, China

Corresponding Author:
Xiaoni Zhong, MSc
School of Public Health and Management
Chongqing Medical University
1 Yi Xue Lu
Yu Zhong District
Chongqing, 400016
China
Phone: 86 (023)68485008
Email: zhongxiaoni@cqmu.edu.cn

Abstract

Background: Mobile health (mHealth) technology is increasingly used in disease management. Using mHealth tools to integrate and streamline care has improved clinical outcomes of patients with atrial fibrillation (AF).

Objective: The aim of this study was to investigate the potential clinical and health economic outcomes of mHealth-based integrated care for AF from the perspective of a public health care provider in China.

Methods: A Markov model was designed to compare outcomes of mHealth-based care and usual care in a hypothetical cohort of patients with AF in China. The time horizon was 30 years with monthly cycles. Model outcomes measured were direct medical cost, quality-adjusted life years (QALYs), and incremental cost-effectiveness ratio (ICER). Sensitivity analyses were performed to examine the robustness of the base-case results.

Results: In the base-case analysis, mHealth-based care gained higher QALYs of 0.0730 with an incurred cost of US $1090. Using US $33,438 per QALY (three times the gross domestic product) as the willingness-to-pay threshold, mHealth-based care was cost-effective, with an ICER of US $14,936 per QALY. In one-way sensitivity analysis, no influential factor with a threshold value was identified. In probabilistic sensitivity analysis, mHealth-based care was accepted as cost-effective in 92.33% of 10,000 iterations.

Conclusions: This study assessed the expected cost-effectiveness of applying mHealth-based integrated care for AF according to a model-based health economic evaluation. The exploration suggested the potential cost-effective use of mHealth apps in streamlining and integrating care via the Atrial fibrillation Better Care (ABC) pathway for AF in China. Future economic evaluation alongside randomized clinical trials is highly warranted to verify the suggestion and investigate affecting factors such as geographical variations in patient characteristics, identification of subgroups, and constraints on local implementation.

(J Med Internet Res 2022;24(4):e29408) doi:10.2196/29408

KEYWORDS

mobile health; integrated care; ABC pathway; atrial fibrillation; model-based; cost-effectiveness; health economic evaluation

Introduction

Atrial fibrillation (AF) is the most frequent cardiac rhythm disorder. Approximately 2% of the population is affected by AF in European and North American countries, with the prevalence varying from 0.1% to 7.2% across different regions [1,2]. AF is less common in China than in Western countries, with a prevalence of 0.71% among adults above 35 years old [3]. Yet, the prevalence is higher than previously reported (0.65% for the population above 30 years) a decade ago, which is likely due to low awareness and huge treatment gaps [3-6]. According to a recent nationwide survey, the estimated number of patients with AF in China was 7.9 million [5]. For adults above 55 years of age, the lifetime risk of developing AF was approximately 1 in 5.3 [7]. Driven by a growing and aging population, the prevalence is predicted to at least double in the next 30 years [8].

The increasing prevalence of AF is associated with more health care utilization and health care expenditure. Stroke is the most common subsequent outcome, and patients with AF have a 5-fold increased risk of stroke [9]. Stroke-related care is costly, with incurred hospitalization costs of US $3000 to US $10,000 per patient in China [10]. Stroke patients with AF were more likely to have comorbidities, and their stroke-related costs were 20% more costly than those of stroke patients without AF [11,12]. Stroke prevention, proactive management of comorbidities, and lifestyle changes are essential priorities in AF care. Therefore, the Atrial fibrillation Better Care (ABC) pathway, a holistic and integrated approach, has been proposed to monitor anticoagulant therapy and manage the cardiovascular risks of patients with AF [13].

Adoption of the ABC pathway was found to be effective in reducing clinical adverse events and related health care costs in the Atherosclerosis in Atrial Fibrillation trial [14]. Another study applying the ABC pathway in the mobile health (mHealth) context also supported the favored effectiveness of integrated care for AF [15]. In this study, 1261 subjects who received mHealth technology–supported care were followed up over 1 year and had a lower risk of composite outcomes of "ischemic stroke (IS)/systemic thromboembolism, death, and rehospitalization" compared with that of their counterparts receiving usual care. mHealth-based care has been demonstrated to be cost-effective in managing diabetes, hypertension, and heart failure; however, the health economic impact of mHealth for patients with AF remains unknown [16-18].

A Markov model is a well-established analytic framework in the economic evaluation of health care interventions by using mutually exclusive disease states to represent all possible consequences. The purpose of this study was to perform a cost-effectiveness analysis via a Markov model to examine the clinical and health economic outcomes of mHealth-based integrated care for patients with AF in China.

Methods

Model Structure

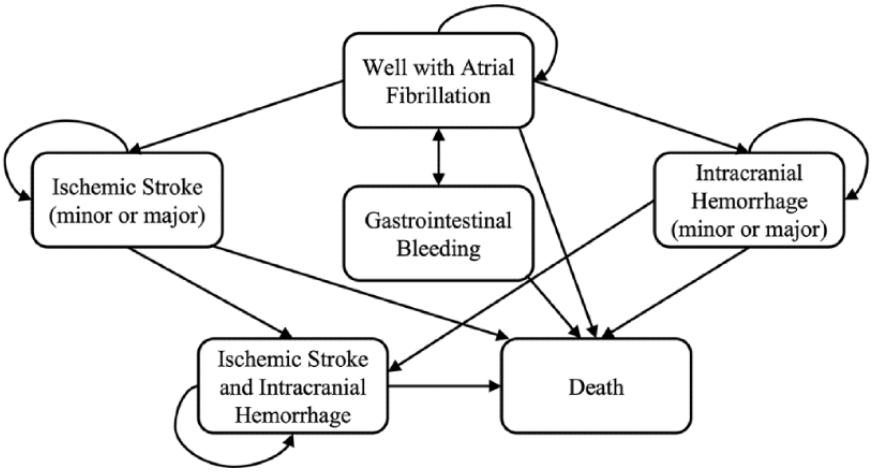
A Markov model was developed to evaluate the cost-effectiveness of mHealth-based care integrating the ABC pathway for patients with AF from the perspective of the public health care provider in China (Figure 1). The cycle length for the model was monthly cycles with a 30-year time frame to estimate the long-term effects. The baseline model population consisted of patients with AF with a mean age of 68 years and a median CHA_2DS_2-VASc (congestive heart failure, hypertension, age≥75 years [doubled], diabetes, stroke/transient ischemic attack/thromboembolism [doubled], vascular disease [prior myocardial infarction, peripheral artery disease, or aortic plaque], age 65 years, sex category [female]) score of 3 [15]. The patient characteristics were derived from the Mobile Atrial Fibrillation App (mAFA)-II trial, a cluster-randomized trial examining the first mHealth technology–based program for following patients with AF based on the ABC pathway in China. The clinical effectiveness from the mAFA-II trial was applied as the key model input after entering a hypothetical cohort of AF patients to the model. The model structure was adapted from the study published by Shah et al [19], where a Markov model was developed to model the prognosis of patients with AF using similar events of interest regarding stroke adopted in this study. Specifically, the strategies examined were mHealth-based care and usual care for patients with AF. The outcome measures were direct medical cost, quality-adjusted life years (QALYs), and incremental cost-effectiveness ratio (ICER).

The model consisted of the following Markov states: well with AF, minor and major IS, minor and major intracranial hemorrhage (ICH), IS and ICH, and death, with a temporary health state of gastrointestinal bleeding (GIB). With either mHealth-based care or usual care, all patients entered the model at the health state of being well and transitioned to another health state in the next cycle. The events of IS and ICH could each be of two types: minor and major. Once ICH or GIB occurred, patients would discontinue the anticoagulant therapy and switch to aspirin for the remaining life years. After stroke, patients might experience recurrent events. They might stay in the same health state or proceed to the health state of "IS and ICH." Consistent with Shah et al [19], we assumed patients would advance to corresponding major events after two minor neurological events and that two major events would lead to death. Patients, in whichever state, could proceed to death. Patients in the arm of mHealth-based care might not be adherent to the mobile technology and would receive the same intervention as patients with usual care. Once the event occurred, these patients would receive mHealth-based care again in the next cycle.

Patients with usual care would receive treatment following the Chinese Stroke Association guideline [20]. Patients with mHealth-based care would receive integrated management based on the ABC pathway. The ABC pathway components consisted of Avoid Stroke with Anticoagulation (A), Better Symptom Management (B), and Cardiovascular Risk and Comorbidity Management (C). Specifically, patients allocated to the

mHealth-based care would install a mobile app connected to a local public hospital via the internet. For Avoid Stroke with Anticoagulation (A), the laboratory results (eg, international normalized ratio, renal/hepatic function) tested in hospitals could be uploaded in the mobile app. The validated algorithm with confirmation from doctors would provide patients with data on anticoagulant monitoring, bleeding risk assessment, and guideline-based dosage adjustment. For Better Symptom Management (B), patients would receive a photoplethysmography smart device connected to the mobile app. They could send the cardiac rhythm monitoring data along with other symptoms such as headache and chest pain to the on-call doctors by the in-built communication function. Advice on rate or rhythm control would be given in a timely manner. Once a patient's condition deteriorates, the management would be escalated to inpatient care. For Cardiovascular Risk and Comorbidity Management (C), patients' comorbidities (eg, blood pressure) would be monitored with treatment optimized for blood pressure <140/85 mmHg. Lifestyle recommendations would also be given by educational articles, videos, and game-playing in the mobile app.

Figure 1. Schematic representation of the Markov model.



Clinical Probabilities

All model inputs are listed in Table 1. The clinical inputs were retrieved from the published reports written in English, identified from a literature search on Medline throughout 2000-2021. Epidemiology or disease burden in the Chinese population, randomized clinical trials, and meta-analyses were preferred sources for clinical model inputs.

The probabilities of clinical events (IS, ICH, and GIB) after AF in usual care, the effectiveness (measured in hazard ratios) of mHealth-based care (vs usual care), and the compliance of mHealth support were retrieved from the mAFA-II trial (N=2473 patients) [15]. In the trial, a structured program of holistic and integrated care based on the ABC pathway via a mobile app was compared to usual care in patients with AF. The 18-month incidence of IS (4.12%), ICH (0.41%), and GIB (0.58%) in the usual-care arm was converted into the monthly probability (0.244%, 0.024%, and 0.034%, respectively) using the equation $p=1–e^{–rt}$ (where p is probability, r is the event rate, and t is the cycle length) from the practical guide for Markov models [21]. The hazard ratio (mHealth-based care vs usual care) was 0.11 (95% CI 0.05-0.27, *P*<.001) for IS and 0.37 (95% CI 0.20-0.70, *P*=.002) for GIB [15]. No event of ICH was reported in the mHealth-based care group over 1-year follow-up. Considering the feasibility in long-term practice, the change of mHealth-based care on ICH incidence was assumed to be 0.5 in the base case, with a range of 0-1. The incidence of IS, ICH, and GIB for patients managed via mHealth technology was approximated by the incidence in usual care and the corresponding hazard ratio, as recommended by the Guide to the Methods of Technology Appraisals 2013 [22]. The proportion of minor, major, and fatal IS/ICH was estimated at 51.6%/49.5%, 40.2%/14.1%, and 8.2%/36.4%, respectively, from a study comparing outcomes of five oral anticoagulants for stroke prevention [19]. The incidences of recurrence and proportion of types of events (ICH or IS) were estimated from a 9-year community-based study of 0.5 million Chinese adults assessing the recurrent events after the first incident stroke [23]. The reported 5-year recurrence was 41% for IS (91% IS and 9% ICH) and 44% for ICH (44% IS and 56% ICH). Using the equation $p=1–e^{–rt}$, the monthly probability was approximated to be 0.68% for IS and 0.73% for ICH [22]. The age-specific mortality rates were retrieved from a nationwide survey conducted in China [24]. The relative risk of death (with vs without GIB) was 3.5, with a range of 2.8 to 4.2 [25]. The mortality of GIB was calculated using age-specific mortality and the relative risk [22]. A population-based prospective study in an elderly (≥60 years) Chinese population identified an increased risk of all-cause mortality for AF, with a hazard ratio of 1.87 (95% CI 1.09-3.20, *P*=.02) [26]. The mortality of AF was estimated by the hazard ratio and age-specific mortality [22].

Table 1. Model inputs of clinical probabilities, utilities, and costs.

Variables	Base-case input (range)	Distribution	Reference
Clinical variables			
Probability of an event in usual care (monthly)			
IS[a]	0.244 (0.20-0.29)	Beta	Guo et al [15]
ICH[b]	0.024 (0.02-0.03)	Beta	Guo et al [15]
GIB[c]	0.034 (0.03-0.04)	Beta	Guo et al [15]
Hazard ratio of events (mobile health–based care vs usual care)			
IS	0.11 (0.05-0.27)	Lognormal	Guo et al [15]
ICH	0.5 (0-1)	Triangular	Guo et al [15], assumption
GIB	0.37 (0.2-0.7)	Lognormal	Guo et al [15]
Compliance of mobile health–based case, %	70.8 (50-100)	Beta	Guo et al [15]
Proportion of events, %			
IS minor	51.6 (43.9-55.8)	Dirichlet	Shah et al [19]
IS major	40.2 (40.2-41.7)	Dirichlet	Shah et al [19]
IS fatal	8.2 (2.5-16.3)	Dirichlet	Shah et al [19]
ICH minor	49.5 (33-63)	Dirichlet	Shah et al [19]
ICH major	14.1 (9-21.4)	Dirichlet	Shah et al [19]
ICH fatal	36.4 (15.6-58.0)	Dirichlet	Shah et al [19]
Probability of stroke recurrence (monthly)			
IS	0.68 (0.68-0.70)	Beta	Chen et al [23]
ICH	0.73 (0.70-0.76)	Beta	Chen et al [23]
Proportion of recurrent events, %			
IS after IS	91 (73-100)	Beta	Chen et al [23]
ICH after IS	9 (0-27)	Beta	Chen et al [23]
IS after ICH	44 (33-55)	Beta	Chen et al [23]
ICH after ICH	56 (45-67)	Beta	Chen et al [23]
Age-specific (years) mortality (monthly), %			
65-69	0.10 (0.08-0.12)	Triangular	National Bureau of Statistics [24]
70-74	0.26 (0.20-0.31)	Triangular	National Bureau of Statistics [24]
75-79	0.41 (0.33-0.50)	Triangular	National Bureau of Statistics [24]
80-84	0.71 (0.57-0.85)	Triangular	National Bureau of Statistics [24]
85-89	1.06 (0.85-1.27)	Triangular	National Bureau of Statistics [24]
90-94	1.59 (1.27-1.91)	Triangular	National Bureau of Statistics [24]
>95	1.81 (1.45-2.17)	Triangular	National Bureau of Statistics [24]
Utilities			
Event-free AF[d]	0.9 (0.8-1)	Uniform	Shah et al [19]
Minor IS	0.75 (0.6-0.92)	Uniform	Shah et al [19]
Major IS	0.39 (0.31-0.47)	Uniform	Shah et al [19]
Minor ICH	0.75 (0.6-0.92)	Uniform	Shah et al [19]
Major ICH	0.39 (0.31-0.47)	Uniform	Shah et al [19]
Utility decrement of GIB	0.16 (0.13-0.19)	Uniform	Shah et al [19]
Costs (US $)			

Variables	Base-case input (range)	Distribution	Reference
Event-related costs (per episode)			
Minor IS	3277 (2622-3932)	Lognormal	Chang et al [10]
Major IS	6676 (5341-8012)	Lognormal	Chang et al [10]
Minor ICH	5284 (4227-6340)	Lognormal	Chang et al [10]
Major ICH	10567 (8454-12,680)	Lognormal	Chang et al [10]
GIB	3443 (2754-4131)	Lognormal	Chang et al [10]
All-cause death	5849 (4679-7019)	Lognormal	Chang et al [10]
Follow-up cost (per month)			
Anticoagulation therapy	249 (199-299)	Lognormal	MENET [27]
Minor IS	328 (262-393)	Lognormal	Experts' opinion
Major IS	668 (534-801)	Lognormal	Experts' opinion
Minor ICH	528 (422-634)	Lognormal	Experts' opinion
Major ICH	1057 (845-1268)	Lognormal	Experts' opinion
Cost of site implementation per patient (one-time cost)	80 (64-96)	Lognormal	Boodoo et al [28]
Cost of managing per month	15 (12-18)	Lognormal	Zhang and Liu [29]

[a]IS: ischemic stroke.

[b]ICH: intracranial hemorrhage.

[c]GIB: gastrointestinal bleeding.

[d]AF: atrial fibrillation.

Utility and Utility Adjustment

Literature-based utilities were assigned to each health state to calculate QALYs (Table 1) [19]. Major or minor neurological events were associated with permanent disutility. Temporary disutility was applied for GIB with a duration of 14 days. The expected QALYs were estimated by the utility in each state and the time spent. The QALYs gained were discounted at 3.5% per annum, as recommended by the Guide to the Methods of Technology Appraisals 2013 [22].

Resource Use and Costs

All costs were considered from a public health care provider's perspective in China and only direct medical costs were included (Table 1). All cost inputs were retrieved from public data. The assumption (if necessary) was made in consultation with local experts, which was considered a legitimate source of information for decision-analytic modeling [30]. The event costs of IS (minor or major), ICH (minor or major), and GIB were estimated from the in-hospital direct costs for thromboembolism and bleeding in Chinese patients with AF, which were collected from seven representative tertiary referral hospitals and three secondary-care hospitals [10]. The cost of all-cause death was approximated by the mean of the event-related cost. The cost of medication was estimated by the frequency and the median cost of anticoagulants in the Menet database [27]. Since limited data were available on the cost of postevent follow-up, the cost was assumed to be 1/10 the event cost based on local cardiac specialists' advice. The one-off charge for site implementation, including a wristband-type wireless photoplethysmographic device, was estimated from the reported cost of a smartphone-based system for heart failure [28]. We assumed that the clinical doctor in charge would spend 1 hour monitoring a patient every month. The monthly cost for subscribing to the mHealth-based service was estimated from physicians' hourly rate and the time spent on the service [29]. All costs were adjusted to the year 2021, with an annual discount rate of 3.5%, based on the recommendations from the Guide to the Methods of Technology Appraisals 2013 [22].

Analytic Methods

All model parameters were used to generate the cohort model. Model validation was performed by comparing the estimated incidence of events to the reported outcomes in the mAFA-II trial and comparing the simulated 5-year survival rate to that reported for the Chinese elderly population with AF [31]. The direct medical costs and QALYs of each comparator were calculated over a 30-year time horizon. ICERs were estimated and compared against the willingness-to-pay (WTP) threshold. The WTP threshold was defined as three times the gross domestic product per capita in China, according to the World Health Organization recommendation [32]. The gross domestic product per capita was US $11,146 (US $1=RMB 6.5); thus, the WTP threshold was US $33,438 per QALY [33].

One-way sensitivity analysis was performed to assess the robustness of the base-case results. The literature-available ranges were adopted (Table 1). Otherwise, ±20% of the base-case values were used to examine the impact of parameters on the ICER. Parameter uncertainty was determined using 10,000 Monte Carlo simulations by varying all inputs simultaneously with random draws from each specified distribution. The results of probabilistic sensitivity analysis are

presented in a scatterplot in the form of incremental costs against incremental QALYs. The probability of each strategy to be preferred was determined in the cost-effectiveness acceptability curve over US $0-50,000 per QALY. All analyses were performed using Excel 2016 software (Microsoft Corp).

Results

Model Validation

To examine the predictive validity of the model, the simulated event rates (IS, ICH, and GIB) in usual care and mHealth-based care were compared to the findings reported in the mAFA-II trial [15]. As shown in Table 2, all simulated events in both arms were within 10% of relative difference when compared with the reported data from the mAFA-II trial. The simulated 5-year survival rate determined by the model (73.5%) was also compared to that reported (68.9%) for the Chinese elderly population with AF (CHA_2DS_2-VASc score of 3), and the relative difference was found to be 6.67% [31].

Table 2. Model validation.

Variable	Usual care (median follow-up 546 days)			mHealth[a]-based care (median follow-up 701 days)		
	Trial	Model	Difference	Trial	Model	Difference
IS[b]	4.12%	3.86%	–6.31%	0.48%	0.43%	–10.42%
ICH[c]	0.41%	0.40%	–2.44%	—[d]	—	—
GIB[e]	0.58%	0.55%	–5.17%	0.40%	0.39%	–2.50%

[a]mHealth: mobile health.

[b]IS: ischemic stroke.

[c]ICH: intracranial hemorrhage.

[d]The incidence of ICH in the trial was reported to be 0 within follow-up. The model simulated the long-term impacts and assumed that the relative risk of mHealth-based care (vs usual care) regarding ICH was 0.5, with a range of 0-1. Therefore, the difference in the ICH incidence is not presented.

[e]GIB: gastrointestinal bleeding.

Base-Case Analysis

Over a 30-year time horizon, the total costs of mHealth-based care and usual care were US $35,691 and US $34,601, respectively. The expected QALY gain was 7.2749 for mHealth-based care and 7.2019 for usual care. Compared with usual care, mHealth-based care gained additional QALYs of 0.0730 with an incurred cost of US $1090. The ICER was US $14,936 per QALY, which was below the WTP threshold of US $33,438 per QALY. These results indicated mHealth-based care as a cost-effective strategy in the base-case analysis.

Sensitivity Analyses

The ICERs of mHealth-based care were all below the WTP threshold throughout the one-way variation. No parameter with a threshold value was found (Figure 2). The analysis showed that the model results were most sensitive to compliance to mHealth-based care, the monthly cost of follow-up after major IS, utility after major IS, the proportion of recurrent IS after IS, the hazard ratio of all-cause mortality (AF vs no AF), and utility of AF. As the compliance of mHealth-based care had the most significant impact on the ICER, an extended one-way sensitivity analysis was performed. Once the probability of compliance to the mobile technology exceeded 99%, mHealth-based care was a cost-saving option with QALYs gained when compared with usual care.

The Monto Carlo simulations of incremental costs versus incremental QALYs gained by mHealth-based care are shown in Figure 3. The mHealth-based care gained average QALYs of 0.0842 (95% CI 0.0832-0.0851, $P<.001$), with a mean additional cost of US $1053 (95% CI US $1033-1073, $P<.001$). Of 10,000 iterations, mHealth-based care gained higher QALYs at a higher cost with the ICER below the WTP threshold 89.44% of the time. The probability of mHealth-based care being effective in QALYs gain with cost savings was 2.89%.

The probability of each comparator being preferred as cost-effective is shown in the acceptability curve of Figure 4. The mHealth-based care and usual care shared the same probability (50%) of being cost-effective at the WTP threshold of US $10,699 per QALY. mHealth-based care was accepted to be cost-effective 92.33% of the time at the WTP threshold of US $33,438 per QALY.

Figure 2. Tornado diagram of one-way sensitivity analysis summarizing the effect of parameters on the ICER. ICER: incremental cost-effectiveness ratio; IS: ischemic stroke; AF: atrial fibrillation.

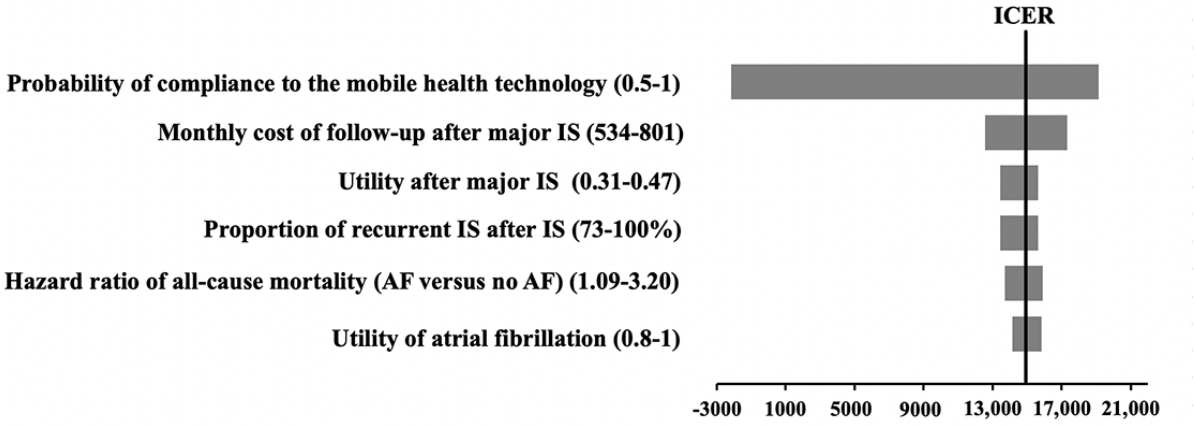


Figure 3. Incremental cost-effectiveness scatterplot: probabilistic sensitivity analysis for mobile health–based care versus usual care. QALYs: quality-adjusted life-years.

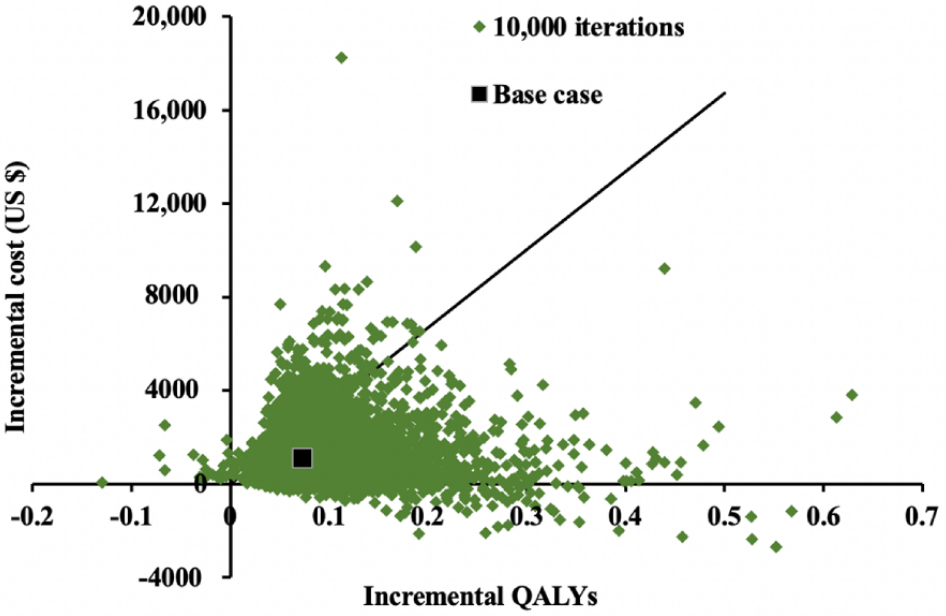
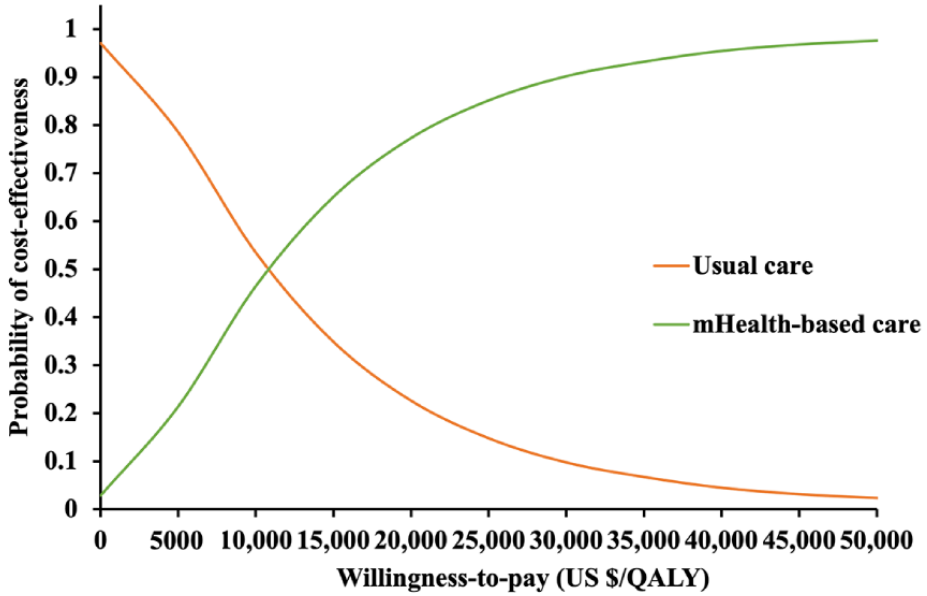


Figure 4. Cost-effectiveness acceptability curve. mHealth: mobile health. QALY: quality-adjusted life year.



Discussion

Principal Results

This is the first cost-effectiveness analysis examining mHealth-based integrated care using the ABC pathway to manage patients with AF. Compared to usual care, mHealth-based care was cost-effective from the public health care provider in China, with an ICER (US $14,936 per QALY) below the WTP threshold (US $33,438 per QALY). No parameter varying the ICER with a threshold value was found in one-way sensitivity analysis, indicating the robustness of the base-case results. In probabilistic sensitivity analysis, the probability of mHealth-based care being preferred was high throughout the WTP threshold variation, which further supported the cost-effective application of managing patients with AF via mHealth technology. To our best knowledge, no cost-effectiveness analysis has previously been performed to examine the application of mHealth technology in care after an AF diagnosis, although such an analysis has been performed for screening [34]. Our findings are in line with the results of previous cost-effectiveness analyses investigating mHealth tools for other cardiovascular diseases, indicating the cost-effective use of the mHealth support system [16-18]. The small improvement in QALYs by the mHealth-based care for AF, similar to other digital health technologies, may be driven by a small, estimated gain in survival and reflect the indirect effects of the technologies on mortality.

A recent community-based multicenter study investigating the prevalence of untreated AF found a noticeable treatment gap in urban China. Only 20.3% (28/138) of patients with AF qualifying for guideline-recommended anticoagulant therapy commenced the treatment [6]. The undertreated AF resulted from patients' preference to attend local community centers over specialist clinics, community health center physicians' lack of knowledge regarding evidence-based management (only antiplatelet drugs and traditional Chinese medicine prescribed), and specialists' low adherence to AF guidelines. The condition is likely to be worse in rural China owing to the lower awareness in rural residents than in urban areas [6]. To optimize the management of patients with AF, a mobile technology–supported program adopting the ABC pathway was initiated in China [15,35]. The program encompassed guideline-adherent recommendation and monitoring on anticoagulants, patient-centered symptom-directed decisions for rate or rhythm control, and comorbidity management. The results showed that mHealth-based care was associated with improved patient outcomes: better rate/rhythm control; increased use of anticoagulants; and reduced composite outcomes of IS/systemic thromboembolism, death, and rehospitalization.

As China's population is aging, there is a shift in the disease burden to chronic noncommunicable diseases. The Chinese government is actively seeking ways to reduce the health care expenditure on chronic diseases. The delivery of quality health services in a cost-effective manner is the key direction [36]. In this landscape, mHealth apps offer unique opportunities for improving the quality of care while reducing the cost of care by outsourcing the proactive patient monitoring to a clinically validated algorithm, enabling early diagnosis and intervention to patients, and saving clinicians' time for more urgent cases. Due to a lack of economic evidence, the innovative use of mobile technology has not yet been adequately integrated into the health care system. Therefore, cost-effective assessment of mHealth-based care is the essential process of considering this new technology in China. Our study demonstrated that using mobile technology to streamline and integrate care for patients with AF is likely to be cost-effective from the perspective of public health care providers in China. A recent study suggested patients' problems of seeking routine care under the context of the COVID-19 pandemic and increasing use of internet-based medical services in China [37,38]. In this regard, internet-based interventions with cost-effectiveness, such as mHealth-based care for AF patients, should be considered part of routine care for chronic diseases.

mHealth-based care involves patients' engagement more than conventional care. Patients' compliance and persistence are significant barriers and challenges for advancing these new technologies [39]. The realization of the benefits of mHealth technology will only occur when high compliance is achieved. Key factors to improve patients' compliance encompass user training, active human support, and telehealth implementation style [40]. To examine the compliance on the findings, an extended one-way sensitivity analysis varying the parameter from 0%-100% (base-case value 70.8%) was performed. Once full adherence (≥99%) is achieved, the mHealth-based care would be a cost-saving option with QALY gains. Compliance is a nontransferrable parameter among different health care systems. Local patients' willingness to uptake and adhere to the technology should be evaluated in a pilot study before incorporating mHealth support into AF care.

Limitations

The results of this model should be interpreted while considering the following limitations. First, the model was developed based on a cluster-randomized clinical trial (mAFA-II trial), which studied an adult population diagnosed with AF who were followed up over 1 year. The baseline characteristics were adopted for the model population [15]. In the trial, the intervention group consisted of 1261 subjects (mean age 67 years, median CHA_2DS_2-VASc score 3, 34.1% women) and 1212 subjects (mean age 70 years, median CHA_2DS_2-VASc score 3, 42.1% women). A community-based survey of 47,841 adults (aged ≥45 years) in seven geographic regions of China showed the characteristic of confirmed AF patients (mean age 67.6 years, CHA_2DS_2-VASc score of 2.13 for men; mean age 66.6 years, CHA_2DS_2-VASc score of 3.08 for women), which indicated the similarity of Chinese AF patients and the model population [5]. However, fewer women were identified in the mAFA-II trial than reported in the community-based survey (54.7%). This likely resulted from the selection bias with more male, younger subjects included during AF screening phases, which limited the generalizability of the results [15].

Second, the effectiveness estimates, both probabilities and hazard ratios, were approximated from a follow-up much shorter than the lifetime horizon and used to examine the long-term effects of mHealth technology. To examine the model's

predictive validity, the simulated events were compared to rates reported in the trial, and the 5-year survival rate was also compared to that reported for the Chinese elderly population with AF. The model development, data conversion, and approximation followed the practical guide for Markov models and the guide to the methods of technology appraisal 2013 [21,22]. The model demonstrated accuracy with acceptable differences between simulated and reported rates. The higher 5-year survival rate generated in the present model was likely due to the model population being less elderly and having fewer comorbidities compared to patient characteristics in the cohort study [31]. For a more precise estimation of the cost-effectiveness over a lifetime scale, a clinical trial investigating the long-term effectiveness of mHealth-based care is highly warranted.

Third, this is a model-based health economic evaluation using model inputs from multiple sources with similar patient characteristics as the model population. Nevertheless, the data availability partially limited the data retrieval and the generalizability of the results to a large population with different region-specific patient characteristics in China. The variation was considered in one-way and probabilistic sensitivity analyses using the reported ranges or 20% of the base-case values. No threshold value was identified in one-way sensitivity analysis, and a strong likelihood of being cost-effective in probabilistic sensitivity analysis supported the applicability of the results. Future studies should investigate the use of mHealth technology for AF in randomized clinical trials considering geographical variations in patient characteristics, identification of subgroups, and constraints on local implementation, along with a trial-based economic evaluation assessing the incremental cost-effectiveness of mHealth tools.

Fourth, the cost inputs were retrieved from multiple sources according to data availability, including expert opinions on the monthly cost of postevent follow-up due to limited data. No identified threshold value of cost inputs indicated the robustness of the base-case result. Future studies investigating the cost of postevent follow-up in Chinese patients with stroke are warranted.

Conclusion

This study assessed the expected cost-effectiveness of applying mHealth-based integrated care for AF by a model-based health economic evaluation. The exploration suggested the potential cost-effective use of mHealth apps in streamlining and integrating care via the ABC pathway for AF in China. Future economic evaluation alongside randomized clinical trials is highly warranted to verify the suggestion and investigate affecting factors such as geographical variations in patient characteristics, identification of subgroups, and constraints on local implementation.

Acknowledgments

We acknowledge the financial support received from the National Natural Science Foundation of China (72074220), Humanities and Social Science Research Program of Chongqing Municipal Education Commission (21SKGH283), Research Program of Innovative Research Demonstration Base of Children's Medical Security (NCRCCHD-2019-HP-12), and Scientific and Technological Research Program of Chongqing Municipal Education Commission (KJ1901535).

Conflicts of Interest

None declared.

References

1. Zoni-Berisso M, Lercari F, Carazza T, Domenicucci S. Epidemiology of atrial fibrillation: European perspective. Clin Epidemiol 2014;6:213-220. [doi: 10.2147/CLEP.S47385] [Medline: 24966695]
2. Pistoia F, Sacco S, Tiseo C, Degan D, Ornello R, Carolei A. The epidemiology of atrial fibrillation and stroke. Cardiol Clin 2016 May;34(2):255-268. [doi: 10.1016/j.ccl.2015.12.002] [Medline: 27150174]
3. Wang Z, Chen Z, Wang X, Zhang L, Li S, Tian Y, for China Hypertension Survey Group. The disease burden of atrial fibrillation in China from a national cross-sectional survey. Am J Cardiol 2018 Sep 01;122(5):793-798. [doi: 10.1016/j.amjcard.2018.05.015] [Medline: 30049467]
4. Zhou Z, Hu D. An epidemiological study on the prevalence of atrial fibrillation in the Chinese population of mainland China. J Epidemiol 2008;18(5):209-216 [FREE Full text] [doi: 10.2188/jea.je2008021] [Medline: 18776706]
5. Du X, Guo L, Xia S, Du J, Anderson C, Arima H, et al. Atrial fibrillation prevalence, awareness and management in a nationwide survey of adults in China. Heart 2021 Jan 28;107(7):535-541 [FREE Full text] [doi: 10.1136/heartjnl-2020-317915] [Medline: 33509976]
6. Chen Y, Huang Q, Sheng C, Zhang W, Shao S, Wang D, et al. Detection rate and treatment gap for atrial fibrillation identified through screening in community health centers in China (AF-CATCH): A prospective multicenter study. PLoS Med 2020 Jul 16;17(7):e1003146 [FREE Full text] [doi: 10.1371/journal.pmed.1003146] [Medline: 32673305]
7. Guo Y, Tian Y, Wang H, Si Q, Wang Y, Lip GY. Prevalence, incidence, and lifetime risk of atrial fibrillation in China: new insights into the global burden of atrial fibrillation. Chest 2015 Jan;147(1):109-119. [doi: 10.1378/chest.14-0321] [Medline: 24921459]

8. Go AS, Hylek EM, Phillips KA, Chang Y, Henault LE, Selby JV, et al. Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study. JAMA 2001 May 09;285(18):2370-2375. [doi: 10.1001/jama.285.18.2370] [Medline: 11343485]
9. Hart RG, Pearce LA, Aguilar MI. Meta-analysis: antithrombotic therapy to prevent stroke in patients who have nonvalvular atrial fibrillation. Ann Intern Med 2007 Jun 19;146(12):857-867. [doi: 10.7326/0003-4819-146-12-200706190-00007] [Medline: 17577005]
10. Chang S, Wu J, Liu Y, Zhang T, Du X, Dong J, et al. In-hospital direct costs for thromboembolism and bleeding in Chinese patients with atrial fibrillation. Chronic Dis Transl Med 2018 Jun;4(2):127-134 [FREE Full text] [doi: 10.1016/j.cdtm.2018.01.001] [Medline: 29988956]
11. Wen L, Wu J, Feng L, Yang L, Qian F. Comparing the economic burden of ischemic stroke patients with and without atrial fibrillation: a retrospective study in Beijing, China. Curr Med Res Opin 2017 Oct 09;33(10):1789-1794. [doi: 10.1080/03007995.2017.1348345] [Medline: 28657348]
12. Sussman M, Menzin J, Lin I, Kwong WJ, Munsell M, Friedman M, et al. Impact of Atrial Fibrillation on Stroke - Related Healthcare Costs. JAHA 2013 Nov 18;2(6):1-7. [doi: 10.1161/jaha.113.000479]
13. Lip GYH. The ABC pathway: an integrated approach to improve AF management. Nat Rev Cardiol 2017 Nov 29;14(11):627-628. [doi: 10.1038/nrcardio.2017.153] [Medline: 28960189]
14. Pastori D, Farcomeni A, Pignatelli P, Violi F, Lip GY. ABC (Atrial fibrillation Better Care) pathway and healthcare costs in atrial fibrillation: The ATHERO-AF Study. Am J Med 2019 Jul;132(7):856-861. [doi: 10.1016/j.amjmed.2019.01.003] [Medline: 30659810]
15. Guo Y, Guo J, Shi X, Yao Y, Sun Y, Xia Y, mAF-App II Trial investigators. Mobile health technology-supported atrial fibrillation screening and integrated care: A report from the mAFA-II trial Long-term Extension Cohort. Eur J Intern Med 2020 Dec;82:105-111 [FREE Full text] [doi: 10.1016/j.ejim.2020.09.024] [Medline: 33067121]
16. Rinaldi G, Hijazi A, Haghparast-Bidgoli H. Cost and cost-effectiveness of mHealth interventions for the prevention and control of type 2 diabetes mellitus: A systematic review. Diabetes Res Clin Pract 2020 Apr;162:108084. [doi: 10.1016/j.diabres.2020.108084] [Medline: 32061819]
17. Piera-Jiménez J, Winters M, Broers E, Valero-Bover D, Habibovic M, Widdershoven JWMG, et al. Changing the health behavior of patients with cardiovascular disease through an electronic health intervention in three different countries: cost-effectiveness study in the Do Cardiac Health: Advanced New Generation Ecosystem (Do CHANGE) 2 randomized controlled trial. J Med Internet Res 2020 Jul 28;22(7):e17351 [FREE Full text] [doi: 10.2196/17351] [Medline: 32720908]
18. Jiang X, Yao J, You JH. Cost-effectiveness of a telemonitoring program for patients with heart failure during the COVID-19 pandemic in Hong Kong: model development and data analysis. J Med Internet Res 2021 Mar 03;23(3):e26516 [FREE Full text] [doi: 10.2196/26516] [Medline: 33656440]
19. Shah A, Shewale A, Hayes CJ, Martin BC. Cost-effectiveness of oral anticoagulants for ischemic stroke prophylaxis among nonvalvular atrial fibrillation patients. Stroke 2016 Jun;47(6):1555-1561. [doi: 10.1161/strokeaha.115.012325]
20. Cao Y, Yu S, Zhang Q, Yu T, Liu Y, Sun Z, Chinese Stroke Association Stroke Council Guideline. Chinese Stroke Association guidelines for clinical management of cerebrovascular disorders: executive summary and 2019 update of clinical management of intracerebral haemorrhage. Stroke Vasc Neurol 2020 Dec 01;5(4):396-402 [FREE Full text] [doi: 10.1136/svn-2020-000433] [Medline: 33262243]
21. Sonnenberg FA, Beck JR. Markov models in medical decision making. Med Decis Making 2016 Jul 02;13(4):322-338. [doi: 10.1177/0272989x9301300409]
22. Guide to the methods of technology appraisal. National Institute for Health and Care Excellence. 2013 Apr 04. URL: https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781 [accessed 2021-08-08]
23. Chen Y, Wright N, Guo Y, Turnbull I, Kartsonaki C, Yang L, China Kadoorie Biobank Collaborative Group. Mortality and recurrent vascular events after first incident stroke: a 9-year community-based study of 0·5 million Chinese adults. Lancet Glob Health 2020 Apr;8(4):e580-e590 [FREE Full text] [doi: 10.1016/S2214-109X(20)30069-3] [Medline: 32199124]
24. The 2010 Population Census of The People's Republic of China. National Bureau of Statistics. 2021. URL: http://www.stats.gov.cn/tjsj/pcsj/rkpc/6rp/indexch.htm [accessed 2021-04-05]
25. Sun X, He S, Lin Z, Zhang L, Wang Y, Zeng L, et al. Radiofrequency catheter ablation versus cryoballoon ablation in the treatment of paroxysmal atrial fibrillation: a cost-effectiveness analysis in China. Clin Ther 2019 Jan;41(1):78-91. [doi: 10.1016/j.clinthera.2018.11.008] [Medline: 30559005]
26. Li L, Sheng C, Hu B, Huang Q, Zeng W, Li G, et al. The prevalence, incidence, management and risks of atrial fibrillation in an elderly Chinese population: a prospective study. BMC Cardiovasc Disord 2015 May 08;15(1):31-37 [FREE Full text] [doi: 10.1186/s12872-015-0023-3] [Medline: 25953603]
27. The bid price of drugs. Menet. URL: https://shuju.menet.com.cn/NEWVIPZone/index.jsp [accessed 2021-04-18]
28. Boodoo C, Zhang Q, Ross HJ, Alba AC, Laporte A, Seto E. Evaluation of a heart failure telemonitoring program through a microsimulation model: cost-utility analysis. J Med Internet Res 2020 Oct 06;22(10):e18917 [FREE Full text] [doi: 10.2196/18917] [Medline: 33021485]

29. Zhang C, Liu Y. The salary of physicians in Chinese public tertiary hospitals: a national cross-sectional and follow-up study. BMC Health Serv Res 2018 Aug 24;18(1):661-669 [FREE Full text] [doi: 10.1186/s12913-018-3461-7] [Medline: 30143042]
30. Leal J, Wordsworth S, Legood R, Blair E. Eliciting expert opinion for economic models: an applied example. Value Health 2007 May;10(3):195-203 [FREE Full text] [doi: 10.1111/j.1524-4733.2007.00169.x] [Medline: 17532812]
31. Wu Y, Wang G, Dong L, Qin L, Li J, Yan H, et al. Assessment of the CHADS-VASc score for the prediction of death in elderly patients with coronary artery disease and atrial fibrillation. Front Cardiovasc Med 2021 Dec 24;8:805234. [doi: 10.3389/fcvm.2021.805234] [Medline: 35004915]
32. The World Health Report 2002. World Health Organization. 2002. URL: https://www.who.int/whr/2002/en/whr02_en.pdf?ua=1 [accessed 2020-11-24]
33. Data on Gross Domestic Product. National Bureau of Statistics. URL: https://data.stats.gov.cn/search.htm?s=%E4%BA%BA%E5%9D%87GDP [accessed 2021-03-12]
34. Giebel GD. Use of mHealth devices to screen for atrial fibrillation: cost-effectiveness analysis. JMIR Mhealth Uhealth 2020 Oct 06;8(10):e20496 [FREE Full text] [doi: 10.2196/20496] [Medline: 33021489]
35. Guo Y, Lane D, Wang L, Zhang H, Wang H, Zhang W, mAF-App II Trial Investigators. Mobile health technology to improve care for patients with atrial fibrillation. J Am Coll Cardiol 2020 Apr 07;75(13):1523-1534 [FREE Full text] [doi: 10.1016/j.jacc.2020.01.052] [Medline: 32241367]
36. Huang J, Dai T. Public hospital reforms in China: the perspective of hospital directors. BMC Health Serv Res 2019 Feb 28;19(1):142-147 [FREE Full text] [doi: 10.1186/s12913-019-3954-z] [Medline: 30819157]
37. Sun S, Xie Z, Yu K, Jiang B, Zheng S, Pan X. COVID-19 and healthcare system in China: challenges and progression for a sustainable future. Global Health 2021 Jan 21;17(1):14-18 [FREE Full text] [doi: 10.1186/s12992-021-00665-9] [Medline: 33478558]
38. Xu X, Cai Y, Wu S, Guo J, Yang L, Lan J, et al. Assessment of internet hospitals in China during the COVID-19 pandemic: national cross-sectional data analysis study. J Med Internet Res 2021 Jan 20;23(1):e21825 [FREE Full text] [doi: 10.2196/21825] [Medline: 33417586]
39. Telehealth. Agency for Healthcare Research Quality. Digital Healthcare Research Archive. URL: https://digital.ahrq.gov/ahrq-funded-projects/emerging-lessons/telehealth [accessed 2021-03-20]
40. Maeder A, Poultney N, Morgan G, Lippiatt R. Patient compliance in home-based self-care telehealth projects. J Telemed Telecare 2015 Dec;21(8):439-442. [doi: 10.1177/1357633X15612382] [Medline: 26556057]

Abbreviations

ABC: Atrial fibrillation Better Care
AF: atrial fibrillation
CHA_2DS_2-VASc: congestive heart failure, hypertension, age ≥75 years (doubled), diabetes, stroke/transient ischemic attack/thromboembolism (doubled), vascular disease (prior myocardial infarction, peripheral artery disease, or aortic plaque), age 65 years, sex category (female)
GIB: gastrointestinal bleeding
ICER: incremental cost-effective ratio
ICH: intracranial hemorrhage
IS: ischemic stroke
mAFA-II: Mobile Atrial Fibrillation App trial
mHealth: mobile health
QALY: quality-adjusted life year
WTP: willingness to pay

Edited by A Mavragani; submitted 07.04.21; peer-reviewed by J Roca, F Liu, Q Zhang, R Halkes; comments to author 07.07.21; revised version received 23.09.21; accepted 24.02.22; published 19.04.22.

Please cite as:
Luo X, Xu W, Ming WK, Jiang X, Yuan Q, Lai H, Huang C, Zhong X
Cost-Effectiveness of Mobile Health–Based Integrated Care for Atrial Fibrillation: Model Development and Data Analysis
J Med Internet Res 2022;24(4):e29408
URL: https://www.jmir.org/2022/4/e29408
doi:10.2196/29408
PMID:35438646

Original Paper

The Box—eHealth in the Outpatient Clinic Follow-up of Patients With Acute Myocardial Infarction: Cost-Utility Analysis

Roderick Willem Treskes[1], MD, MSc, PhD; M Elske van den Akker-van Marle[1], MSc, PhD; Louise van Winden[1], MSc; Nicole van Keulen[1], MANP; Enno Tjeerd van der Velde[1], MSc, PhD; Saskia Beeres[1], MD, PhD; Douwe Atsma[1], MD, PhD; Martin Jan Schalij[1], MD, PhD

Department of Cardiology, Leiden University Medical Center, Leiden, Netherlands

Corresponding Author:
Martin Jan Schalij, MD, PhD
Department of Cardiology, Leiden University Medical Center
Albinusdreef 2
Leiden, 2333 ZA
Netherlands
Phone: 31 715262020
Fax: 31 715262020
Email: m.j.schalij@lumc.nl

Abstract

Background: Smartphone compatible wearables have been released on the consumers market, enabling remote monitoring. Remote monitoring is often named as a tool to reduce the cost of care.

Objective: The primary purpose of this paper is to describe a cost-utility analysis of an eHealth intervention compared to regular follow-up in patients with acute myocardial infarction (AMI).

Methods: In this trial, of which clinical results have been published previously, patients with an AMI were randomized in a 1:1 fashion between an eHealth intervention and regular follow-up. The remote monitoring intervention consisted of a blood pressure monitor, weight scale, electrocardiogram device, and step counter. Furthermore, two in-office outpatient clinic visits were replaced by e-visits. The control group received regular care. The differences in mean costs and quality of life per patient between both groups during one-year follow-up were calculated.

Results: Mean costs per patient were €2417±2043 (US $2657±2246) for the intervention and €2888±2961 (US $3175±3255) for the control group. This yielded a cost reduction of €471 (US $518) per patient. This difference was not statistically significant (95% CI –€275 to €1217; *P*=.22, US $–302 to $1338). The average quality-adjusted life years in the first year of follow-up was 0.74 for the intervention group and 0.69 for the control (difference –0.05, 95% CI –0.09 to –0.01; *P*=.01).

Conclusions: eHealth in the outpatient clinic setting for patients who suffered from AMI is likely to be cost-effective compared to regular follow-up. Further research should be done to corroborate these findings in other patient populations and different care settings.

Trial Registration: ClinicalTrials.gov NCT02976376; https://clinicaltrials.gov/ct2/show/NCT02976376

International Registered Report Identifier (IRRID): RR2-10.2196/resprot.8038

(J Med Internet Res 2022;24(4):e30236) doi:10.2196/30236

KEYWORDS

smart technology; myocardial infarction; cost-utility; outpatients; cost-effectiveness; eHealth; remote monitoring; cost of care; quality of life

Introduction

eHealth, broadly speaking, the delivery of medicine using information technology, has been suggested as a cost-saving tool to deliver health care [1,2]. It can be delivered using personal computers, mobile phones, or tablets. One advantage of delivering health care through these mobile devices is that it uses an already existing infrastructure. The vast majority of the western world population has internet access or possesses a smartphone. Recent statistics showed that 92% of the Dutch

population (aged ≥12 years) uses the internet, and 89% of the population owns a smartphone [3].

Mobile technology might be cheaper than conventional health care technology. Furthermore, an eHealth intervention can be delivered to more patients at the same time using information technology [4]. This also allows health care delivery in low- and middle-income countries. In addition, if it decreases the costs of health care delivery in high-income countries, it may increase equality.

Accordingly, smartphone-compatible devices might be clinically effective and cost-saving tools to deliver health care to acute myocardial infarction (AMI) patients. In The Box trial, a trial randomizing 200 patients to either an eHealth intervention or regular follow-up, it was found that there was no difference in clinical endpoints. A cost-effectiveness analysis of this trial was not included [5]. It is, therefore, the primary purpose of this study to describe a cost-utility analysis of an eHealth intervention (The Box) compared to regular follow-up in the outpatient care setting of patients who have been treated for AMI with primary percutaneous coronary intervention (PCI), with or without ST elevation, using data from The Box trial.

Methods

Overview

"The Box" was a single-center open-label randomized controlled trial (RCT) conducted at the Department of Cardiology of the Leiden University Medical Center (LUMC) in Leiden, the Netherlands, between May 2016 and December 2018 (NCT02976376) [5]. The current paper describes a trial-based cost-utility analysis of the intervention.

Intervention

Details about the trial protocol and the results of the clinical trial have been published previously [5,6]. In brief, patients who were admitted to the cardiac care unit (CCU) of the LUMC with an AMI, as defined by European Society of Cardiology (ESC) guidelines [7,8], were approached for participation. Both patients with ST-elevation myocardial infarction (STEMI) and patients presenting with non-ST-elevation myocardial infarction (NSTEMI) were eligible for participation. Therefore, according to the ESC guidelines, every patient admitted to the CCU, with symptoms of AMI, elevated troponin levels, and a more than 90% occlusion on coronary angiography, which was treated with primary PCI ≤48 hours after onset of symptoms, was considered for participation [9]. Patients were excluded if they were ≤18 years old, pregnant, unwilling or unable to sign the informed consent form, included in another RCT, or unable to communicate in English or Dutch at a sufficient level. After inclusion, patients were randomized to either the intervention group or the control group. When randomized to the control group, patients were followed-up according to the department's AMI follow-up protocol (called MISSION! protocol) [10]. Patients visited the outpatient clinic 1 month, 3 months, 6 months, and 12 months after they were treated for AMI. At each visit, a 12-lead electrocardiogram (ECG) was obtained, and blood pressure (BP) was measured by a nurse practitioner (NP) with ample training using a handheld sphygmomanometer. At 3 months, a stress echocardiogram was done, and a 24 Holter monitor was attached to the patient. At 6 months, a transthoracic echocardiogram (TTE) was done, and a 24 Holter monitor was performed. At 12 months, a TTE was done. Patients were not monitored in between outpatient clinic visits. When randomized to the intervention group, patients received a box containing a smartphone-compatible weight scale, a BP monitor, a step counter (all three by Nokia Health, Nokia), and an ECG device (Kardia, AliveCor Inc). Patients were asked to record their weight, BP, and ECG once daily and to record their steps taken continuously. Data were automatically transferred from the patient's smartphone to the department's dedicated hospital information system (EPD-Vision), and data were checked multiple times per week. In case of possible abnormalities (high BP, possible arrhythmias, or a sudden increase or decrease in weight), patients were contacted by a doctor or NP, and the therapeutic regimen could eventually be adjusted. Furthermore, the outpatient clinic visits 1 month and 6 months after AMI were replaced by an e-visit, in which the patient contacted the hospital via a secured video connection (Starleaf Breeze, Starleaf). The ECG at the 1- and 6-month outpatient clinic visit, as well as the TTE and the 24 hour Holter monitor at the 6-month outpatient clinic visit, were not performed in the intervention group. In case of technical difficulties, patients could contact a project dedicated PhD Student for technical support. This technical support was primarily delivered via telephone or a secured video connection. If problems could not be solved, a computer expert would visit the patient at their home.

Dutch Health Care System

Detailed information (in English) on the Dutch health care system is published elsewhere [11]. In brief, the law that covers the payment for hospital care is called the Health Care Insurance Act (in Dutch: "Zorgverzekeringswet"). The system combines aspects of private and public insurance. Health care insurers are private companies that are not-for-profit. The health care insurance act demands that healthcare insurers accept all customers, regardless of their health care condition. Insurers are furthermore forbidden to charge different premiums for the same package. Finally, insurers are obliged to make health care that is part of the government decided basic health package available to all customers. All residents of the Netherlands are obliged to take health insurance. Health care insurers negotiate prices and volumes of care with hospitals, focusing on affordability and quality of care. Hospitals either employ doctors or doctors are working on a fee-for-service basis. Patients pay a deductible of €385 (US $423) and a fixed premium price per month [11].

In this trial, costs of The Box are covered by the Department of Cardiology of the LUMC. The diagnosis-related group of a PCI in AMI is more than €385 (US $423). As such, patients had to pay their deductible. Therefore, as premium prices are fixed and do not depend on the amount of health care consumed, there was no difference in the amount of money patients had to pay for health care between patients who participated in the intervention group or the control group.

Trial Based Analysis

The trial-based cost-utility analysis was performed from a department of cardiology's perspective with a time horizon of 1 year. All costs are reported in 2020 euros. The general Dutch consumer price index was used to convert costs to 2020 price levels [12]. The analysis was performed on a modified intention-to-treat population. To create such a population, 12 (12%)patients in the intervention group and 8 (8%) patients in the control group were included in the trial but dropped out within two weeks due to various reasons and, therefore, not following the protocol as planned, were excluded from the analysis. For the base-case analysis, only health care consumption at the cardiology department (defined as the cardiac care unit, emergency room, ward, and outpatient clinic) was taken into consideration. The intervention primarily intervenes with follow-up of cardiac care and targets some specific risk factors for cardiovascular disease. It is therefore assumed there is no causal relationship between our intervention and health care utilization of other departments or outside the hospital. These costs are therefore not taken into account. All calculations were done in Excel and SPSS (version 23.0, IBM Corp) and IBM SPSS Statistics for Windows (version 22; IBM Corp).

Cost Calculations

Costs of The Box

Costs for the intervention (The Box) and the technical support of The Box were copied from bills received by the department. Costs for procedures performed as part of the study protocol (stress echocardiogram, transthoracic echocardiogram, 24-Hour Holter monitor, digital outpatient clinic visit, and in-office outpatient clinic visit) were at LUMC prices. Extra consultations for adjustment of the therapeutic regimen as a consequence of irregularities in Box data were included in the intervention group. A consult of 10 minutes was multiplied by the hourly wage of a nurse practitioner. All cost prices are given in Table 1.

Table 1. Costs of The Box, follow-up, and major adverse cardiac events.

Item	Price	Source
The Box (ECG[a] monitor, blood pressure monitor, weight scale, cardboard box, manual)	€318[b]	Bills
Follow-up		
Stress echo	€542	LUMC[c]
Echo (outpatient clinic)	€117	LUMC
24-hour Holter monitor	€152	LUMC
E-visit outpatient clinic visit	€44	LUMC
Normal outpatient clinic visits	€96	LUMC
Technical support	€1758	Bills
Contact NP[d] due to Box measurements (per contact)	€4	Dutch costing guidelines
Coronary angiogram	€2037	NZA[e]
Revascularization (elective), 1-vessel disease (1VD), with admission	€5999	NZA
Revascularization (elective), multivessel disease (MVD), with admission	€6428	NZA
Admission, unspecified (price per night)	€684	Dutch costing guidelines
In hospital technical support (including training of patients to use devices and checking of data)	€15,367	UMC[f] gross salary of 0.5 FTE[g]

[a]ECG: echocardiogram.

[b]A currency exchange rate of €1=US $1.0994 is applicable.

[c]LUMC: Leiden University Medical Center.

[d]NP: nurse practitioner.

[e]NZA: Nederlandse Zorgautoriteit (Dutch Healthcare Authority).

[f]UMC: University Medical Center.

[g]FTE: full-time equivalent.

Follow-up Hospitalization Costs

The following events were taken into account: during follow-up, cardiac care utilization due to nonfatal adverse cardiac events (defined as any hospital visit for myocardial infarction, elective revascularization, arrhythmia, or heart failure) was noted. The following events were defined: coronary angiography without intervention, revascularization (elective), recurrent STEMI, recurrent non-ST elevation acute coronary syndrome, acute heart failure, and hospital admission to the cardiac care unit or cardiology ward for other reasons than the above. Costs for revascularization were taken from the Dutch Health care Authority (in Dutch: "Nederlandse Zorgautoriteit"; NZA)

costing lists [13]. These lists distinguish four types of elective revascularization: single vessel revascularization with and without an overnight stay in the hospital and multivessel revascularization with or without an overnight stay in the hospital. Costs for revascularization with an overnight stay in the hospital were used. Costs of hospitalizations not due to revascularizations and emergency care visits were derived from the Dutch Costing Manual [14].

Costs of the Outpatient Clinic Visits

As a reference cost price of an e-visit was not available, the e-visits and in-office outpatient clinics were calculated by top-down micro-costing. In the base-case, patient-related costs were not taken into account. For the in-office outpatient clinic (ie, ECG), administrative and NP costs were taken into account. The last two were taken from the Dutch costing manual, whereas costs for the ECG were taken from the NZA list of maximum prices [14]. Hospital costs were multiplied by 1.44 (44% overhead), in accordance with the Dutch costing manual [14]. For the e-visit, administrative, video connection system, and consultation costs were taken into account. Costs of the video system were calculated by dividing the yearly subscription costs by the system's full capacity (11 patients per half-day, 110 patients a week times 50 weeks of outpatient clinic, summing up to 5500 e-visits per year). These were multiplied by 1.22 (22% overhead). A 22% overhead was assumed because of a lack of cleaning and a decrease in housing costs. The costs of an e-visit amounted to €44 (US $48), and an in-office visit cost €96 (US $106).

Quality-Adjusted Life Years

Utilities were derived from the Short Form Health Survey (SF-36) questionnaire [15]. These questionnaires were administered via the computer, smartphone, or tablet. Patients received an email with a URL to a web page where the SF-36 could be filled in digitally. During The Box trial, patients in both groups were asked to fill in the SF-36 three times: at 1 month, 6 months, and 12 months after inclusion. Results of the SF-36 were converted into health utilities (1=perfect health, 0= health as bad as dead) by using the established UK-based utility algorithm obtained through the University of Sheffield Licensing [16]. Multiple imputation was used to assess missing utility values. Baseline characteristics such as, but not limited to, age, gender, index event (STEMI vs NSTEMI), maximum troponin levels, and previous utilities were taken into account. Subsequently, quality-adjusted life years (QALYs) were calculated using the area under the curve method.

Cost-Effectiveness Acceptability Curve

Nonparametric bootstrapping was used (involving 1000 replications) to calculate uncertainty around the costs and effects estimates. Based on these results, a cost-effectiveness acceptability curve was constructed by plotting the proportion of costs and effects pairs for which the intervention is cost-effective compared to regular follow-up for a range of values of the willingness to pay for a QALY. The willingness-to-pay threshold in the Netherlands is between €20,000 (US $21,989) and €80,000 (US $87,957) per QALY [17].

Sensitivity Analysis of Patient-Related Costs

To analyze the potential effect of The Box on hospital and patient-related costs, a sensitivity analysis was performed. In this analysis, the costs of e-visits and in-office outpatient clinic visits were altered, as patient-related costs were included in the calculation. To calculate patient-related costs, the following costs were assumed for the in-office outpatient clinic visit: travel costs, parking costs, and 4.5 hours of loss of economic productivity multiplied by an hourly wage. For the digital outpatient clinic visit, half an hour of loss of economic productivity was assumed. The median age of the population was 59 years (IQR 53-66); therefore, it was assumed that 70% of the study population was still economically productive. The hourly wage of “economically productive” patients was €37.05 (US $40.76), whereas the hourly wage of “non-economically productive” patients (eg, retired patients) was €13.33 (US $14.66) [18]. The vast majority of the myocardial infarction population of the LUMC lives within 10 kilometers of the hospital. As such, an average distance of 7 kilometers for both groups was assumed, according to the Dutch Costing manual [14]. Parking costs were assumed to be €3.20 (US $3.52) for one hospital visit, again in accordance with the Dutch Costing manual [14]. Time for using The Box was multiplied by the hourly tariff of non-economically productive patients, as patients used this during non-office hours. A total of 10 minutes per week to take measurements was included. All costs taken into account are given in Table 1.

Statistical Analysis

The sample size calculation of 200 patients has been published previously. It was calculated using R statistical software version 3.2.0 for Windows (R Project for Statistical Computing). It was assumed that 95% of patients in the intervention group would have regulated BP against 75% in the control group. An α of .05, a β of .20, and a margin of 0.07 were chosen. Costs were calculated in 2020 euros and are presented as mean±SD. Differences between the intervention and control group in mean costs per category were tested for statistical significance using an independent student’s *t* test. *P* values and confidence intervals were calculated using SPSS (IBM Corp) and IBM SPSS Statistics for Windows (version 25.0; IBM Corp).

Results

Patient Population

In total, 200 patients (median age 59 years, 78% male) were included in the trial. These patients served as the base-case for the trial-based analysis. Of these patients, 100 (from now: 100% of the intervention group) were randomized to The Box, and 100 (from now: 100% of the control group) were randomized to the control group. In total, 12 patients (12%) in the intervention group and 8 (12%) in the control group were lost to follow-up and were not included in the base-case analysis. In both groups, 2 patients per group passed away (21% each). These patients were included in this cost-utility analysis. In total, the intervention group consisted of 88 (88%) patients and the control group of 92 (92%) patients. Results on primary and secondary outcomes have been published previously in detail [5].

Base-Case Analysis

Mean total costs per patient were €2417±2043 (US $2657±2246) for the intervention group and €2888±2961 (US $3175±3255) for the control group. On average, costs were €471 (US $518) lower in the intervention group. This difference was, however, not statistically significant (95% CI –€275 to €1217; *P*=.22, US $–302 to $1338). Statistical significance of differences in mean costs per item per patient are given in Table 2.

Mean utilities per randomization group were noted at 1, 6, and 12 months. Differences between utilities were not statistically significant at 1 month (–0.03; 95% CI –0.07 to 0.01; *P*=.16), 6 months (–0.05; 95% CI –0.11 to 0.001; *P*=.06), and 12 months (–0.05; 95% CI –0.11 to 0.01). When converting the utilities to QALYs, the mean QALY per patient was 0.74 for the intervention group and 0.69 for the control group. This difference was statistically significant (95% CI –0.09 to –0.01; *P*=.028). Utilities are graphically represented in Figure 1.

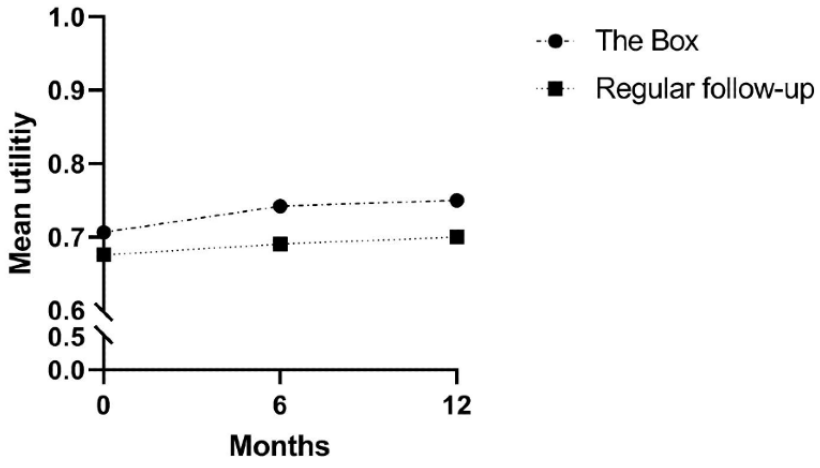
Table 2. Pooled costs per patient per item for both the intervention group and regular follow-up.

Item	Intervention group, numbers	Intervention group, costs, €μ±SD	Regular follow-up, numbers	Regular follow-up, costs, €μ±SD	Difference	95% CI	*P* value
The Box	88	€318±0[a]	0	€0±0	–€318	–€318 to –€318	<.001
E-visit	148	€34±13	0	€0±0	–€34	–€36 to –€31	<.001
In-office outpatient clinic visit	181	€146±58	373	€288±46	€142	€126 to €157	<.001
Holter	89	€196±108	171	€360±84	€164	€136 to €192	<.001
Transthoracic echocardiogram	100	€151±76	178	€256±58	€105	€86 to €126	<.001
Stress echocardio-gram	76	€468±204	85	€500±144	€32	–€19 to €84	.12
Emergency care visit	12	€38±104	25	€75±221	€37	–€13 to €88	.15
Hospitalization	18	€139±697	25	€186±984	€47	–€204 to €296	.72
Catheterization	13	€308±745	22	€499±946	€191	–€59 to €440	.13
Single vessel PCI[b]	10	€273±1256	4	€652±2077	€379	–€125 to €884	.14
Multivessel PCI	2	€146±963	1	€69±670	–€77	–€318 to €165	.54
Box support	1	€195±0	0	€0±0	–€195	–€195 to –€195	<.001

[a]A currency exchange rate of €1=US $1.0994 is applicable.

[b]PCI: percutaneous coronary intervention.

Figure 1. Mean pooled utilities per randomization group at one, six and twelve months after study inclusion.



Sensitivity Analysis

Mean patient-related costs were €426±114 (US $468±125) per patient in the intervention group, while mean patient-related costs in the control group were €570±92 (US $627±101). The difference of €144 (US $158) was statistically significant (95% CI €115 to €175; *P*<.001, US $127 to $193). In the sensitivity analysis, mean total costs per patient were €2842±2047 (US $3127±2252) for the intervention group and €3458±2974 (US $3805±3273) for the control group. This difference was not

statistically significant (95% CI –€133 to €1365; *P*=.11, US $146 to $1365).

Cost-Effectiveness

Bootstrap results of the base-case and sensitivity analysis, including patient-related costs, are presented in the cost-effectiveness planes shown in Figures 2 and 3, respectively. The cost-effectiveness acceptability curves of both the base-case and the sensitivity analysis show that the probability that the Box is cost-effective compared to usual care is very high, above 0.9 for all values of the willingness to pay for a QALY (Figure 4).

Figure 2. Scatter plot of incremental costs and incremental quality-adjusted life years in the base-case analysis. QALY: quality-adjusted life years.

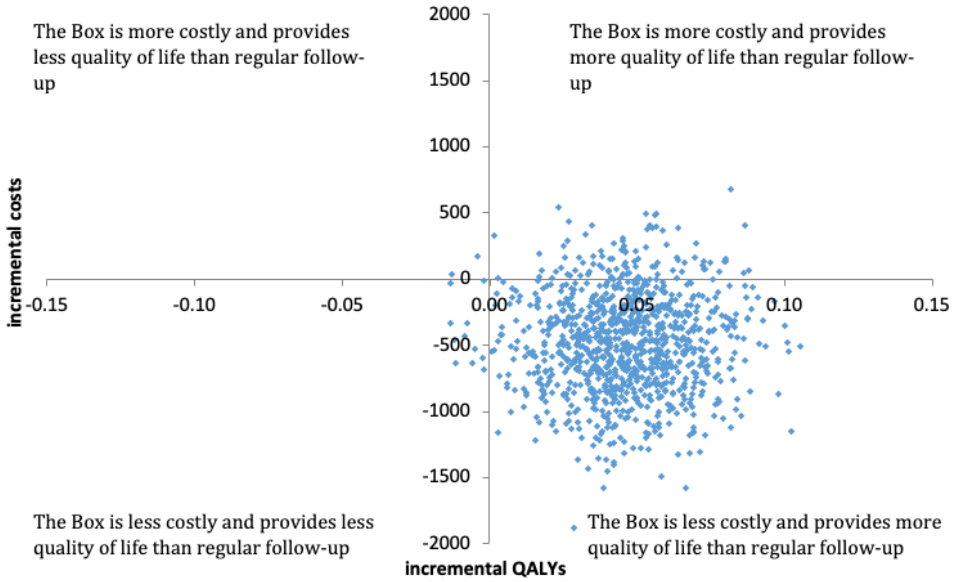


Figure 3. Scatter plot of incremental costs and incremental quality-adjusted life years in the sensitivity analysis. QALY: quality-adjusted life years.

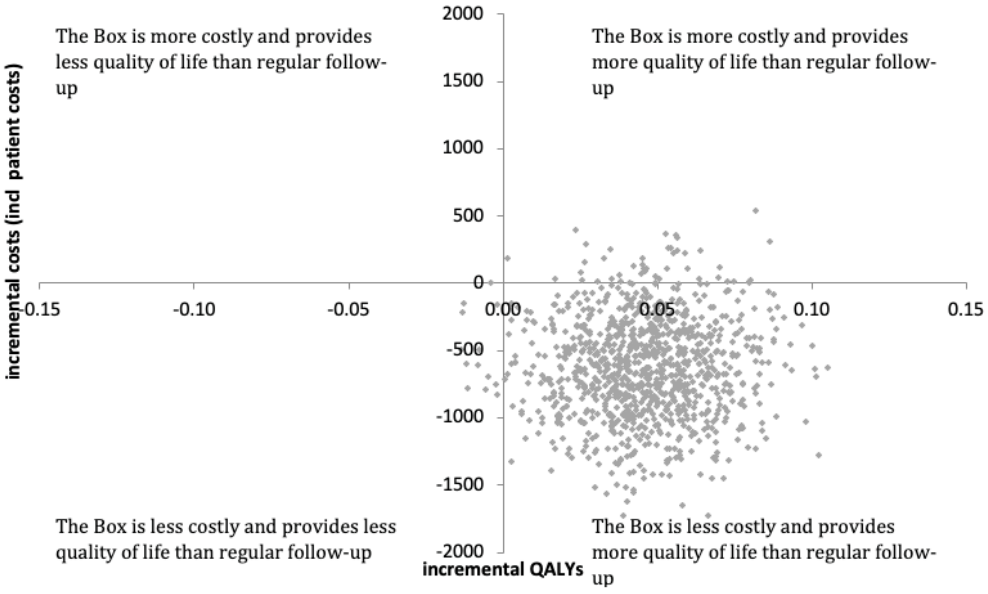
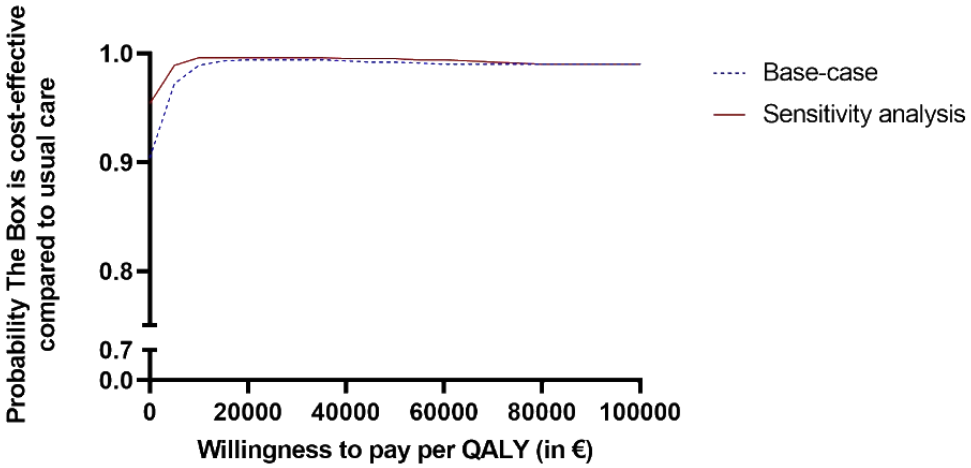


Figure 4. Cost-effectiveness acceptability curve. A currency exchange rate of €1=US $1.0994 is applicable. QALY: quality-adjusted life years.



Discussion

Principal Findings

In this paper, an RCT analyzing the cost-utility of an eHealth intervention in post-AMI patients was presented. Important findings of this paper were that, on average, costs per patient in The Box group were nonsignificant lower than in the control group and quality of life showed a small but statistically significant difference. These findings were corroborated in a sensitivity analysis.

The results from this paper can contribute to the ongoing discussions regarding telemonitoring and telerehabilitation in patients with cardiovascular disease. Rising health care costs are putting pressure on budgets for health care in all developed countries. In the Netherlands, health care costs are an estimated 11% of the total gross domestic product (GDP). Costs are growing faster than the economy. Without a significant change in the way health care is delivered by 2040, it is expected that 30% of the GDP will be spent on health care [19]. This increase in costs has been attributed to increased volumes, patients with multi-morbidity, as well as the use of more sophisticated (and therefore expensive) clinical technology [20]. eHealth has been identified as a tool to lower costs while at the same time increasing quality by focusing more on preventing disease (instead of treating). Moreover, it could reduce costs by helping to integrate care by easing communication between care providers and reducing duplication of diagnostic testing. Lastly, it could reduce costs as patients are enabled to perform some of their diagnostic tests by themselves instead of by trained health care staff [21]. Although these are rather general remarks, the results of this study support some of this theory. In this study, patients were able to measure their own BP, ECG, and weight and transfer it to the hospital. This enabled the replacement of two in-office outpatient clinic visits with two digital outpatient clinic visits, with consequently cost reductions, as the price of an e-visit is about half the price of an in-office outpatient clinic. Potentially, with 34,000 AMI patients in the Netherlands per year, the eHealth intervention could save an estimated €16.1 million euros (US $17.7 million US dollars) in health care costs for cardiology departments annually [18,22].

External Validity

This RCT was performed in Dutch patients who suffered from AMI. In the Netherlands, distances are known to be small. The average distance between the hospital and the patient's home was 7 kilometers [14]. Moreover, in this study, it was estimated that 30% of patients were retired. In a population, however, where more patients are still working and distances are larger, cost savings due to eHealth might be higher. A sensitivity analysis, taking into account the patient-related costs of an e-visit, demonstrated that cost savings of The Box could be higher. The costs of devices of The Box should be incorporated as well. These costs contribute significantly to the total costs of the intervention group. In larger populations, a cost reduction could be achieved by reducing the price per device due to larger volumes. A reduction in costs for The Box could result in a statistically significant reduction in total costs per patient compared to the control group. It could be expected that in such a scenario, the cost reduction in the intervention group could reach statistical significance. Moreover, further selection of subpopulations that are most likely to benefit from The Box could improve cost-effectiveness as well.

Literature

To our knowledge, this is the first paper to evaluate the cost-utility of remote monitoring compared to regular follow-up in the outpatient care setting of post-AMI patients. eHealth is a rather broad term, encompassing almost all use of information technologies in health care. It is a relatively new concept. Few RCTs have been performed. A recent systematic review found 16 cost evaluations of RCTs in eHealth, ranging from internet-based cognitive behavioral therapy for depression to telemonitoring for patients with congestive heart failure [23]. As these patient populations and interventions differ from our patient population and intervention, comparing the results is difficult. Previous studies mainly found eHealth to be cost-effective but predominantly leading to an increase in costs. Our study found a (nonsignificant) reduction in costs, which is very likely to be due to the design of this study; eHealth was used to partially replace regular care, while in most eHealth studies, it is provided on top of regular care. One study evaluated telerehabilitation in post-AMI patients [24] and found the

intervention to be cost-saving. A Dutch study with a comparable design and patient population corroborated these findings [25]. However, although this study evaluated to some extent a comparable patient population, a different intervention was performed. The intervention involved a telerehabilitation program and focused on exercise. Digital outpatient clinic visits were not part of the intervention [24]. These factors could explain the differences in the cost reductions found in both studies.

Limitations

For the interpretation of the results, some limitations have to be taken into account. For the trial-based analysis, only data from the Department of Cardiology of the LUMC were used.

Secondly, this study was performed in a tertiary care center by a project dedicated team. It is therefore unknown if the percentages found in this trial will be similar in other medical centers, where care for post-AMI patients is delivered by cardiologists, and a project dedicated team is unfeasible due to lower volumes and other financing structures. Thirdly, this cost-utility calculation was done from a department's perspective. Costs generated in other departments were not taken into account. Therefore, total costs could be underestimated. However, as it is assumed this is equally distributed, it has most likely a limited effect on the difference between the intervention and control group. In the base-case analysis, furthermore, patient-related costs were not taken into account as well. This might have led to an underestimation of costs in the control group, as patient-related costs are assumed to be higher. The sensitivity analysis indicated that if patient-related costs are included, The Box is likely to be more cost-effective. Fourthly, to convert SF-36 scores into utilities, the UK algorithm was used. As such, as there might be subtle differences between the UK and Dutch patient populations in basic SF-36 scores, this might have skewed the utility data slightly. Nevertheless, as the algorithm was used for both the intervention group and control group, there is no reason to believe another algorithm would have changed the conclusion that is based on the utility data.

Conclusions

The most important conclusion is that remote monitoring in post-AMI patients is likely to be cost-effective compared to usual care (and at least not more expensive). This intervention in the outpatient care setting of post-AMI patients could be a valuable additive in restraining rising health care costs or in situations where physical outpatient clinic visits are undesirable.

Acknowledgments

No external funding was received for this study.

Data Availability

The data underlying this article will be shared following a reasonable request to the corresponding author.

Conflicts of Interest

RT reports receiving personal fees from Boston Scientific, Pfizer, and Sanofi outside the submitted work. SB reports receiving personal fees from Boston Scientific outside the submitted work. All other authors report no conflicts of interest.

References

1. Ball MJ, Lillis J. E-health: transforming the physician/patient relationship. Int J Med Inform 2001 Apr;61(1):1-10. [doi: 10.1016/s1386-5056(00)00130-1] [Medline: 11248599]
2. Saner H, van der Velde E. eHealth in cardiovascular medicine: a clinical update. Eur J Prev Cardiol 2016 Oct;23(2 suppl):5-12. [doi: 10.1177/2047487316670256] [Medline: 27892420]
3. Internet: access, use and facilities. StatLine. 2019 Aug 05. URL: http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=83429NED&D1=0,2-5&D2=0,3-6&D3=0&D4=a&HDR=T&STB=G1,G2,G3&VW=T [accessed 2022-03-30]
4. Keyserling TC, Sheridan SL, Draeger LB, Finkelstein EA, Gizlice Z, Kruger E, et al. A comparison of live counseling with a web-based lifestyle and medication intervention to reduce coronary heart disease risk: a randomized clinical trial. JAMA Intern Med 2014 Jul;174(7):1144-1157 [FREE Full text] [doi: 10.1001/jamainternmed.2014.1984] [Medline: 24861959]
5. Treskes RW, van Winden LAM, van Keulen N, van der Velde ET, Beeres SLMA, Atsma DE, et al. Effect of smartphone-enabled health monitoring devices vs regular follow-up on blood pressure control among patients after myocardial infarction: a randomized clinical trial. JAMA Netw Open 2020 Apr 01;3(4):e202165 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.2165] [Medline: 32297946]
6. Treskes RW, van Winden LA, van Keulen N, Atsma DE, van der Velde ET, van den Akker-van Marle E, et al. Using smart technology to improve outcomes in myocardial infarction patients: rationale and design of a protocol for a randomized controlled trial, the box. JMIR Res Protoc 2017 Sep 22;6(9):e186 [FREE Full text] [doi: 10.2196/resprot.8038] [Medline: 28939546]
7. Task Force on the management of ST-segment elevation acute myocardial infarction of the European Society of Cardiology (ESC), Steg PHG, James SK, Atar D, Badano LP, Blömstrom-Lundqvist C, et al. ESC Guidelines for the management of

acute myocardial infarction in patients presenting with ST-segment elevation. Eur Heart J 2012 Oct;33(20):2569-2619. [doi: 10.1093/eurheartj/ehs215] [Medline: 22922416]
8. Roffi M, Patrono C, Collet JP, Mueller C, Valgimigli M, Andreotti F, ESC Scientific Document Group. 2015 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: Task Force for the Management of Acute Coronary Syndromes in Patients Presenting without Persistent ST-Segment Elevation of the European Society of Cardiology (ESC). Eur Heart J 2016 Jan 14;37(3):267-315. [doi: 10.1093/eurheartj/ehv320] [Medline: 26320110]
9. Ibanez B, James S, Agewall S, Antunes MJ, Bucciarelli-Ducci C, Bueno H, ESC Scientific Document Group. 2017 ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: The Task Force for the management of acute myocardial infarction in patients presenting with ST-segment elevation of the European Society of Cardiology (ESC). Eur Heart J 2018 Jan 07;39(2):119-177. [doi: 10.1093/eurheartj/ehx393] [Medline: 28886621]
10. Liem SS, van der Hoeven BL, Oemrawsingh PV, Bax JJ, van der Bom JG, Bosch J, et al. MISSION!: optimization of acute and chronic care for patients with acute myocardial infarction. Am Heart J 2007 Jan;153(1):14.e1-14.11. [doi: 10.1016/j.ahj.2006.10.002] [Medline: 17174628]
11. Health care in the Netherlands. National Health Care Institute. URL: https://english.zorginstituutnederland.nl/about-us/publications/publications/2016/01/31/healthcare-in-the-netherlands [accessed 2022-03-30]
12. Jaarmutatie consumentenprijsindex; vanaf 1963. StatLine. URL: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70936ned/table?ts=1534453580289 [accessed 2019-08-05]
13. Tarieventabel dbc-zorgproducten en overige zorgproducten per 1 januari 2018. PUC publicatie Platform. URL: https://puc.overheid.nl/nza/doc/PUC_13274_22/1/ [accessed 2022-03-30]
14. Kanters TA, Bouwmans CAM, van der Linden N, Tan SS, Hakkaart-van Roijen L. Update of the Dutch manual for costing studies in health care. PLoS One 2017;12(11):e0187477 [FREE Full text] [doi: 10.1371/journal.pone.0187477] [Medline: 29121647]
15. Brazier JE, Harper R, Jones NM, O'Cathain A, Thomas KJ, Usherwood T, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. BMJ 1992 Jul 18;305(6846):160-164 [FREE Full text] [doi: 10.1136/bmj.305.6846.160] [Medline: 1285753]
16. SF-6D: calculating QALYs from the SF-36 and SF-12. Sheffield Licensing: University of Sheffield. URL: https://licensing.sheffield.ac.uk/i/health-outcomes/SF-6D.html [accessed 2020-09-11]
17. Rechtvaardige en duurzame zorg [Equitable and sustainable care]. Den Haag: RVZ; 2007.
18. Koek JL, de Bruin A, Gast A, Gevers E, Kardaun JWPF, Reitsma JB, et al. Incidence of first acute myocardial infarction in the Netherlands. Neth J Med 2007 Dec;65(11):434-441 [FREE Full text] [Medline: 18079566]
19. Brinkman A. RIVM: Zorgkosten stijgen naar 174 miljard in 2040. Zorgvisie. URL: https://www.zorgvisie.nl/rivm-zorgkosten-stijgen-naar-174-miljard-in-2040/ [accessed 2018-08-13]
20. Dieleman JL, Squires E, Bui AL, Campbell M, Chapin A, Hamavid H, et al. Factors associated with increases in US health care spending, 1996-2013. JAMA 2017 Nov 07;318(17):1668-1678 [FREE Full text] [doi: 10.1001/jama.2017.15927] [Medline: 29114831]
21. Topol E. The Creative Destruction of Medicine. New York, NY: Basic Books; 2013.
22. Hart- en vaatziekten in Nederland 2018. Hartstichting. URL: https://www.hartstichting.nl/getmedia/a6e15c10-2710-41b9-bcf8-8185feaf54b2/cijferboek-hartstichting-hart-vaatziekten-nederland-2018.pdf [accessed 2019-08-31]
23. Sanyal C, Stolee P, Juzwishin D, Husereau D. Economic evaluations of eHealth technologies: a systematic review. PLoS One 2018;13(6):e0198112 [FREE Full text] [doi: 10.1371/journal.pone.0198112] [Medline: 29897921]
24. Frederix I, Hansen D, Coninx K, Vandervoort P, Vandijck D, Hens N, et al. Effect of comprehensive cardiac telerehabilitation on one-year cardiovascular rehospitalization rate, medical costs and quality of life: a cost-effectiveness analysis. Eur J Prev Cardiol 2016 May;23(7):674-682. [doi: 10.1177/2047487315602257] [Medline: 26289723]
25. Kraal JJ, Van den Akker-Van Marle ME, Abu-Hanna A, Stut W, Peek N, Kemps HM. Clinical and cost-effectiveness of home-based cardiac rehabilitation compared to conventional, centre-based cardiac rehabilitation: results of the FIT@Home study. Eur J Prev Cardiol 2017 Aug;24(12):1260-1273 [FREE Full text] [doi: 10.1177/2047487317710803] [Medline: 28534417]

Abbreviations

AMI: acute myocardial infarction
BP: blood pressure
CCU: cardiac care unit
ECG: electrocardiogram
ESC: European Society of Cardiology
GDP: gross domestic product
LUMC: Leiden University Medical Center
NP: nurse practitioner

NSTEMI: non-ST-elevation myocardial infarction
NZA: Nederlandse Zorgautoriteit
PCI: percutaneous coronary intervention
QALY: quality-adjusted life years
RCT: randomized controlled trial
SF-36: Short Form Health Survey-36
STEMI: ST-elevation myocardial infarction
TTE: transthoracic echocardiogram

Edited by A Mavragani; submitted 06.05.21; peer-reviewed by KK Lim, R Tonino, M Mars, B Cher; comments to author 30.07.21; revised version received 23.12.21; accepted 05.02.22; published 25.04.22.

Please cite as:
Treskes RW, van den Akker-van Marle ME, van Winden L, van Keulen N, van der Velde ET, Beeres S, Atsma D, Schalij MJ
The Box—eHealth in the Outpatient Clinic Follow-up of Patients With Acute Myocardial Infarction: Cost-Utility Analysis
J Med Internet Res 2022;24(4):e30236
URL: https://www.jmir.org/2022/4/e30236
doi:10.2196/30236
PMID:35468091

Original Paper

Social Networking Service, Patient-Generated Health Data, and Population Health Informatics: National Cross-sectional Study of Patterns and Implications of Leveraging Digital Technologies to Support Mental Health and Well-being

Jiancheng Ye[1]; Zidan Wang[2]; Jiarui Hai[3]

[1]Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[2]Department of Statistics, Northwestern University, Evanston, IL, United States

[3]Department of Hydraulic Engineering, Tsinghua University, Beijing, China

Corresponding Author:
Jiancheng Ye
Feinberg School of Medicine
Northwestern University
633 N. Saint Clair St
Chicago, IL, 60611
United States
Phone: 1 312 503 3690
Email: jiancheng.ye@u.northwestern.edu

Abstract

Background: The emerging health technologies and digital services provide effective ways of collecting health information and gathering patient-generated health data (PGHD), which provide a more holistic view of a patient's health and quality of life over time, increase visibility into a patient's adherence to a treatment plan or study protocol, and enable timely intervention before a costly care episode.

Objective: Through a national cross-sectional survey in the United States, we aimed to describe and compare the characteristics of populations with and without mental health issues (depression or anxiety disorders), including physical health, sleep, and alcohol use. We also examined the patterns of social networking service use, PGHD, and attitudes toward health information sharing and activities among the participants, which provided nationally representative estimates.

Methods: We drew data from the 2019 Health Information National Trends Survey of the National Cancer Institute. The participants were divided into 2 groups according to mental health status. Then, we described and compared the characteristics of the social determinants of health, health status, sleeping and drinking behaviors, and patterns of social networking service use and health information data sharing between the 2 groups. Multivariable logistic regression models were applied to assess the predictors of mental health. All the analyses were weighted to provide nationally representative estimates.

Results: Participants with mental health issues were significantly more likely to be younger, White, female, and lower-income; have a history of chronic diseases; and be less capable of taking care of their own health. Regarding behavioral health, they slept <6 hours on average, had worse sleep quality, and consumed more alcohol. In addition, they were more likely to visit and share health information on social networking sites, write online diary blogs, participate in online forums or support groups, and watch health-related videos.

Conclusions: This study illustrates that individuals with mental health issues have inequitable social determinants of health, poor physical health, and poor behavioral health. However, they are more likely to use social networking platforms and services, share their health information, and actively engage with PGHD. Leveraging these digital technologies and services could be beneficial for developing tailored and effective strategies for self-monitoring and self-management.

(J Med Internet Res 2022;24(4):e30898) doi:10.2196/30898

KEYWORDS

patient-generated health data; social network; population health informatics; mental health; social determinants of health; health data sharing; technology acceptability; mobile phone; mobile health

Introduction

Background

Mental health issues such as depression and anxiety disorders are severe psychiatric diseases with high prevalence and elevated risks of recurrence and chronicity [1]. There are >260 million people of all ages who have experienced mental illnesses worldwide, which are a leading cause of disability worldwide and a major contributor to the overall global burden of disease [2]. Studies have demonstrated that mental health issues are a strong indicator of poor general health, unhealthy alcohol use, and sleep problems [3,4]. Poor sleep quality has been linked to an increased motivation to drink, especially for young adults [5]. It is critical for patients with mental health issues to receive appropriate health care and social services.

In recent years, there has been increasing acknowledgment of the important role that mental health plays in achieving improved population health. Understanding how these fundamental factors (physical and behavioral health, mental health, and technologies) relate to one another may yield important insights for novel approaches to designing prevention programs and enhancing services for mental health support. Digital health technologies such as smartphone apps and social media provide opportunities to continuously collect objective information on behavior in the context of people's real lives, generating a rich data set that can provide insights into the extent and timing of mental health needs in individuals [6].

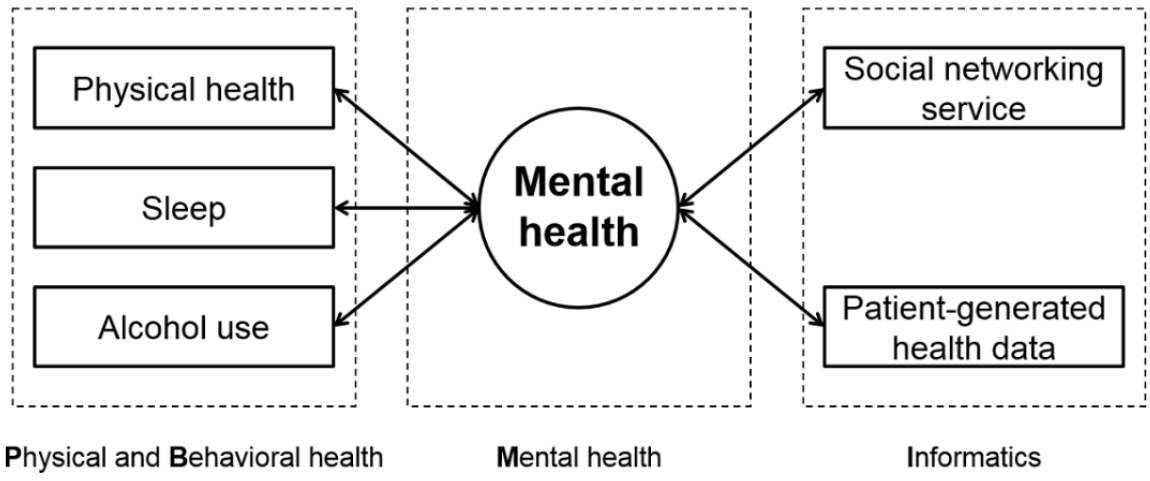
However, long-standing problems have hampered the efforts to improve mental health care delivery, quality of care, and social support. For example, if mental health conditions are assessed exclusively on patients' self-reporting, it may be burdensome to collect and subjective for clinical decision support. Currently, mental health services are mainly provided at times chosen by the practitioner rather than at the patient's time of greatest need [7]. The ideal way of providing support is to conduct regular assessments, which is useful for capturing the temporal dynamics of symptoms and crucial for both diagnosis and treatment planning [8]. However, this could contribute to burnout among health care providers and patients [9].

The emerging health technologies and digital services provide effective ways of collecting human behavior information, gathering patient-generated health data (PGHD), and sharing health-related information outside clinical settings in a systematic way, thus making interventions timely. Coupled with population health informatics tools, these technologies can track people's digital exhaust, which includes PGHD and social networking platform use [10]. Social networking services are web-based platforms that people use to build social networks or social relationships with other people who share similar personal interests, activities, backgrounds, or real-life connections [11]. The rich real-time data enable researchers to gain insights into aspects of behavior that are well-established building blocks of mental health and illness, such as mood, social communication, sleep, alcohol use, and physical activity.

Objectives

This study had 2 aims. The first aim was to provide a conceptual framework that will be used to describe the relationship between physical and behavioral health, mental health, and informatics. Figure 1 demonstrates the conceptual framework—Physical and Behavioral Health, Mental Health, and Informatics (PBMI)—for this study. The results could provide a comprehensive understanding of the relationship between health, behavior, and informatics, which could be useful for developing tailored and effective strategies to support mental health management. The second aim was to describe and compare characteristics of populations with and without mental health issues (depression or anxiety disorders), including physical health, sleep, and alcohol use, based on the proposed PBMI framework. We also examined the patterns of social networking service use, PGHD, and attitudes toward health information sharing and activities.

Figure 1. Physical and Behavioral Health, Mental Health, and Informatics (PBMI) framework.



Methods

Study Design

Data for this study were drawn from the 2019 Health Information National Trends Survey (HINTS) of the National Cancer Institute. HINTS is a nationally representative survey administered every year by the National Cancer Institute that provides a comprehensive assessment of the American public's current access to and use of health information [12]. The HINTS target population is civilian, noninstitutionalized adults aged ≥18 years living in the United States. In this study, we investigated the relationships between mental health, physical health, behavioral health, and social networking service use. Social networking services include sharing health information, writing online diary blogs, participating in online forums or health-related groups, and watching health-related videos.

Study Participants

The data used in this study were from the third round of data collection for HINTS 5 (cycle 3), which was conducted from January 22, 2019, to April 30, 2019. Cycle 3 received 5590 questionnaires, of which 5438 (97.28%) were determined to be eligible after excluding blank, incomplete, and duplicate surveys.

In this study, the primary outcome was the presence of mental health issues, which was determined by the participant's status of depression or anxiety disorder based on the results of the question *Has a doctor or other health professional ever told you that you had depression or anxiety disorder (yes/no)?* Of the 5438 eligible respondents, 1139 (20.95%) reported *yes*, 4168 (76.65%) reported *no*, and 131 (2.41%) were missing and omitted from our analyses.

Measures

Social Determinants of Health

The sample was divided into 2 groups according to mental health status. Participants with depression or anxiety disorders were classified as the group with mental health issues, and the others were classified as the group with no mental health issues. We used the participants' self-reported information on age, sex, race, ethnicity, level of education, annual income, and usual source of care as our sociodemographic variables. We transformed the continuous variable of age into a categorical variable by classifying age into four groups: (1) 18 to 34 years, (2) 35 to 49 years, (3) 50 to 64 years, and (4) ≥65 years. Education level was recategorized as less than college (including post–high school training), some college, college graduate, and postgraduate degree. Annual income level was recategorized as ≤US $20,000, US $20,000 to $35,000, US $35,000 to $50,000, US $50,000 to $75,000, and >US $75,000. We examined the participants' history of chronic conditions using four questions (all with yes or no responses): *Has a doctor or other health professionals ever told you that you had (1) diabetes, (2) high blood pressure, (3) a heart condition, and (4) chronic lung disease?*

Health-Related Information

To assess the participants' general health, we considered their answers to questions related to physical health and mental status, including the ability to take care of their health, emotion control by changing the way of thinking, and future consideration. We also included the Patient Health Questionnaire–4 (PHQ-4), which was a derived composite from the participants' responses to questions on lack of interest in doing things, presence of depressed feelings, nervousness and anxiousness, and uncontrolled worry [13].

PGHD and Social Networking Service Use

We examined the participants' use of the internet for health-related reasons using the following five survey questions (all with yes or no responses): *In the past 12 months, have you used the internet to (1) visit a social networking site, such as Facebook or LinkedIn, (2) share health information on social networking sites, such as Facebook or Twitter, (3) write in an online diary or blog, (4) participate in an online forum or support group for people with similar health or medical issue, or (5) watch a health-related video on YouTube?*

We also inspected the first source of health information of the participants using their responses to the following question—*The most recent time you looked for information about health or medical topics, where did you go first?*—where the respondent could select one of 12 options. We further grouped the options into five main categories: internet, health professionals, family and friends, print materials, and others. In addition, we investigated the participants' attitudes toward sharing health information, such as avoidance of physician visits and talking about health with family and friends.

Alcohol Consumption and Sleep

We examined the participants' alcohol consumption using two questions: the number of days with at least one alcoholic drink per week and the average number of drinks per day. We assessed the participants' sleep hours and quality using two questions: the average number of hours of sleep per night and the self-rated overall sleep quality. We transformed the continuous variable of average sleep per night into a categorical variable by classifying sleep hours as (1) 0 to 6 hours, (2) 7 to 8 hours, and (3) ≥9 hours.

Statistical Analysis

We used the survey package in the R programming language (version 4.0.5; R Foundation for Statistical Computing) to account for the complex sampling design used in HINTS and incorporated the Taylor series (linear approximation) [14] to generate accurate variance estimation. All analyses used weighted data based on the Taylor series method to calculate population estimates. Pairwise deletion was used to deal with missing data to preserve more information.

To assess sociodemographic characteristics, general health, chronic diseases, social networking service use, alcohol consumption, and sleeping variables, we generated weighted 2-way cross-tabulation tables, which were tested with a Pearson chi-square test of association [15].

A univariate logistic regression was built to examine the association between each predictor and mental health. We then performed multivariate logistic regression analyses using a survey-weighted generalized linear modeling function in R [16].

The variables included the participants' sociodemographic and clinical characteristics. Odds ratios (ORs) and 95% CIs for both models were calculated. All reported *P* values were 2-tailed, and a cutoff of $P<.05$ was used to determine statistical significance for all analyses.

Ethics Approval

The data for this study are publicly available.

Results

Population Characteristics

Table 1 reports the sociodemographic and clinical characteristics of the participants. Respondents with mental health issues were significantly more likely to be younger ($P=.004$), White ($P=.005$), and female ($P<.001$); have a lower income ($P<.001$); and have a usual source of care ($P<.001$). They were also more likely to have a history of diabetes ($P=.001$) and lung disease ($P<.001$). There were no significant differences between the 2 groups regarding the characteristics of history of hypertension ($P=.10$), heart condition ($P=.40$), and cancer ($P=.13$).

Table 1. Unweighted and weighted prevalence estimates for sample sociodemographic characteristics, Health Information National Trends Survey 5, cycle 3.

Characteristic	Overall (n=5438), unweighted %	Overall (n=252,070,495), weighted %	Having mental health issues (n=57,953,433), weighted %	No mental health issues (n=189,456,090), weighted %	*P* value
Age (years)					.004
18 to 34	13	24.3	28.1	23.1	
35 to 49	18.3	24.5	26.3	24.2	
50 to 64	31.6	31.1	32.8	30.7	
≥65	37.1	20.2	12.8	22	
Sex (male)	42.1	48.8	37.9	52.5	<.001
Race[a]					.005
White	73.9	77	82.9	75	
Black	16.5	13	9.9	13.9	
Asian	5	5.8	3.4	6.7	
Others	4.7	4.2	3.9	4.4	
Ethnicity (Hispanic)	14.9	16.7	13.9	17.6	.07
Education					.41
High school diploma or less	54.4	70.5	72.5	69.7	
College degree	26.5	17.3	16.1	17.8	
Postgraduate degree	19.1	12.2	11.4	12.5	
Income (US $)					<.001
<20,000	18.8	18.5	26.2	15.6	
20,000 to 34,999	12.8	11	11.6	10.7	
35,000 to 49,999	13.1	13.5	14.2	13.4	
50,000 to 74,999	17.7	17.4	16.8	17.7	
≥75,000	37.6	39.6	31.3	42.6	
Insurance					.18
Public	44.1	35.2	38.9	33.5	
Private	42.3	48.9	44.8	50.8	
Uninsured	4.8	7.7	7.2	7.8	
Others[b]	8.8	8.2	9.2	8	
Has a usual source of care	69.8	64.5	73.7	61.7	<.001
History of cancer	16.1	9.5	7.9	9.7	.13
History of lung disease	11.8	11.2	2	8.2	<.001
History of heart condition	16.1	8.1	9.2	7.8	.40
History of diabetes	21.7	17	22.4	15.2	.001
History of hypertension	45	36	39.1	34.9	.10

[a]Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, and other Asian were collapsed into the *Asian* category. Race categories other than White, Black, and Asian were reclassified as *Others*.

[b]*Others* include coverage under the spouse, coverage under parents, and low-income beneficiary.

Health Information and Social Networking Service

Table 2 shows the characteristics of health information source, health information sharing, and social networking service use. Participants with mental health issues were more likely to have a worse general health status (*P*<.001), less confidence in taking care of their own health (*P*<.001), and a higher PHQ-4 score (*P*<.001). Individuals with mental health issues were also less

likely to control emotions by changing the way they thought about situations (*P*<.001) and try to influence things in the future with day-to-day behavior (*P*<.001). In addition, Table 2 shows that those with mental health issues were significantly more likely to visit social networking sites (*P*=.04), share health information on social networking sites (*P*=.001), write online diary blogs (*P*=.007), participate in an online forum or support group (*P*<.001), and watch health-related videos (*P*=.009). There were no significant differences between the 2 groups in the first source of health information (*P*=.23), using wellness apps (*P*=.33), avoidance of physician visits (*P*=.15), and talking about health with family or friends (*P*=.08). Multimedia Appendix 1 shows the results of the multivariate logistic regression of health information and social networking service use.

Table 2. Prevalence estimates for characteristics of health information and social networking service use.

Characteristic	Overall (n=5438), unweighted %	Overall (n=252,070,495), weighted %	Having mental health issues (n=57,953,433), weighted %	No mental health issues (n=189,456,090), weighted %	*P* value
Source of health information					.23
Internet	42.9	46.1	50.2	45.2	
Health professionals	48.9	44.6	39.9	45.9	
Family or friends	4.1	5.2	4.6	5.3	
Print materials	2.3	2.2	2.4	2.1	
Others	1.8	2	2.9	1.5	
Use of health apps					.33
Yes	52.4	54.8	57.7	54.2	
No	42.2	39.6	35.9	40.6	
Do not know	5.4	5.5	6.4	5.2	
Good health status	47.9	49.4	33.7	54.2	<.001
Have ability to take care of health	72.2	71.5	56	76.3	<.001
Avoid visiting physician	25	30.6	33.9	29.6	.15
Talks about health with family or friends	81	78.5	81.9	77.4	.08
PHQ-4[a]					<.001
0	50.3	45.6	13.6	54.9	
1	12.3	13.2	9.3	14.4	
2	9.9	9.6	11.5	9.1	
≥3	27.6	31.6	65.5	21.6	
Can control emotions	85.1	84.5	77.6	86.7	<.001
Consider future	84.5	84.7	79.2	86.6	<.001
Visit social networking sites	65	71.5	76	70.8	.04
Share health information	11.9	14.6	19.7	13.1	.001
Write online diary blog	3.6	5	8	4	.007
Participate in online forum or health-related group	7	8.1	13.3	6.6	<.001
Watch health-related videos	32.8	37.3	42.5	35.9	.009

[a]PHQ-4: Patient Health Questionnaire–4.

Alcohol Consumption and Sleep

Table 3 shows the characteristics of behavioral health, including sleep and alcohol use, of the 2 groups. Individuals with mental health issues were more likely to sleep <6 hours or >9 hours (*P*=.01), have worse sleep quality (*P*<.001), and consume more alcohol per day (*P*=.03). There was no significant difference in the number of days of alcohol consumption per week between the 2 groups.

Figure 2 illustrates the difference in sleep quality among individuals who slept ≤6 hours, 7 to 8 hours, and >9 hours per night between the 2 groups. For individuals with mental health issues, 52% of those who slept ≤6 hours per night had a poor sleep quality. Among individuals without mental health issues,

only 9.1% of those who slept 7 to 8 hours per night had a poor sleep quality, which is significantly less than that of individuals with mental health issues who slept the same hours.

We examined whether sleep quality was the same for populations with and without mental health issues separately within the 3 sleep hour categories (0-6 hours, 7-8 hours, and ≥9 hours). As the normality assumption is unjustified, we conducted the Mann-Whitney *U* test. For people who slept 0 to 6 hours (*P*<.001) and 7 to 8 hours (*P*<.001), there was a significant difference in sleep quality between the 2 groups. There was no significant difference in sleep quality for people who slept >9 hours (*P*=.14) between the 2 groups. We found that individuals without mental health issues slept 7 to 8 hours with higher quality, whereas patients with mental health issues slept <6 hours or >9 hours with poor quality.

Figure 3 illustrates the difference in the amount of alcohol consumed per week between the groups stratified by sex and mental health status. Approximately 43% (38,696,406/89,991,643) of women without mental health issues consumed one or more drinks per week, whereas 44.8% (16,123,109/35,989,082) of women with mental health issues consumed the same number of drinks per week. We found no significant difference in drink amount between women (*P*=.66) and men (*P*=.23) regardless of mental health status. However, for individuals without mental health issues, men drank significantly more than women (*P*<.001).

Table 3. Prevalence estimates for characteristics of sleep and alcohol use.

Characteristic	Overall (n=5438), unweighted %	Overall (n=252,070,495), weighted %	Having mental health issues (n=57,953,433), weighted %	No mental health issues (n=189,456,090), weighted %	*P* value
Sleep hours per night					.01
0 to 6	38.9	38.9	41.4	38.2	
7 to 8	53.3	54.1	49	55.7	
≥9	7.8	7	9.6	6.1	
Overall sleep quality					<.001
Very good	20	18.8	10.5	21.1	
Fairly good	58.2	58.7	55.7	59.8	
Fairly bad	17.8	18.3	25.7	16.1	
Very bad	4	4.2	8	3	
Days consuming alcohol per week					.52
0	50.5	51.3	54.2	50.3	
1 to 2	26.5	27.3	25.6	27.9	
3 to 4	10.8	11.1	9.8	11.5	
≥5	12.2	10.3	10.5	10.3	
Alcohol drinks per day					.03
0 to 1	43.2	37.7	29.3	40.1	
2 to 3	43.9	45.3	51.4	43.6	
≥4	12.9	17	19.3	16.4	

Figure 2. Sleep patterns between the 2 mental health groups.

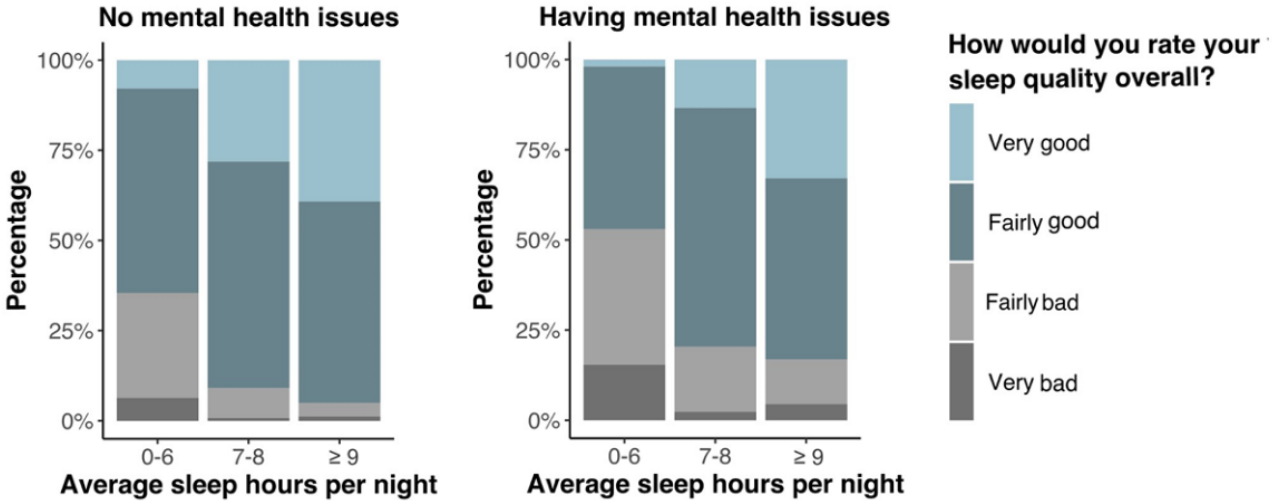
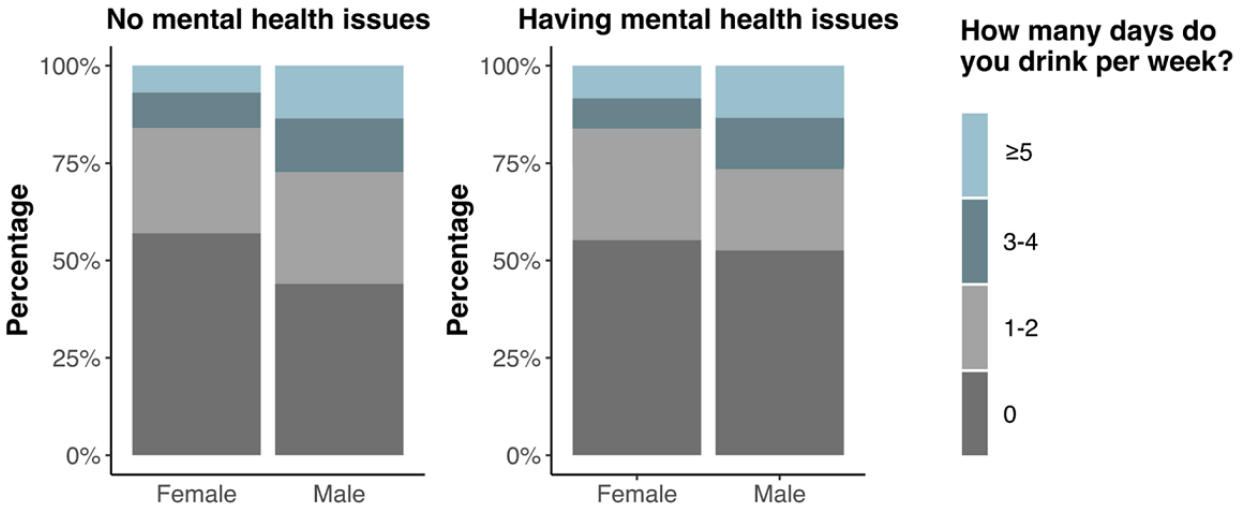


Figure 3. Alcohol use patterns stratified by sex and mental health status.



Social Determinants of Health and Mental Health

Table 4 shows the results of logistic regression analyses. In the unadjusted logistic regression model, most covariates were associated with mental health. In the adjusted model, those aged ≥65 years had a reduced likelihood (OR 0.20, 95% CI 0.11-0.35) of having mental health issues compared with those aged 18 to 34 years. Men were less likely (OR 0.52, 95% CI 0.40-0.68) to have mental health issues. The Black population had a reduced likelihood (OR 0.41, 95% CI 0.27-0.63) of having mental health issues compared with the White population.

Individuals who had an annual family income <US $20,000 were more likely (OR 2.39, 95% CI 1.48-3.84) to have mental health issues than those whose income was >US $75,000. Those having a usual source of care were more likely (OR 1.72, 95% CI 1.24-2.39) to have mental health issues. As expected, those with a history of lung disease (OR 2.17, 95% CI 1.54-3.05), diabetes (OR 1.42, 95% CI 1.03-1.95), and hypertension (OR 1.40, 95% CI 1.07-1.84) were more likely to have mental health issues. The results also indicated that ethnicity, education, insurance type, history of cancer, and history of heart condition had no association with mental health status.

Table 4. Crude and adjusted odds from logistic regression analyses of associations between social determinants of health and mental health.

Predictor	Unadjusted		Adjusted	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value
Age (years)				
18 to 34	Reference	Reference	Reference	Reference
35 to 49	0.89 (0.61-1.31)	.56	0.99 (0.64-1.54)	.98
50 to 64	0.88 (0.60-1.29)	.51	0.60 (0.38-0.96)	.04
≥65	0.48 (0.32-0.71)	<.001	0.20 (0.11-0.35)	<.001
Sex (male)	0.55 (0.43-0.70)	<.001	0.52 (0.40-0.68)	<.001
Race				
White	Reference	Reference	Reference	Reference
Asian	0.46 (0.25-0.86)	.02	0.51 (0.25-1.03)	.06
Black	0.64 (0.46-0.90)	.01	0.41 (0.27-0.63)	<.001
Others	0.79 (0.48-1.32)	.37	0.52 (0.25-1.08)	.08
Ethnicity (Hispanic)	0.76 (0.56-1.02)	.07	0.73 (0.48-1.11)	.14
Education				
High school diploma or less	Reference	Reference	Reference	Reference
College degree	0.87 (0.68-1.12)	.28	0.91 (0.64-1.31)	.63
Postgraduate degree	0.87 (0.65-1.17)	.37	0.92 (0.59-1.42)	.69
Income (US $)				
<20,000	2.29 (1.65-3.17)	<.001	2.39 (1.48-3.84)	<.001
20,000 to 34,999	1.48 (1.00-2.19)	.05	1.59 (1.00-2.55)	.05
35,000 to 49,999	1.44 (0.97-2.14)	.07	1.34 (0.83-2.15)	.23
50,000 to 74,999	1.29 (0.89-1.86)	.18	1.32 (0.88-1.98)	.19
≥75,000	Reference	Reference	Reference	Reference
Insurance				
Public	Reference	Reference	Reference	Reference
Private	0.76 (0.59-0.98)	.03	0.71 (0.47-1.07)	.11
Uninsured	0.79 (0.48-1.31)	.37	0.56 (0.29-1.09)	.09
Others	0.99 (0.65-1.51)	.98	0.94 (0.59-1.51)	.81
Has a usual source of care	1.74 (1.36-2.24)	<.001	1.72 (1.24-2.39)	.001
History of cancer	0.81 (0.61-1.07)	.14	0.88 (0.61-1.28)	.51
History of lung disease	2.99 (2.21-4.04)	<.001	2.17 (1.54-3.05)	<.001
History of heart condition	1.19 (0.79-1.80)	.40	0.94 (0.60-1.47)	.79
History of diabetes	1.61 (1.23-2.10)	<.001	1.42 (1.03-1.95)	.03
History of hypertension	1.20 (0.97-1.48)	.10	1.40 (1.07-1.84)	.01

[a]OR: odds ratio.

Discussion

Principal Findings

This study aimed to describe and compare the characteristics of populations with and without mental health issues (depression or anxiety disorders), including physical health, sleep, and alcohol use. We examined the patterns of social networking service use, PGHD, and attitudes toward health information sharing and activities. We found that participants who were younger, White, and female; had a lower income; had a history of chronic disease; and had a higher PHQ-4 score were more likely to have mental health problems, which is consistent with previous findings [17]. Overall, social determinants of health such as age, race, income, insurance status, and chronic diseases, including lung disease, diabetes, and hypertension, were associated with mental health. Participants with mental illness

were more likely to visit social networking sites, share health information on social networking sites, write online diary blogs, participate in online forums or support groups, and watch health-related videos. We also found that participants with mental illness slept less with worse sleep quality and consumed more alcohol per day.

Health disparities exist between women and men and among different races with regard to mental health. Mental health issues result in less sleep with poor quality and unhealthy alcohol consumption behaviors. Individuals with mental health issues are more likely to use social networking platforms, share their health information, and actively engage in PGHD. The results provide important insights into the interplay between three vital health-related domains—physical health, behavioral health (sleep and alcohol use), and social networking service use and their patterns in populations with mental health issues.

In recent years, there has been increasing acknowledgment of the important role that mental health plays in achieving improved population health. Understanding how these fundamental factors (physical and behavioral health, mental health, and technologies) relate to one another may yield important insights for novel approaches to designing prevention programs and enhancing services for mental health support. Digital health technologies such as smartphone apps and social media provide opportunities to continuously collect objective information on behavior in the context of people's real lives, generating a rich data set that can provide insights into the extent and timing of mental and physical health needs in individuals [6].

Social Networking Service

Individuals who have depression and anxiety are more likely to use social networking platforms, especially younger people. They also tend to be less likely to control emotions by changing the way they think about situations and try to influence things in the future with day-to-day behavior. Social networking plays an important role for this population to find ways to reduce loneliness or symptoms of mental health problems.

We also found that women had a higher level of vulnerability to poor mental health compared with men, which aligned with previous findings [18]. There is an ongoing debate on whether the use of mobile health technologies such as social media is detrimental to mental health [19]. Interestingly, those with depression or anxiety disorders were significantly more likely to visit social networking sites, write online diary blogs, participate in an online forum or support group, and watch health-related videos. These social networking platforms could potentially provide effective strategies to intervene in mental illness. We acknowledge that safe limited use of social media is beneficial, but it could introduce harmful influences if people spend too much time in this digital and internet-based world [20]. Further research is needed to understand the quantitative and dynamic patterns of social media use to measure its benefits and harmful effects and inform evidence-based approaches to clinical interventions, practices, policy, education, and regulation [21]. If we take advantage of the social networking services and data-gathering functions of digital platforms in the right ways, we may achieve breakthroughs in the technologies' ability to support mental health and well-being.

Mental Health and PGHD

This study found that individuals with depression or anxiety disorders were willing to share health information on social networking sites, which offers an opportunity to provide interventions that are timely, personalized, and scalable. Coupled with telehealth or remote management platforms [22-24], practitioners could provide mental health services and support in a timelier manner and at each individual's time of greatest need. Digital health platforms and PGHD are facilitating the development of a wave of timely interventions for mental health care and support [7]. Big data technologies are facilitating the integration of PGHD and electronic health records, which will encourage the use of predictive analytics and artificial intelligence such as natural language processing and machine learning on structured and unstructured data to help health care providers, hospitals, and patients make their data more meaningful [25,26]. These findings may be useful for stakeholders such as health care providers, researchers, public health practitioners, and mobile health and social media companies and encourage them to work jointly to design and provide *precision social networking service* with higher personalized and participatory levels, thus improving population health [27].

Mental Health, Alcohol Use, and Sleep

This study found that individuals without mental health issues slept 7 to 8 hours with higher quality, whereas patients with mental health issues slept <6 hours or >9 hours with poor quality. Scientific guidelines for sleep suggest that ≥7 hours of sleep per night are appropriate for adults aged 18 to 60 years, 7 to 9 hours are appropriate for adults aged 61 to 64 years, and 7 to 8 hours are appropriate for adults aged ≥65 years [28,29]. Although the amount of sleep is important, other aspects of sleep also contribute to health and well-being. Good sleep quality is also essential. We found that patients with mental health issues were more likely to sleep too much, which is not recommended by health professionals. Previous studies have shown that adolescents and young adults are prone to both mental health and sleep problems [30]. Sleep quality may be particularly important for young adults such as college students with poor mental health who, compared with their peers, tend to lack protective social support networks [31].

Among individuals who are already susceptible to alcohol use, inadequate sleep may further weaken their cognitive capacity to make safer drinking-related decisions or their self-protective behaviors irrespective of consumption levels. Further investigations are needed to examine how poor mental health relates to both alcohol consumption and consequences as well as the extent to which alcohol consumption may mediate the relationship between mental health and consequences. Digital social platforms play a vital role in educating people on alternative coping or harm-reduction skills to use in drinking contexts.

Given the important role of different types of drinking motives in the connection between mental health and drinking outcomes,

it is important to examine drinking motivations as mediators of this relationship [32]. Furthermore, event-level methods that simultaneously account for individuals' sleep and alcohol use behaviors may be helpful for future longitudinal research.

Limitations

The sample consisted of missing data regarding health outcomes and covariates, which may not be missing completely at random. Those who did not respond to questions may be less active and, thus, our estimates may be subject to bias. As the survey was cross-sectional, we could not examine causality among the variables. Meanwhile, given the limitations of the data set, we did not have information about the use frequency and duration of social networking platforms. Despite these limitations, this study provides a better understanding of the effects and patterns of social networking service use, PGHD, social determinants of health, and mental health.

Conclusions

This study provided a conceptual framework—PBMI—that could be used to describe the relationship between physical and behavioral health, mental health, and informatics. With this framework, we described the health disparities that existed between women and men and among individuals of different races with regard to mental health, patterns of using social networking platforms, sharing health information, and engagement in PGHD. Leveraging digital platforms and population informatics such as mobile health and social media along with PGHD could offer unique opportunities to develop effective self-monitoring and self-management strategies for supporting patients with mental health issues.

Authors' Contributions

JY conceived and designed the study and contributed to the analyses. ZW contributed to the analyses. JY, ZW, and JH contributed to the interpretation of the results and drafting and revision of the manuscript. All the authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Multivariable logistic regression of health information and social networking service use.
[DOCX File , 14 KB - jmir_v24i4e30898_app1.docx]

References

1. Tiller JW. Depression and anxiety. Med J Aust 2013 Sep 16;199(S6):S28-S31. [doi: 10.5694/mja12.10628] [Medline: 25370281]
2. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1789-1858 [FREE Full text] [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
3. Miller BE, Miller MN, Verhegge R, Linville HH, Pumariega AJ. Alcohol misuse among college athletes: self-medication for psychiatric symptoms. J Drug Educ 2002;32(1):41-52. [doi: 10.2190/JDFM-AVAK-G9FV-0MYY] [Medline: 12096556]
4. Ye J. Pediatric mental and behavioral health in the period of quarantine and social distancing with COVID-19. JMIR Pediatr Parent 2020 Jul 28;3(2):e19867 [FREE Full text] [doi: 10.2196/19867] [Medline: 32634105]
5. Digdon N, Landry K. University students' motives for drinking alcohol are related to evening preference, poor sleep, and ways of coping with stress. Biol Rhythm Res 2013 Feb;44(1):1-11. [doi: 10.1080/09291016.2011.632235]
6. Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 2017 May 08;13:23-47 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
7. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIS) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
8. Ye J, Zhang R, Bannon JE, Wang AA, Walunas TL, Kho AN, et al. Identifying practice facilitation delays and barriers in primary care quality improvement. J Am Board Fam Med 2020;33(5):655-664 [FREE Full text] [doi: 10.3122/jabfm.2020.05.200058] [Medline: 32989060]
9. 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]
10. Ye J. The role of health technology and informatics in a global public health emergency: practices and implications from the COVID-19 pandemic. JMIR Med Inform 2020 Jul 14;8(7):e19866 [FREE Full text] [doi: 10.2196/19866] [Medline: 32568725]

11. Boyd DM, Ellison NB. Social network sites: definition, history, and scholarship. J Comput Mediat Commun 2007;13(1):210-230. [doi: 10.1111/j.1083-6101.2007.00393.x]
12. Nelson DE, Kreps GL, Hesse BW, Croyle RT, Willis G, Arora NK, et al. The health information national trends survey (HINTS): development, design, and dissemination. J Health Commun 2004;9(5):443-460. [doi: 10.1080/10810730490504233] [Medline: 15513791]
13. Kroenke K, Spitzer RL, Williams JB, Löwe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. Psychosomatics 2009;50(6):613-621. [doi: 10.1176/appi.psy.50.6.613] [Medline: 19996233]
14. Taylor B. Methodus incrementorum directa & inversa. London, UK: typis Pearsonianis: prostant apud Gul. Innys; 1715.
15. Ye J, Ma Q. The effects and patterns among mobile health, social determinants, and physical activity: a nationally representative cross-sectional study. AMIA Jt Summits Transl Sci Proc 2021 May 17;2021:653-662 [FREE Full text] [Medline: 34457181]
16. Ye J, Ren Z. Using multivariate models to examine the impact of COVID-19 pandemic and gender differences on health and health care. medRXiv 2021 Sep 21. [doi: 10.1101/2021.09.02.21263055]
17. Zender R, Olshansky E. Women's mental health: depression and anxiety. Nurs Clin North Am 2009 Sep;44(3):355-364. [doi: 10.1016/j.cnur.2009.06.002] [Medline: 19683096]
18. Rosenquist JN, Fowler JH, Christakis NA. Social network determinants of depression. Mol Psychiatry 2011 Mar;16(3):273-281 [FREE Full text] [doi: 10.1038/mp.2010.13] [Medline: 20231839]
19. Haidt J, Allen N. Scrutinizing the effects of digital technology on mental health. Nature 2020 Feb;578(7794):226-227. [doi: 10.1038/d41586-020-00296-x] [Medline: 32042091]
20. Nesi J, Choukas-Bradley S, Prinstein MJ. Transformation of adolescent peer relations in the social media context: part 1-a theoretical framework and application to dyadic peer relationships. Clin Child Fam Psychol Rev 2018 Sep;21(3):267-294 [FREE Full text] [doi: 10.1007/s10567-018-0261-x] [Medline: 29627907]
21. Orben A. Teenagers, screens and social media: a narrative review of reviews and key studies. Soc Psychiatry Psychiatr Epidemiol 2020 Apr;55(4):407-414. [doi: 10.1007/s00127-019-01825-4] [Medline: 31925481]
22. Ye J, Li N, Lu Y, Cheng J, Xu Y. A portable urine analyzer based on colorimetric detection. Anal Methods 2017;9(16):2464-2471. [doi: 10.1039/c7ay00780a]
23. Greenwood DA, Litchman ML, Isaacs D, Blanchette JE, Dickinson JK, Hughes A, et al. A new taxonomy for technology-enabled diabetes self-management interventions: results of an umbrella review. J Diabetes Sci Technol 2021 Aug 11:19322968211036430. [doi: 10.1177/19322968211036430] [Medline: 34378424]
24. Ye J. Health information system's responses to COVID-19 pandemic in China: a national cross-sectional study. Appl Clin Inform 2021 Mar 19;12(2):399-406. [doi: 10.1055/s-0041-1728770] [Medline: 34010976]
25. Ye J, Yao L, Shen J, Janarthanam R, Luo Y. Predicting mortality in critically ill patients with diabetes using machine learning and clinical notes. BMC Med Inform Decis Mak 2020 Dec 30;20(Suppl 11):295 [FREE Full text] [doi: 10.1186/s12911-020-01318-4] [Medline: 33380338]
26. Ye J, Sanchez-Pinto LN. Three data-driven phenotypes of multiple organ dysfunction syndrome preserved from early childhood to middle adulthood. AMIA Annu Symp Proc 2021 Jan 25;2020:1345-1353 [FREE Full text] [Medline: 33936511]
27. Ye J. Advancing mental health and psychological support for health care workers using digital technologies and platforms. JMIR Form Res 2021 Jun 30;5(6):e22075 [FREE Full text] [doi: 10.2196/22075] [Medline: 34106874]
28. Hirshkowitz M, Whiton K, Albert SM, Alessi C, Bruni O, DonCarlos L, et al. National Sleep Foundation's sleep time duration recommendations: methodology and results summary. Sleep Health 2015 Mar;1(1):40-43. [doi: 10.1016/j.sleh.2014.12.010] [Medline: 29073412]
29. Consensus Conference Panel, Watson NF, Badr MS, Belenky G, Bliwise DL, Buxton OM, Non-Participating Observers, American Academy of Sleep Medicine Staff, et al. Recommended amount of sleep for a healthy adult: a joint consensus statement of the American Academy of Sleep Medicine and Sleep Research Society. J Clin Sleep Med 2015 Jun 15;11(6):591-592 [FREE Full text] [doi: 10.5664/jcsm.4758] [Medline: 25979105]
30. Ehlers CL, Gilder DA, Criado JR, Caetano R. Sleep quality and alcohol-use disorders in a select population of young-adult Mexican Americans. J Stud Alcohol Drugs 2010 Nov;71(6):879-884 [FREE Full text] [doi: 10.15288/jsad.2010.71.879] [Medline: 20946745]
31. Pauley PM, Hesse C. The effects of social support, depression, and stress on drinking behaviors in a college student sample. Communication Studies 2009 Oct 21;60(5):493-508. [doi: 10.1080/10510970903260335]
32. Kenney SR, Lac A, Labrie JW, Hummer JF, Pham A. Mental health, sleep quality, drinking motives, and alcohol-related consequences: a path-analytic model. J Stud Alcohol Drugs 2013 Nov;74(6):841-851 [FREE Full text] [doi: 10.15288/jsad.2013.74.841] [Medline: 24172110]

Abbreviations

HINTS: Health Information National Trends Survey
OR: odds ratio
PBMI: Physical and Behavioral Health, Mental Health, and Informatics

PGHD: patient-generated health data
PHQ-4: Patient Health Questionnaire–4

Edited by R Kukafka; submitted 02.06.21; peer-reviewed by W Tang, F Mughal; comments to author 14.10.21; revised version received 14.03.22; accepted 21.03.22; published 29.04.22.

Please cite as:
Ye J, Wang Z, Hai J
Social Networking Service, Patient-Generated Health Data, and Population Health Informatics: National Cross-sectional Study of Patterns and Implications of Leveraging Digital Technologies to Support Mental Health and Well-being
J Med Internet Res 2022;24(4):e30898
URL: https://www.jmir.org/2022/4/e30898
doi:10.2196/30898
PMID:35486428

Original Paper

The Effects of Health Care Chatbot Personas With Different Social Roles on the Client-Chatbot Bond and Usage Intentions: Development of a Design Codebook and Web-Based Study

Marcia Nißen[1], PhD; Dominik Rüegger[1,2], PhD; Mirjam Stieger[3,4,5], PhD; Christoph Flückiger[5], PhD; Mathias Allemand[5,6], PhD; Florian v Wangenheim[1], PhD; Tobias Kowatsch[1,7], PhD

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

[2]Pathmate Technologies AG, Zurich, Switzerland

[3]Department of Psychology, Brandeis University, Waltham, MA, United States

[4]Institute of Communication and Marketing, Lucerne University of Applied Sciences and Arts, Lucerne, Switzerland

[5]Department of Psychology, University of Zurich, Zurich, Switzerland

[6]University Research Priority Programs, Dynamics of Healthy Aging, University of Zurich, Zurich, Switzerland

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

Corresponding Author:
Marcia Nißen, PhD
Centre for Digital Health Interventions
Department of Management, Technology, and Economics
ETH Zurich
Weinbergstrasse 56/58
Zurich, 8092
Switzerland
Phone: 41 44 632 65 66
Email: mnissen@ethz.ch

Abstract

Background: The working alliance refers to an important relationship quality between health professionals and clients that robustly links to treatment success. Recent research shows that clients can develop an affective bond with chatbots. However, few research studies have investigated whether this perceived relationship is affected by the social roles of differing closeness a chatbot can impersonate and by allowing users to choose the social role of a chatbot.

Objective: This study aimed at understanding how the social role of a chatbot can be expressed using a set of interpersonal closeness cues and examining how these social roles affect clients' experiences and the development of an affective bond with the chatbot, depending on clients' characteristics (ie, age and gender) and whether they can freely choose a chatbot's social role.

Methods: Informed by the social role theory and the social response theory, we developed a design codebook for chatbots with different social roles along an interpersonal closeness continuum. Based on this codebook, we manipulated a fictitious health care chatbot to impersonate one of four distinct social roles common in health care settings—institution, expert, peer, and dialogical self—and examined effects on perceived affective bond and usage intentions in a web-based lab study. The study included a total of 251 participants, whose mean age was 41.15 (SD 13.87) years; 57.0% (143/251) of the participants were female. Participants were either randomly assigned to one of the chatbot conditions (no choice: n=202, 80.5%) or could freely choose to interact with one of these chatbot personas (free choice: n=49, 19.5%). Separate multivariate analyses of variance were performed to analyze differences (1) between the chatbot personas within the no-choice group and (2) between the no-choice and the free-choice groups.

Results: While the main effect of the chatbot persona on affective bond and usage intentions was insignificant (P=.87), we found differences based on participants' demographic profiles: main effects for gender (P=.04, η_p^2=0.115) and age (P<.001, η_p^2=0.192) and a significant interaction effect of persona and age (P=.01, η_p^2=0.102). Participants younger than 40 years reported higher scores for affective bond and usage intentions for the interpersonally more distant expert and institution chatbots; participants 40 years or older reported higher outcomes for the closer peer and dialogical-self chatbots. The option to freely choose a persona

significantly benefited perceptions of the peer chatbot further (eg, free-choice group affective bond: mean 5.28, SD 0.89; no-choice group affective bond: mean 4.54, SD 1.10; P=.003, η_p^2=0.117).

Conclusions: Manipulating a chatbot's social role is a possible avenue for health care chatbot designers to tailor clients' chatbot experiences using user-specific demographic factors and to improve clients' perceptions and behavioral intentions toward the chatbot. Our results also emphasize the benefits of letting clients freely choose between chatbots.

(J Med Internet Res 2022;24(4):e32630) doi:10.2196/32630

KEYWORDS

chatbot; conversational agent; social roles; interpersonal closeness; social role theory; working alliance; design; persona; digital health intervention; web-based experiment

Introduction

Motivation

In health care, the ongoing and active engagement of clients in their treatment and care is paramount to achieving optimal health outcomes [1] (eg, in the management of chronic diseases, such as depression and diabetes) but also to promote sustained behavioral and lifestyle changes in preventive medicine. Yet, compared to other settings, such as retail or hospitality, the health care context is arguably unique in at least two respects: (1) the highly personal, sensitive, emotional, and potentially high-stakes nature of most health matters and (2) the interpersonal nature of many therapeutic approaches and clients' encounters with various social health care roles [2].

Since personnel and financial resources of health care professionals to monitor and foster clients' engagement in their therapy are limited [3], the development of a broad range of novel digital health technologies is increasingly supporting efforts to empower clients to take charge of their own health outside clinical settings at every stage of the patient journey. Among these digital innovations, health care chatbots (ie, text-based conversational agents), in particular, have been received with enthusiasm. Chatbots are said to simplify and "humanize" access to digital health care services, even in longitudinal settings [4,5], since they mimic natural physician-patient dialogue, where clients engage in natural, text- or voice-based interpersonal exchanges via messenger-like user interfaces [6]. Taking over the role of "digital coaches" [7], they can develop an *affective bond* with clients and support them in their everyday lives and therapeutic settings beyond on-site consultations anywhere and anytime [8]. However, maintaining engagement with a chatbot aimed at fostering patient engagement brings new challenges and raises the question of which design choices foster engagement and an affective bond with the chatbot in the first place [9,10].

Originally developed for psychotherapeutic settings, the working alliance—often also referred to as the therapeutic or helping alliance—is a key construct in the therapist-client collaboration in the clinical context of mental health disorders, reflecting the collaborative quality between clients and health professionals [11]. The working alliance is robustly linked to patient engagement, retention [12,13], and therapeutic success [14,15] and encompasses three subdimensions [16]: *therapeutic goals* and *therapeutic tasks* that the health professional and client jointly agree upon "in the context of an *affective bond* or positive attachment" [17]. This affective bond subsumes a sense of sympathy, interpersonal closeness, familiarity, trust, or common purpose and understanding between a mental health professional and a client [14]. In social psychology, bond, relatedness, attachment, intimacy, and closeness between two social actors are often subsumed under the psychological concept of interpersonal closeness [18]. The development of interpersonal closeness and a sustainable affective bond between mental health professionals and clients strongly depends on continuous personal encounters in either face-to-face psychotherapy [14,15], online therapy [19,20], or remote consultations [21].

A recent study among 36,070 users of the text-based chatbot Woebot found that clients may develop a working alliance with a chatbot within 5 days [22]. However, while a body of work describes how or which design choices affect the development of an affective bond with *embodied* conversational agents (which can use more nonverbal cues such as mimicry and gestures to convey interpersonal closeness [23]), it is still unclear which design choices foster an affective bond with *text-based* chatbots [24].

Research under the social response theory—also known as the *computers are social actors* paradigm or the media equation—has demonstrated that individuals anthropomorphize conversational agents that exhibit human-like social cues [25], treat them as social actors [26], and are capable of developing relationships with them [22]. Previous work suggests that a chatbot's social role constitutes one such cue that could "cause people to make inferences about social presence in a computing product" [27], which could help elicit social responses [28].

However, in practice, chatbots are often mindlessly designed "to perform social roles traditionally associated with humans" [29], for instance, a therapist role, in the case of Woebot, or a nurse role, in the case of the Florence chatbot. Theoretically, however, it is not yet well understood how a chatbot's impersonated social role actually affects a client's relationship with the chatbot and, thus, the affective bond between the client and the chatbot and the client's intentions to use the chatbot.

With this study, we aimed to close this gap by, first, investigating which design choices allow the manifestation of the social role of a chatbot (research question 1) and how a chatbot's social role affects users' affective bond with the chatbot and their intentions to use it (research question 2). Second, we explored how an individual's demographic profile (ie, gender and age; research question 3) and the option to freely

choose the social role of a chatbot affect these evaluations (research question 4).

Designing Engaging Health Care Chatbots With Human-Like Social Roles

Social role theory "concerns itself with a triad of concepts: patterned and characteristic social behaviors, parts or identities that are assumed by social participants, and scripts or expectations for behavior that are understood by all and adhered to by the performer" [30]. Social role theory has long been applied in health care to disentangle communication and power dynamics in the physician-client relationship [31,32] and to better understand how specific social communication scripts are used by physicians and clients to navigate their different social roles and interactions [33-37].

Considering chatbots as social actors (cf, social response theory), individuals can be expected to apply readily available learned human social scripts to interactions with a chatbot as well, especially when specific cues signal that it enacts a particular role [27,38]. Prior conceptual conversational agent studies have developed taxonomies, typologies, and classifications of different types of chatbots, for example, differentiating chatbots for domain-specific or for general-purpose use [39], for specific applications (eg, business-to-business customer services [40] and health care [41]), for different purposes [42,43], for single- or multiple-user use cases [44,45], or for specific periods [46]. However, relatively few conceptual studies to date have addressed how a chatbot can impersonate a holistic, domain-specific social role and how such a social role affects user assessments. In e-commerce contexts, McGoldrick et al [47] identified and investigated three possible roles for virtual sales agents (ie, friend, personal buyer, or helper) and found that the helper role was most widely preferred, followed by the friend role. Similarly, Rhee and Choi [48] investigated effects of two possible roles (ie, friend or secretary) for a voice-enabled recommendation agent and found that consumers rated a product more favorably when it was recommended by a friend role. Another study on perceptions of Apple's voice assistant Siri found that participants expressed more favorable attitudes toward Siri when they thought of it as a coworker versus a supervisor or a friend [49]. In digital health care settings, however, we are not aware of any previous study that has investigated how individuals' attitudes toward social roles that are impersonated by a health care chatbot promote or antagonize the development of an affective bond with the chatbot and their intention to use it.

The American Psychological Association Dictionary of Psychology defines a social role as "the set of attitudes and characteristic behaviors expected of an individual who occupies a specific position or performs a particular function in a social context, such as being a spouse or acting as a caregiver for an aging parent" [50]. To this end, social role theory suggests that the perception of a particular social role is triggered by a set of role-typical social cues (eg, use of jargon and business attire) that individuals subconsciously look for to orient themselves and to understand potential outcomes or goals of a relationship with one another [51]. To fulfill a social role in a particular encounter—be it a client with a therapist, a nurse with a client, a boss with employees, etc—people mindlessly exhibit ritualized social behaviors that they have learned to be appropriate for that type of encounter [51].

Typical social roles that clients encounter in health care settings in the real world include, for example, the medical experts of their condition, supportive peers with the same condition, the institution providing the context for their health care services, and the clients themselves who have to adopt their own social roles to this situation. The conceptualization and operationalization of the latter role draws from people's tendency to talk to themselves through internal dialogues (ie, to their "dialogical selves"). This is because intrapersonal communication is known to provide self-regulatory functions through monological, goal-directed self-talk (eg, "I believe in you!") or more contemplative and reflective internal dialogues with one or more imaginary interlocutors [52]. Overall, these social roles can be conceptualized on an interpersonal closeness continuum, ranging from extremely distant (ie, the institution), rather distant (ie, expert), and rather close (ie, peer) to extremely close (ie, the client's dialogical self) social roles.

Social, interpersonal closeness cues that are available to conversational agents include (1) visual cues (eg, the avatar), (2) verbal style cues (eg, form of address), (3) nonverbal style cues (eg, emojis), and (4) verbal relational content cues (eg, self-disclosures and jokes) [53]. To provide an overview of empirical studies of interpersonal closeness cues and their effects, we conducted a systematic analysis of the literature to date. Specifically, 116 research articles that were included in two recently published systematic literature reviews on design features of embodied conversational agents [23] and of text-based chatbots [24] were analyzed with regard to the types of cues and the outcome variables investigated in the included studies, respectively (Tables S1 and S2 in Multimedia Appendix 1). Articles were included for detailed analysis if they presented at least one empirical or experimental study that investigated the effects of interpersonal closeness cues on user assessments or the client-agent relationship. Studies that, for example, compared a chatbot with a website or assessed the effects of overall system quality on users' evaluation of conversational agents were excluded from the analysis. In total, 47 studies could be analyzed: 32 studies on embodied conversational agents and 15 studies on text-based conversational agents. Tables S1 and S2 in Multimedia Appendix 1 yield descriptive details of all included studies and the classification result.

Quantitative analysis has demonstrated that the majority of all studies (76.6%) either investigated the effects of visual cues, such as abstract versus human avatars [54]; role-appropriate attire [55,56]; gender, age, or ethnicity of the embodied character (eg, Alsharbi and Richards [57]); or nonverbal style cues, such as empathic versus nonempathic facial expressions [58,59] or emoji (eg, Beattie et al [60] and Fadhil et al [61]). Another considerable share representing 46.8% of all studies analyzed the effects of verbal relational content cues, such as self-disclosures (eg, Ho and Hancock [62] and Kang and Gratch [63]) or empathic feedback (eg, Liu and Sundar [64]). Similarly, 44.7% of the articles investigated verbal style cues, such as the form of address (eg, Bickmore and Picard [65]), T/V (*tu/vos*) distinction (ie, formal and informal forms of the second-person

pronoun "you" [66]), or paralinguistic and backchanneling cues [67].

While some of these articles investigated combinations or interactions of such cues, no previous study has investigated how they can be used to design holistic chatbot personas that are modeled to impersonate a particular human social role based on the adaption of various interpersonal closeness cues.

By unifying social role theory and social response theory at this point, *hypothesis 1* is as follows: It is expected that health care chatbots can be designed such that they impersonate social roles of differing closeness by using combinations of purposefully selected and differently manifested interpersonal closeness cues.

Given that the interpersonal closeness construct is closely related to the affective bond subgoal of the working alliance, and given that interpersonal closeness can be expected to be at least partially determined by the social role that a chatbot adopts in the encounter with a client, *hypothesis 2* is as follows: It is expected that chatbots that impersonate a closer social role (ie, peer or dialogical self), as compared to a more distant social role (ie, expert or institution), will improve perceived interpersonal closeness, the affective bond with the chatbot, and intentions to use the chatbot.

Personalization and Customization of Health Care Chatbots

While personal characteristics of an individual, such as demographic factors [68] or their innate tendency to anthropomorphize objects [69], may decisively influence their capacity to experience interpersonal closeness, in chatbot design, all too often "one-fits-all approaches" still prevail [70]. Yet, personalization and customization are common strategies that companies worldwide are adopting to account for personal characteristics or preferences [71,72].

Personalization refers to the automatic tailoring of service offerings, for instance, to preferences, past usage behaviors [73], or demographic characteristics of customers [70,74], "usually based on previously collected customer data" [72]. While the personalization of chatbots to user characteristics is still in its infancy, it has already been linked to greater user satisfaction, user engagement, and dialogue quality [75] and has been deemed as important for adapting conversational agents to the changing needs of patients through the various stages of a disease [76].

Given that demographic characteristics (ie, age and gender) have been found to affect user assessments of conversational agents in inconsistent ways [47,68,74], *hypothesis 3* is as follows: It is expected that the demographic profile (ie, gender and age) will affect perceived interpersonal closeness, the affective bond with the chatbot, and intentions to use the chatbot.

Customization occurs when individuals have the opportunity to proactively choose between different options, such as different chatbot roles or specific design elements within them [72]. It is generally known for other technologies that the imposed use of self-service technologies (eg, imposed information retrieval via ticketing machines or the internet) has a strong negative impact on users' evaluations and intentions to adopt such a technology [77] and that customization options are linked to greater customer satisfaction and loyalty [78].

Although the link between free choice and perceptions or relationship-building processes with conversational agents has not previously been established, *hypothesis 4* is as follows: It is expected that free, versus imposed, choice will increase, versus decrease, users' perceived interpersonal closeness, the affective bond with the chatbot, and intentions to use the chatbot.

A Design Codebook for Chatbots With Different Social Roles

Overview

To answer our first research question and as a prerequisite to empirically assess the effects of chatbots' impersonated social roles on perceived interpersonal closeness, the affective bond, and the intention to use them, we reviewed literature from social psychology, communication, and human-computer interaction research to develop a design codebook for chatbots with different social roles. A prior version of the design codebook and the study design have been presented at the European Conference on Information Systems 2018 (ECIS 2018) and published as research-in-progress work in the conference proceedings [79].

Given our focus on chatbots that operate in health care settings and given that the affective bond plays a central role in the physician-client relationship [80], we were eager to develop design guidelines that would allow us to manipulate interpersonal closeness with chatbots. Overall, we modeled four distinct chatbot personas embodying typical social roles that clients encounter in health care settings on the proposed interpersonal closeness continuum; the personas were as follows: an institution-like chatbot, an expert-like chatbot, a peer-like chatbot, and a dialogical self–like chatbot. We define chatbot personas as "composite archetypes" of their corresponding human social roles impersonated by the chatbots by different combinations and manifestations of selected interpersonal closeness cues [81].

As a base for the development of a design codebook for chatbots with different social roles, we included seven out of eight suggested "behaviors" from the framework of relational behaviors for embodied conversational agents by Bickmore and Picard [65]: form of address, politeness, social dialogue, meta-relational dialogue, empathy exchanges, humor, and continuity behaviors. We did not include nonverbal immediacy behaviors, which are only applicable to embodied conversational agents. We then compared Bickmore and Picard's [65] framework against the taxonomy of social cues by Feine et al [53] to organize all interpersonal closeness cues into four groups (ie, visual cues, verbal style cues, nonverbal style cues, and verbal relational content cues) and included six new cues from their taxonomy (ie, avatar, display name, professional jargon, emojis, greeting and farewell behavior, and self-disclosures). In total, 13 interpersonal closeness cues were included in the design codebook for chatbots with different social roles (Figure 1 [53,60,65,74,82-89]).

To develop the chatbot personas, we followed Bickmore and Picard's [65] approach and defined how these cues would manifest for each chatbot social role, having in mind the social role theory, social response theory, and the interpersonal closeness continuum along which we had allocated the four social roles. In the following sections, we will outline, per cue category, why we included a particular interpersonal closeness cue and how we derived the different manifestations of each interpersonal closeness cue per chatbot social role.

Figure 1. Design codebook for chatbots with different social roles. A research-in-progress version of this design codebook had been presented at the European Conference on Information Systems 2018 and published in the conference proceedings. The version in this paper represents the latest version. The T/V distinction design cue is only applicable to chatbots that operate in a language in which different pronouns are used for different social contexts; stated pronouns here are in German. T/V: *tu/vos*.

Chatbot persona	Institution	Expert	Peer	Dialogical self
Interpersonal closeness	Low			High
Visual cues				
Avatar [53,82,83]	 Institutional logo	 Avatar, older looking, professional looking	 Avatar, more ordinary looking	 Real photo of the client or silhouette
Verbal style cues				
Display name [53,65]	Coach	Dr. Change	Milo/Miro Mila/Mira	Participant name
Professional jargon [83]	Yes, professional jargon	Yes, professional jargon	No, colloquial lay language	No, colloquial lay language
Form of address [53,65,84]	Nonspecific	Professional, honorific (Dr) + surname	Personal, first name	Personal, first name
T/V distinction [65,85]	Second person, V pronoun ("Sie")	Second person, V pronoun ("Sie")	Second person, T pronoun ("du")	First person pronoun ("ich" or "wir")
Nonverbal style cues				
Emojis [53,60]	No	No	Yes	Yes
Verbal relational content cues				
Greeting and farewell behavior [53,65]	Formal	Formal	Informal	Informal
Social dialogues [65]	No	Infrequent, once a week	Frequent, every day	Frequent, every day
Meta-relational talk [65]	No	No	Yes	Yes
Reciprocal self-disclosure [53,86]	No	No	Yes	Yes
Empathy exchanges [65,74,87]	No	Yes, compassionately detached, clinical empathy	Yes, emotionally involved empathy	Yes, emotionally involved empathy
Humor [65,88]	No	No	Yes	Yes
Continuity behaviors [65,89]	No	Infrequent, once a week	Frequent, every day	Frequent, every day

Visual Cues

Since our research focuses on chatbots that operate in a text-messaging format, we only considered one visual cue: a static avatar. Manipulating the graphical representation of the chatbot persona allowed us to leverage findings from prior research. This research found priming effects of visual cues triggering social and contextual psychological and behavioral reactions [83], which can reduce uncertainty about the chatbot's role and one's own role in an initial interaction, similar to human-human initial encounters, where individuals rely on visual cues to form first impressions of one another [82]. Accordingly, we designed the avatars to convey stereotypical features of their roles by altering age and accessories (eg, glasses) of the avatars to make them look more professional (ie, expert) or ordinary (ie, peer). For the institution-like chatbot, a logo of the university, where the chatbot prototype had been developed, was used. In the dialogical-self condition, participants had the chance to either upload a photo of themselves or to use a default representation of a gendered human-resembling silhouette.

Verbal Style Cues

Verbal style cues are a key aspect for conveying important social and contextual information and, therefore, for facilitating social understanding [90] and, thus, for the creation of artificial agents [53]. Accordingly, we included three verbal style cues available to text-based chatbots: form of address, professional jargon, and T/V distinction [66].

Form of address is considered a sociolinguistic cue that verbally conveys the degree of formality and politeness imposed by the relational closeness between two interlocutors, a phenomenon known as "social deixis" and widely studied in pragmatics and sociolinguistics research [85]. Accordingly, we chose a professional formal address for the expert (ie, honorific [eg, Prof or Dr] + surname; here, "Dr. Change") compared to the more personal, informal form of address by the peer version and the dialogical-self version, who would address users by their first name only (eg, "Hey David"). The institution version was designed to completely avoid the use of direct address and used indirect and passive constructions (eg, "Please indicate how you feel today").

Similar to the form of address, T/V distinction is considered as another socially deictic cue [85]. This cue refers to the use of distinct second-person pronouns to denote less (ie, T form, Latin: "*tu*") or more (ie, V form, Latin: "*vos*") formality, power, or intimacy, depending on the relational status between two interlocutors [66]. Accordingly, since the study was conducted among a German-speaking population, the institution and expert chatbot used the German V form "*Sie*," whereas the peer and dialogical-self chatbot used the German T form "*du*."

Professional jargon refers to the use of a learned and shared, specialized language used within specific occupational groups, which can carry social and relational information about its generator [83]. Accordingly, we implemented the institution- and expert-like chatbots using more abstract, medical language (eg, translated from German: "Please read the following instructions for today's exercise carefully" and "Thank you very much, I will put the note in your files"), whereas the peer- and dialogical self–like chatbots used more colloquial and lay language (eg, "Then I'll explain today's task to you right away" and "Thank you, got it!").

Nonverbal Style Cues

Generally, computer-mediated communication lacks the possibility to convey classical nonverbal style cues, such as gestures or gaze. However, since nonverbal cues are "vital to interpersonal processes [to convey and interpret] feelings and attitudes" [60], we included emojis as available nonverbal cues available in computer-mediated communication. Prior research has shown that the use of emojis (ie, icons displaying emotions, also known as "smileys") has the potential to surrogate other nonverbal cues [60] and that they are indicative of the development of close interpersonal relationships [91]. Accordingly, the peer and dialogical-self chatbots will make use of emojis, whereas the institution and expert chatbots will not.

Verbal Relational Content Cues

Five verbal relational content cues were directly adopted from Bickmore and Picard's [65] framework of relational behavior: social dialogue, meta-relational talk, empathy exchanges, humor, and continuity behaviors. Additionally, we included greeting and farewell behavior and reciprocal self-disclosure as a potential strategy to increase interpersonal closeness, which Bickmore and Picard [65] had discussed but not integrated into their framework. Yet, Feine et al [53] included them in their taxonomy as well.

Greeting and farewell behavior directly reflects the degree of formality and politeness imposed by the social relationship between two interlocutors (see Laver [84] as cited in Bickmore and Picard [65]) and, accordingly, was manipulated gradually from very formal (ie, institution) to very informal (ie, dialogical self).

Social dialogue refers to the use of small talk, which, "on the surface, [may] not seem to move the dialogue forward at all" but is essential "to how humans obtain information about one another's goals and plans and decide whether collaborative work is worth engaging in at all" [92]. Since the depth and breadth of social dialogue are indicative of the level of trust and familiarity between two interlocutors [92], the institution-like chatbot will not engage in any social dialogue, the expert-like chatbot will engage in it only once, and the peer- and dialogical-self chatbots will use it frequently to transition between more task-oriented talk and goal-oriented talk.

Meta-relational talk entails communication about the relationship, such as "discussing the nature of the relationship [or] disclosing one's desires for the relationship" [93]. Research comparing the use of meta-relational talk between friends, lovers, relatives, and others found that the more intimate the relationship, the more individuals talk *about* their relationship [94]. We, thus, included meta-relational dialogues only in the peer and dialogical-self chatbots.

Reciprocal self-disclosure refers to the process of reciprocally revealing increasingly personal and intimate information about

oneself (eg, personal experiences, beliefs, and values). Social penetration theory posits that relationships between humans progress based on how much (ie, breadth) and how intimately (ie, depth) two interlocutors reciprocally disclose information to each other [95]. Hence, self-disclosing was found to be closely linked to liking in human-human relationships [96] as well as in human-chatbot interactions [26]. Accordingly, we included self-disclosures by the chatbots and opportunities for the users to disclose something about themselves in the peer and dialogical-self chatbots but not in the institution and expert chatbots.

Empathy exchanges refer to conveying a sense of understanding and warmth and have been described as "one of the core processes in building and maintaining relationships" [65]. However, while the clinical empathy afforded by physicians is, at best, of a professional nature characterized by "compassionate detachment...keeping sympathy at a reasonable distance to maintain an emotional balance" [97], empathy exchanges in intimate relationships are characterized by more emotional involvement and labor [98]. Accordingly, we model the expert chatbot to display a certain degree of clinical empathy and the peer and dialogical-self chatbots to express empathic concerns in an emotionally more involved fashion.

Humor refers to the use of "incongruous [comments] that [are] recognized by the receiver as an attempt to amuse and that succeeds at amusing" [88]. Since humor and laughing are known to shape interpersonal bonds and liking [99], we manipulated the peer and dialogical-self chatbots to make some self-directed, innocent jokes.

Lastly, continuity behaviors refer to actions aimed at establishing and "[maintaining] a sense of persistence in a relationship" [65], for example, by making retrospective references to the past (eg, "Last time we...") or prospective statements about future encounters (eg, "Next time we..."). Since continuity behaviors are closely linked to relational satisfaction [89] and differ in type and frequency between strangers and friends [100], the peer and dialogical-self chatbots will always (ie, on the simulated day of the intervention in the study) use adequate continuity statements, and the expert chatbot will do so occasionally with less emphasis.

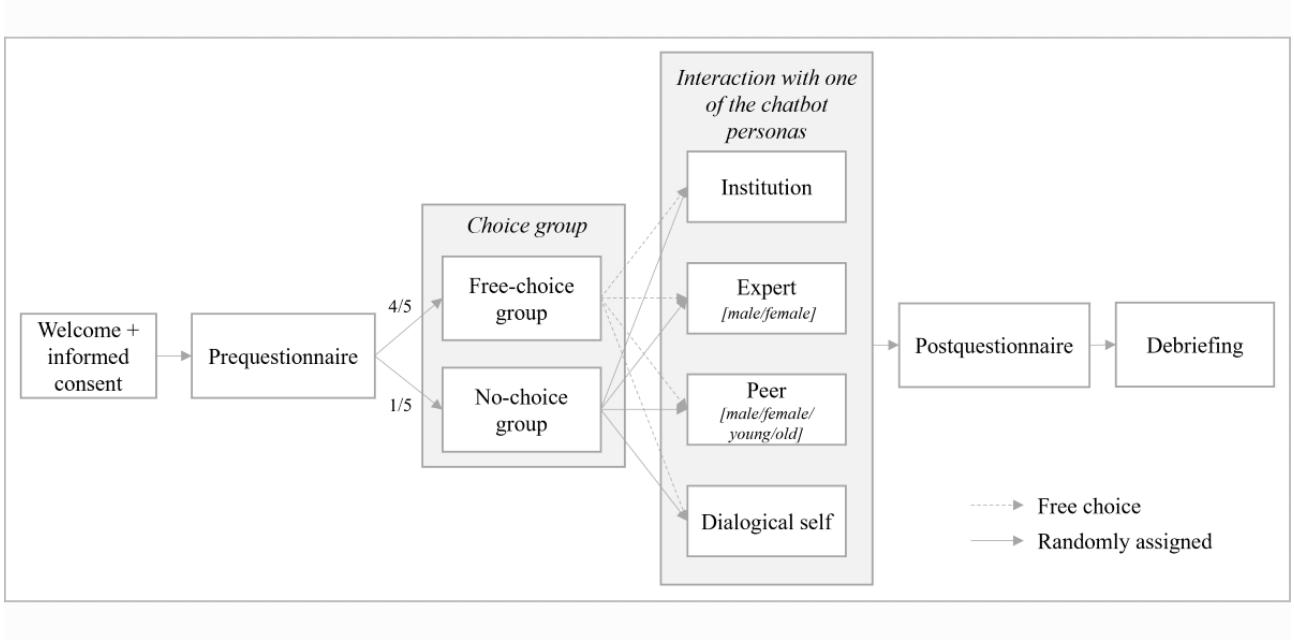
Methods

Study Design

We conducted a web-based experiment to investigate the effects of a health care chatbot's social role (ie, hypothesis 2), the role of participants' demographic profiles (ie, gender and age; hypothesis 3), and free choice of a chatbot persona (ie, hypothesis 4) on perceived interpersonal closeness, the development of an affective bond, and intentions to use a chatbot. Following the Checklist for Reporting of Results of Internet E-Surveys [101], we will outline the study design and procedure in detail.

The experimental design corresponded to a 4 (chatbot personas: institution, expert, peer, or dialogical self) × 2 (participant gender: female or male) × 2 (participant age: younger or older) between-group design. The sample was further divided into two experimental subgroups: (1) 4 out of 5 (80%) participants were randomly assigned to the "no-choice group," in which they were randomly assigned to one of the chatbot conditions, introduced to the scenario (Figure S1 in Multimedia Appendix 1), and then presented with the role description (Table 1), and (2) 1 out of 5 (20%) participants were assigned to a "free-choice group," in which they could read the descriptions first and then freely choose a chatbot persona themselves thereafter (Figure S2 in Multimedia Appendix 1). Both groups were then asked to interact with their assigned or self-chosen chatbot. Figure 2 illustrates the study design.

Figure 2. Study design. Participant numbers are listed on the arrows going into the choice groups.



Procedure

Participants were recruited via the online panel provider Talk Online Panel and rewarded based on the provider's points-based incentive system to compensate for their efforts. The entire study was conducted in Germany in July 2017. Participants were sent an anonymous link to the closed survey via email by the panel provider. After providing informed consent regarding the study conditions, participants were screened for eligibility (ie, 18 years of age or older and native German speakers).

After completing an introductory questionnaire asking for demographic and socioeconomic data, participants were introduced to a scenario (Figure S1 in Multimedia Appendix 1) in which they were asked to test a health care chatbot prototype that promoted a personality change intervention. A digital lifestyle intervention allowed us to work with a more heterogeneous sample of healthy individuals who did not have to imagine themselves being chronically ill.

After the interaction with the respective chatbot, participants were redirected to complete the postquestionnaire, with all outcome variables and manipulation checks, and debriefed as to the experiment's purpose. On average, participants spent 15 to 62 minutes completing the survey (mean 30.23, SD 9.13 minutes).

Development of the Experimental Stimuli

The experimental stimuli were designed based on a prototype of a fictitious health care chatbot promoting a personality change intervention adapted from Stieger et al [102]. For purposes of standardization, the interaction with the chatbot was purely text based (ie, no voice input or output) and followed a rule-based conversational script with predefined answer options. The content of the conversation was based on fictitious scripts for the "first 2 days" and the "last day" of the intervention, which allowed us to encompass essential elements of a comprehensive health intervention (eg, introduction, goal agreement, task agreement, and feedback). All coaching elements were designed in an expert coaching style, which refers to a type of instructional coaching by an experienced coach who helps a "coachee" through providing her or him "with self-assessment tools and constructive feedback" [103]. A bidirectional peer coaching–style intervention would require similar levels of experience and mutual, reciprocal learning goals [103], which does not apply to our scenario.

We first created a minimum viable "skeleton" version of the conversational script encompassing exemplary interventional elements. In a second step, to induce the chatbots' respective social roles, we systematically manipulated the script with respect to the manifestations and frequency of the interpersonal closeness cues defined in the conceptually derived codebook (Figure 1). Except for the different avatars, the experimental manipulation only concerned adaptations of the conversational script by adding, editing, or deleting statements and specific terms or emojis.

In the survey, all personas were introduced as follows (translated into English): "[chatbotName] is a 'digital coach' developed by researchers, experts, and psychotherapists from the University of Zurich, which has been equipped with various skills based on the latest findings from many years of research on 'personality development.'" This sentence was followed by individual role-specific descriptions of the personas, as visible in Table 1.

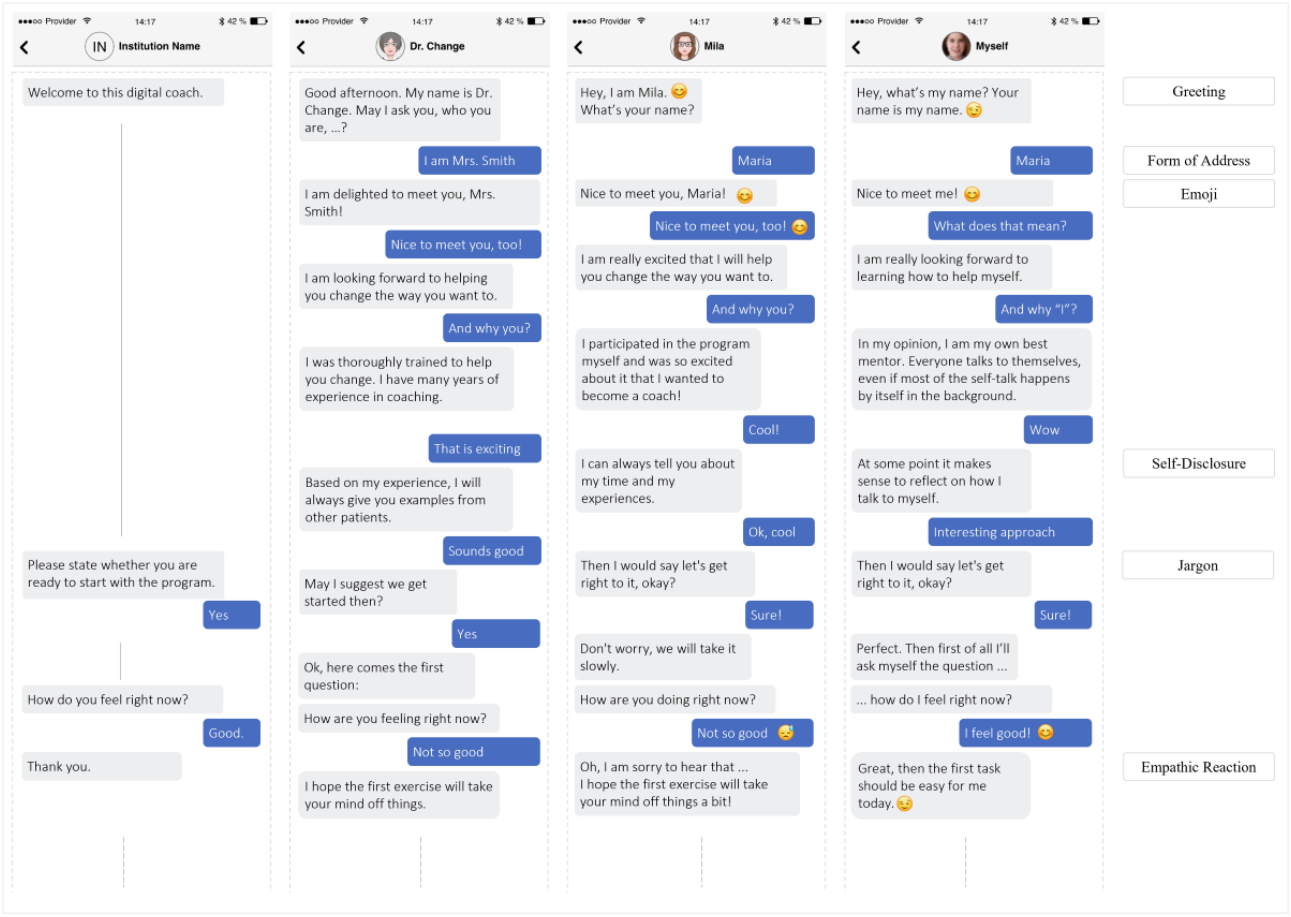
With most interpersonal closeness cues being absent in the institution condition, the script in this condition was significantly shorter than in all other conditions (institution: 51 chatbot statements and 32 user responses; expert: 73 chatbot statements and 46 user responses; peer: 77 chatbot statements and 48 user responses; and dialogical self: 74 chatbot statements and 45 user responses; χ^2_3=10.5, P=.02). Still, an analysis of variance (ANOVA) revealed that there were no significant differences (P=.35) between treatment groups with regard to the average interaction time spent with the respective chatbot (mean 16.0, SD 7.11 minutes). Since interaction with the chatbot was self-paced and unsupervised, this speaks for participants' comparable involvement with all chatbots across conditions. Figure 3 depicts excerpts of the onboarding conversations with the four chatbot personas.

Table 1. Chatbot personas and their introductions within the conversations.

Chatbot persona	Introduction[a]
Institution	"The PersonalityCoach has been programmed to represent the Psychological Institute of the University of Zurich, which enjoys a very good reputation worldwide in the field of personality research."
Expert	"Dr. Change has been programmed to represent a professional psychotherapist who has many years of experience in the field of personality coaching."
Peer	"Milo/Mila has been programmed to represent a peer who had once been a participant in the PersonalityChange program him/herself and who will share his/her own experiences with you."
Dialogical self	"Your MySelfCoach has been programmed to give you the impression of talking to yourself through simulated inner dialogues and, thus, to help you see your experiences in a new light."

[a]Translated into English by the authors for this paper.

Figure 3. Exemplary excerpts of participants' onboarding conversations with the four chatbot personas. The content in this figure has been translated into English by the authors for this paper.



Technical Implementation

The whole study was implemented with the web-based survey tool SoSci Survey. For full customization of the health care chatbots, we built our own prototype chat app with the open-source software platform MobileCoach [104]. The chat app was then deployed in a web-based smartphone simulator using the virtualization service Appetize.io. Hence, the app could be displayed within the questionnaire but on a dedicated, separate survey page; participants did not have to download and install an app on their phone but could focus on the web-based interaction and seamlessly continue within the survey. For a screenshot of the study environment and the study prototype see Figure S3 in Multimedia Appendix 1.

Usability tests were conducted by one computer engineer, two chatbot researchers, and two domain experts who qualitatively assessed and confirmed the realism of the chatbot personas and the study design. Moreover, the complete study setup was pretested with 22 previously uninvolved individuals from the authors' networks over two iterations to ensure comprehensibility and technical functionality.

Sample Characteristics

The power analysis was informed by the few existing studies in the field on embodied conversational agents that had investigated the effects of relational cues on affective bond and intention to use [65]. A priori power analysis using G*Power (version 3.1; Heinrich Heine University Düsseldorf) suggested that we would need to recruit 250 participants to find medium-sized effects (f=0.25) in ANOVAs of between-group fixed effects at the α=.05 level of significance and statistical power of 0.80. After data runs, 251 responses were included in the final analyses (free-choice group: n=49, 19.5%; no-choice group: n=202, 80.5%). Participants ranged in age from 19 to 65 years (mean 41.15, SD 13.87 years), and 57.0% (143/251) of the total sample were female.

Measurements

All measurements were adapted from established multi- or single-item scales; perceived interpersonal closeness was measured using the Inclusion of Other in the Self scale [18], an established and reliable pictorial instrument to measure the subjectively perceived closeness of a relationship [105]. Affective bond (Cronbach α=.900) was measured based on the bond subscale of the Working Alliance Inventory for technology-based health care interventions [106]. Intention to use (α=.952) was measured based on a scale adapted from the technology acceptance model [107]. The full list of measurements is provided in Table S3 in Multimedia Appendix 1. Manipulation check items were measured at the design cue level and carefully drafted and pretested to capture perceptions of each manipulated interpersonal closeness cue individually (Table S4 in Multimedia Appendix 1). Demographics encompassed participants' age, gender, and native language.

Statistical Analysis

Where constructs consisted of multiple items, reliability analyses were carried out to discern Cronbach α with all constructs scoring greater than the .70 threshold [108]. To test the main and interaction effects of the chatbot persona, participant gender, and participant age on all outcome variables (ie, hypotheses 2 and 3), one multivariate ANOVA (MANOVA) with type III sum of squares was conducted, with partial eta squared (η_p^2) indicating the size of the effect. A second MANOVA was conducted to test the effects of choice type, participant gender, and participant age on all outcome measurements (ie, hypothesis 4). Participant age was dummy coded using median split: 1=younger than 40 years (n=122, 48.6%) and 2=40 years or older (n=129, 51.4%). Where significant effects were discerned, univariate ANOVAs per outcome measurement were conducted, followed by Bonferroni-adjusted pairwise comparisons following the procedure described by Field et al [109]. Due to the mixed levels of our hypotheses on main and interaction effects, where significant interactions occurred, we investigated the highest-order interactions and not the lower-order interactions or main effects [110], also following guidance outlined in Field et al [109].

Ethical Considerations

To meet ethical standards, we applied the following procedures. Before the study started, all participants received written information about the research project, benefits, and risks of participation. Furthermore, they were informed about their right to withhold or revoke their consent without giving reasons, their right to withdraw from participating in the study at any time during the study, and their right to receive information at any time in response to further questions when contacting the study team. They also received transparent communication about the main sources of financing for the research project as well as transparent communication that the chatbot they would be interacting with during the study was only a minimal viable prototype. Informed consent was obtained before assessment and before interaction with the prototype. At the end of the study, participants were debriefed regarding the study's actual purpose.

According to the ETH Zurich's Ethics Commission's Compliance Guide regarding human subject research [111], this study did not require ethics approval for the following reasons:

- Besides participants' age and gender, we did not collect personal information. Other social and behavioral data collected in our study were collected completely anonymously. Age and gender information are always only reported in aggregated, anonymous ways.
- Vulnerable or dependent groups were explicitly not included.
- Experimental manipulation on the chatbot did not affect functional aspects of the intervention, but only affected style- and design-related features of the chatbot.
- Experimental manipulations being researched were not likely to upset or disturb participants and did not use socially sensitive topics as a basis for the scenario development.

Results

Manipulation Check

Separate ANOVAs confirmed significant differences between the chatbot personas for the form of address (P=.02, η_p^2=0.052), professional jargon (P=.006, η_p^2=0.063), T/V distinction (P<.001, η_p^2=0.578), small talk (P=.003, η_p^2=0.070), self-disclosure (P<.001, η_p^2=0.289), use of emojis (P<.001, η_p^2=0.320), humor (P<.001, η_p^2=0.173), and meta-relational talk (P=.003, η_p^2=0.072) in the intended directions, but not for empathy exchanges (P=.31, η_p^2=0.018) or greeting (P=.75, η_p^2=0.006). Figure S4 (a-j) in Multimedia Appendix 1 illustrates the perceived differences between the chatbot personas. Since the manipulation of the interpersonal closeness cues worked as intended, except for empathy exchanges and greeting, the results provide partial support for hypothesis 1.

No-Choice Group: Effects of Randomly Assigned Chatbot Personas on Outcome Measures

Main and Interaction Effects

The MANOVA model specified with chatbot persona, participant gender, and participant age on all outcome variables (ie, interpersonal closeness, affective bond, and intention to use) showed no significant main effects for chatbot persona (P=.88) but did so for participant age (Wilks λ=0.897, $F_{3,184}$=7.010, P<.001, η_p^2=0.103) and participant gender (Wilks λ=0.952, $F_{3,184}$=3.095, P=.03, η_p^2=0.048). Moreover, a significant interaction effect was found for chatbot persona and participant age (Wilks λ=0.887, $F_{9,448}$=2.518, P=.01, η_p^2=0.040), showing small effect sizes. Table 2 shows an overview of the MANOVA results.

As the MANOVA discerned significant effects, separate univariate ANOVAs were specified with the same factors as before for each outcome variable following the procedure described in Field [108].

These ANOVAs showed a significant main effect of participant gender on perceived interpersonal closeness ($F_{1,186}$=5.923, P=.02, η_p^2=0.031) and affective bond ($F_{1,186}$=8.081, P=.005, η_p^2=0.042), and they showed a significant main effect of participant age only on intention to use ($F_{1,186}$=4.528, P=.04, η_p^2=0.024).

The interaction effect of participant age and chatbot persona was significant for perceived interpersonal closeness ($F_{1,186}$=3.046, P=.03, η_p^2=0.047) and affective bond ($F_{1,186}$=4.836, P=.003, η_p^2=0.072) but not for the intention to use the chatbot (P=.10). An excerpt of the ANOVA results for the significant effects is depicted in Table 3. Insignificant effects in the ANOVAs were not further analyzed in the post hoc tests described in the following section.

Table 2. Multivariate test results for the specified multivariate analysis of variance model.

Effect	Wilks λ	*F* test (*df*)	*P* value	η_p^2
Main effect				
Chatbot persona	0.976	0.490 (9, 448)	.88	0.008
Participant age (<40 years old)	0.897	7.010 (3, 184)	<.001	0.103
Participant gender	0.952	3.095 (3, 184)	.03	0.048
Two-way interaction effect				
Chatbot persona × participant age	0.887	2.518 (9, 448)	.008	0.040
Chatbot persona × participant gender	0.969	0.655 (9, 448)	.75	0.011
Participant age × participant gender	0.988	0.756 (3, 184)	.52	0.012
Three-way interaction effect				
Chatbot persona × participant gender × participant age	0.968	0.670 (9, 448)	.74	0.011

Table 3. Analysis of variance (ANOVA) model results for the main effects of participant gender and participant age and two-way interaction between chatbot persona and participant age.

Independent variable[a]	Dependent variable								
	Interpersonal closeness			Affective bond			Intention to use		
	F test (*df*)	*P* value	η_p^2	*F* test (*df*)	*P* value	η_p^2	*F* test (*df*)	*P* value	η_p^2
Participant gender	5.923 (1, 186)	.02	0.031	8.081 (1, 186)	.005	0.042	2.170 (1, 186)	.14	0.012
Participant age	2.952 (1, 186)	.09	0.016	1.094 (1, 186)	.30	0.006	4.528 (1, 186)	.04	0.024
Chatbot persona × participant age	3.046 (3, 186)	.03	0.047	4.836 (3, 186)	.003	0.072	2.099 (3, 186)	.10	0.033

[a]Only factors that were significant in the multivariate ANOVA were analyzed in the ANOVAs.

Pairwise Comparisons

Pairwise comparisons using Bonferroni correction showed that, regardless of the chatbot persona, superior outcomes were consistently generated for female participants; for instance, female participants reported 0.450 points higher affective bond than male participants (*P*=.005, 95% CI of the difference 0.138-0.762) and 0.633 points higher perceived interpersonal closeness (*P*=.02, 95% CI of the difference 0.120-1.146). Differences in intention to use (*P*=.14) were not statistically significant. Pairwise comparisons between participant age groups (ie, younger than 40 years and 40 years or older) showed that across all chatbot personas, younger participants were more likely to intend to use the chatbot in the future than older participants (mean difference 0.404, SE 0.190; *P*=.04, 95% CI of the difference 0.029-0.778). Figure 4, A to C, and Figure 5, A to C, depict the interaction graphs for all outcome measurements by chatbot persona and participant gender or age, respectively.

Since the interaction effect between chatbot persona and participant age was significant, we inspected the interaction plots (Figure 5, A-C) and conducted two separate simple-effect analyses (1) for each chatbot persona and (2) for each participant age group, respectively.

The analysis of simple effects for each chatbot persona revealed the following significant differences. For the interpersonally closer dialogical-self chatbot, older participants reported significantly higher interpersonal closeness (mean difference 1.354, SE 0.527; *P*=.01) and higher bond scores (mean difference 1.186, SE 0.321; *P*<.001) than younger participants. For the interpersonally more distant expert chatbot, younger participants reported significantly higher intentions to use it than older participants (mean difference 0.805, SE 0.374; *P*=.03).

The analysis of simple effects for each participant age group confirmed that younger participants reported significantly lower affective bond scores (mean difference −1.015, SE 0.363; *P*=.03) for the interpersonally close dialogical-self chatbot compared to the distant institution chatbot. Other differences between the chatbot personas for each participant age group were not statistically significant. Taken together, hypothesis 2 is partially supported, and hypothesis 3 is fully supported.

Figure 4. Interaction effects of chatbot personas with participant genders on interpersonal closeness (A), affective bond (B), and intention to use (C).

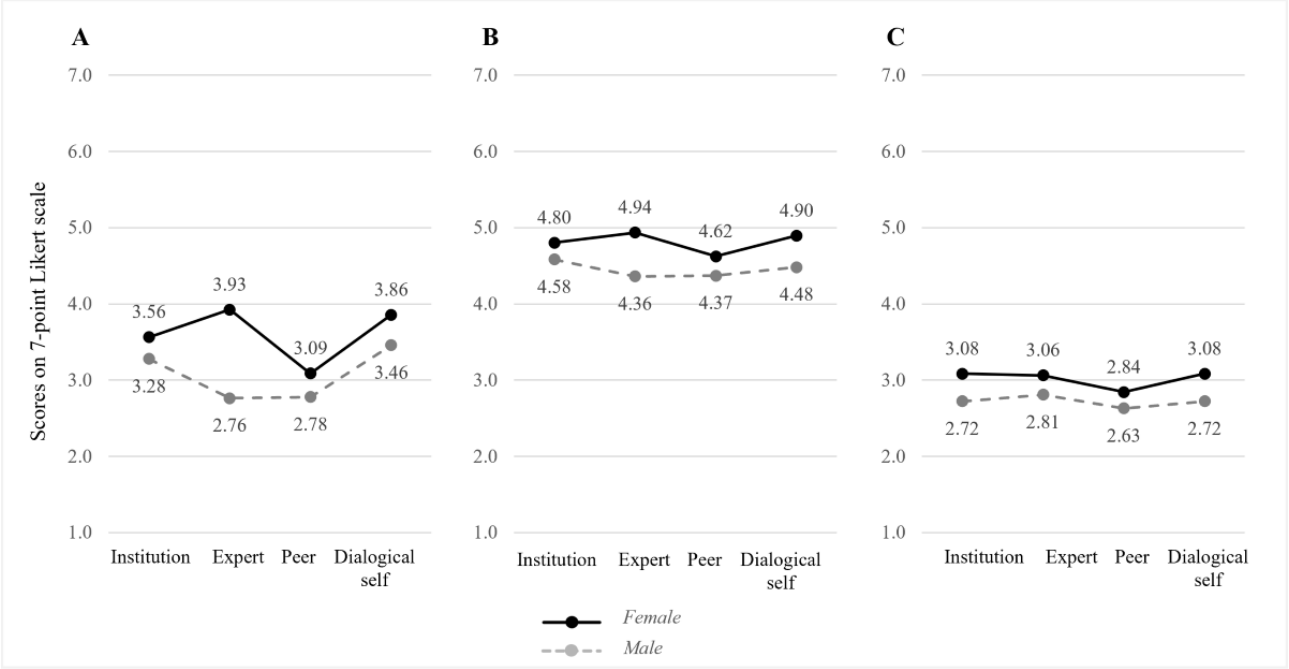
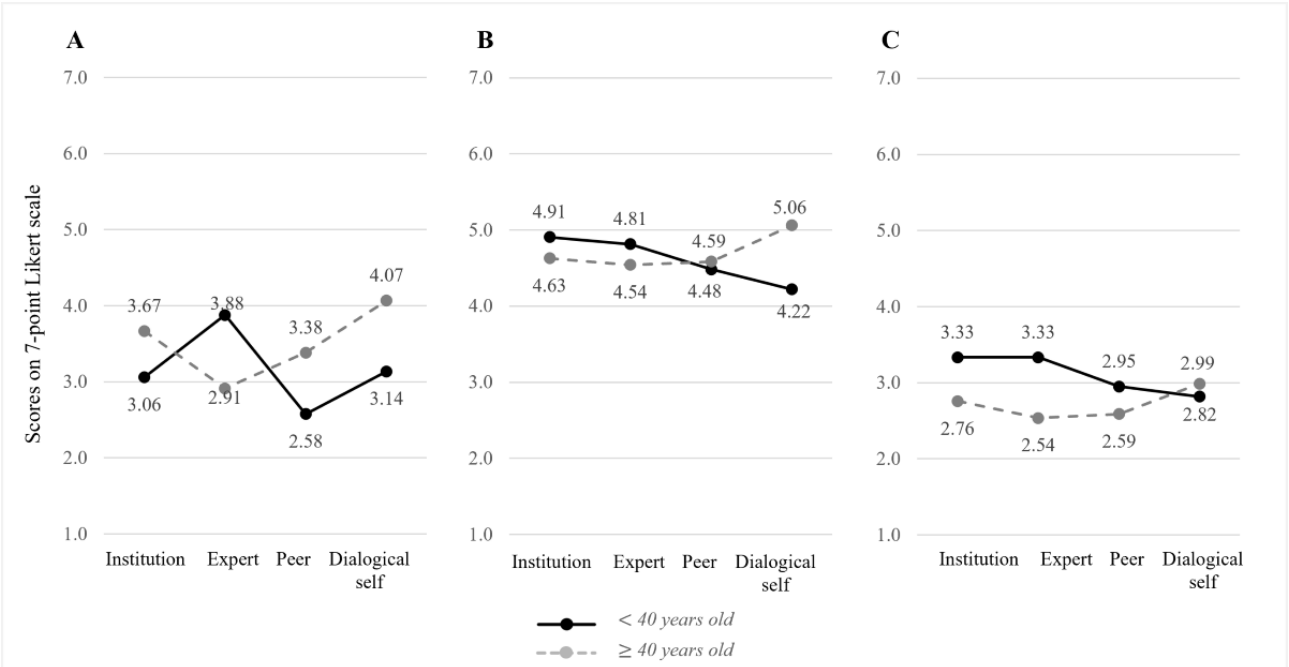


Figure 5. Interaction effects of chatbot personas with participant ages on interpersonal closeness (A), affective bond (B), and intention to use (C).



Free-Choice Versus No-Choice Group: Effects of Free Choice of the Chatbot Persona on Outcome Measures

In the free-choice group, a chi-square goodness-of-fit test showed that preferences for a chatbot persona were statistically significant (χ^2_3=31.4, *P*<.001), with most people (29/49, 59%) choosing the interpersonally close peer chatbot, compared to 8 participants (16%) each choosing the interpersonally more distant institution or expert chatbots. Only 4 participants (8%) chose the extremely close dialogical-self chatbot. Furthermore, cross-gender choice effects were significant in the expert, peer, and dialogical-self versions (χ^2_3=4.1, *P*=.04, Cramer *V*=0.318). A total of 84% (16/19) of all female participants chose a female avatar, whereas only 45% (10/22) of male participants chose a male avatar.

To examine the effect of free versus imposed choice, we specified another MANOVA model with choice type, participant gender, and participant age on the same outcome measures as before; however, we only compared perceptions of the peer chatbots, since group sizes of institution, expert, and dialogical-self chatbots in the free-choice group were arguably small.

The MANOVA revealed a significant main effect for choice type (Wilks λ=0.802, $F_{3,71}$=5.856, *P*=.001, η_p^2=0.198), but no significant main or interaction effects for or with participant

gender (*P*=.55) or age (*P*=.21), respectively. Table S5 in Multimedia Appendix 1 provides an overview of the MANOVA results. Subsequently conducted univariate ANOVAs specified with the same factors on each outcome variable confirmed the significant main effect of choice type on all outcome variables. For an overview, see Table S6 in Multimedia Appendix 1. Post hoc pairwise comparisons revealed that participants who had the option to freely choose the peer chatbot evaluated it consistently better than participants who had been imposed to interact with it; for example, affective bond with the peer chatbot had a mean score of 5.28 (SD 0.89) in the free-choice group compared to a mean score of 4.54 (SD 1.10) in the no-choice group. Similarly, participants who had freely chosen the chatbot felt closer to the chatbot than those who did not choose it (mean difference 1.613, SE 0.466; *P*=.001), and they were more likely to intend to use it (mean difference 1.294, SE 0.324; *P*<.001). Taken together, hypothesis 4 is supported.

Discussion

Principal Findings

Our design codebook for chatbots with different social roles provides a novel approach to design chatbots along an interpersonal closeness continuum that is inspired by clients' encounters with different social health care roles in their client journey: health care institutions, medical experts, peers, and themselves (ie, hypothesis 1). The results from the web-based experiment suggest that a chatbot's impersonated social role affects users' perception and the development of an affective bond contingent on users' demographic profiles. Since the main effect of the chatbot personas on the outcome measures was not significant, our study strengthens the recommendation that it is necessary to take into account user-specific factors before developing generic one-fits-all designs (ie, hypothesis 3).

Specifically, we found a significant age difference in chatbot assessments, this is, older participants rated the dialogical-self character in significantly more positive ways than younger participants, and younger participants consistently preferred the expert character. We assume that whereas younger participants valued guidance from a more distant "authority role" [27], older users seemed more intrigued by the idea to find guidance in themselves. Since older age is also associated with a greater number of life experiences and, consequently, more knowledge and judgment about life and ways of planning, managing, and understanding life [112], older participants might have had more trust in their expertise than younger participants. Thus, older participants' positive reception of the dialogical-self chatbot might generally point to an untapped potential of mimicking self-talk and inner dialogue by health care chatbots. Changing the way people think about and talk to themselves is a fundamental principle of cognitive behavioral therapy [113], an approach commonly implemented in many web-based [114] and chatbot-based mental health interventions [115]. Making a chatbot explicitly mimic one's inner voice represents an innovative approach to help clients experience how positive self-talk can look and feel.

Lastly, our study shows that giving individuals the option to choose between a range of presented chatbot characters can have an effect on their chatbot preferences (ie, hypothesis 4). Specifically, in our study, we found that free choice significantly improved participants' perceptions of the peer chatbot. This strengthens the recommendation that it is worth the extra effort to integrate even simple customization and personalization options [71,75]. Necessary user information, such as gender and age, could be easily elicited during onboarding based on responses to survey questions posed by the chatbot.

Limitations and Future Research Directions

This work has several limitations that point to future research directions upon which researchers can seize.

First, despite the scientific rigor employed to develop four distinct chatbot personas by manipulating a holistic set of interpersonal closeness cues derived from previous research, the conceptualized social roles can only be considered as design archetypes. Future experiments could examine nuances of the conceptualized social roles as well as which design cues are most relevant for clients' perceptions of a social role (research direction 1). Furthermore, our scenario covered a health care chatbot providing a particular type of health intervention. Whereas social role dynamics in a provider-client relationship in lifestyle interventions are likely similar in their essence to other health interventions (ie, highly sensitive, emotional, personal, and interpersonally intense), future research could explicitly compare differences in distinct health contexts, for example, between health care chatbots for different chronic diseases and those for preventive care (research direction 2).

Second, another limitation of our study concerns its limitation to German-speaking people ranging in age from 18 to 65 years only. Digital interventions addressing younger or older clients beyond this age range (eg, older people suffering from dementia [76] or younger clients with child obesity [116]) are likely to require different designs, content, and functionalities of the chatbot (research direction 3). For instance, the interpersonal relationship of an older client with cancer with a peer chatbot is likely to be of another kind than the relationship of a pregnant woman with a "pregnant" peer chatbot. Thus, depending on the health care setting, disease, or the specific stage of a disease [76], future research could explore further interesting specialized social roles that could be impersonated by a health care chatbot, such as professionals (ie, nurses, midwives, physiotherapists, pharmacists, etc) and health care social workers, which could help with alleviating the financial, social, and psychological burden of chronic disease. Other social roles that could be impersonated by a chatbot could be those of family members, for example, in digital pediatric or partner health interventions (research direction 4).

Third, another consideration concerns the length of the prospective client-chatbot relationship. In our experiment, participants were interacting with a chatbot prototype that simulated three nonconsecutive "days" of a fictitious health care intervention for only about 16 minutes in a time-lapsed manner. However, relationships develop and change in depth and breadth over time [95]. Therefore, the effects that we detected likely reflected participants' initial affect levels toward the chatbot, which would explain the smaller effect sizes as well. Future research could examine the impact of a chatbot's

social role in a fully operational prototype (research direction 5) to investigate changes in users' evaluations of the social role over time (research direction 6) or the choice of appropriate social roles at the various stages of the client journey (research direction 7). Future research could also examine the impact of a chatbot's social role, ultimately, on actual therapeutic outcomes (research direction 8), depending on the health care setting. For instance, similar to human-human interactions, some social roles might be more appropriate for communicating a life-threatening diagnosis, whereas another social role might be more appropriate for helping clients monitor a specific vital parameter every day. Eventually, our experimental design only allowed us to measure behavioral *intentions*, which do not necessarily translate into behavior (cf, intention-behavior gap [117]). Future field experiments should examine how the design of the chatbot actually affects, for instance, the likelihood to share sensitive personal health information (research direction 9).

Fourth, even if one result of our research was to match every client with their personalized perfect chatbot, or at least to provide enough freedom of choice between a couple of chatbot characters to improve clients' evaluation of a chatbot, the development of multiple chatbots delivering the same intervention adds complexity to the development of digital health interventions, thereby requiring more financial resources and time. Future research should seek to explore the optimal levels of personalization, customization, and choice options by contrasting them with development resources (research direction 10). In a similar vein, future researchers should also investigate additional individual characteristics as potential control variables, such as people's tendency to anthropomorphize artificial objects or their perceived creepiness when interacting with anthropomorphized technology (research direction 11) [69].

Finally, this study indicated that there were substantial age effects in the perception of different conversational agents within a cross-sectional design. In the case that these age effects can be replicated, future research investigating longitudinal designs is needed (research direction 12) to better understand whether such potential effects are impacted by particular cohorts and particular ages in the context of personality change (eg, Marsh et al [118]) and other digital health interventions.

Comparison With Prior Work

To the best of our knowledge, this is the first study to configure and compare health care chatbots with different impersonated social roles common in health care settings. By studying the impact of chatbots' social roles and integrating knowledge from the social role theory, this work extends prior research that has investigated the effects of relational cues [65], and it takes on prior calls for research to consider an agent's impersonated social role as an important factor in user-chatbot relationships [27,119]. This research thus contributes to the understanding of the design of health care chatbots by defining how a set of interpersonal closeness cues manifest the social role of a chatbot, as well as by determining how these social roles affect clients' experiences and the development of an affective bond with the chatbot.

Furthermore, our study contributes to prior research on the importance of personal characteristics in people's decisions to interact with chatbots [68] and shows that user perceptions of specific social roles depend on a person's demographic profile, namely an individuals' age and gender, and whether individuals could freely choose the chatbot persona or not.

Conclusions

Since chatbots are becoming increasingly prevalent in clients' health care service experiences, health care providers' success depends on their ability to design chatbots effectively. Especially in the context of chronic diseases, where digital health interventions are aimed at accompanying clients for years, it is still open as to which design choices promote the development of a strong affective bond between chatbot and client. To this end, we developed a codebook that allows researchers and practitioners to systematically design health care chatbots with specific social roles common in health care settings, and we explored whether or how these social roles affect the development of an affective bond between the user and the chatbot. Overall, our results suggest that positive effects can come from customizing the chatbot persona to easily accessible user characteristics, such as age and gender, or from allowing clients to choose the social role they feel they need most. Future work is required to investigate role dynamics in client-chatbot relationships during longer-term interventions.

Acknowledgments

The authors gratefully acknowledge all comments from reviewers of the study protocol that was presented at the 26th ECIS 2018 in Portsmouth, United Kingdom, June 23 to 28, 2018. Recruitment of the participants in this study was supported by a grant from the Swiss National Science Foundation (No. 162724; principal investigator: MA).

Authors' Contributions

MN and TK developed the study from conceptualization of the chatbots to study design. MN, DR, and TK were responsible for prototype design, app development, and implementation of the study. MS, CF, MA, and FvW tested and reviewed the study and prototype design. MN was responsible for data collection, curation, and analysis and wrote the manuscript, incorporating critical reviews from all authors. All authors reviewed and approved the manuscript before submission.

Conflicts of Interest

MN, DR, TK, and FvW are affiliated with the Centre for Digital Health Interventions (CDHI), a joint initiative of the Department of Management, Technology, and Economics at ETH Zurich and the Institute of Technology Management at the University of St. Gallen, which is funded in part by the Swiss health insurance company CSS Insurance. However, CSS did not have any role in the study design, data analysis, or data interpretation or in writing, reviewing, or approving the manuscript for publication. TK is the scientific director at the CDHI and FvW cochairs the CDHI. TK is also a cofounder of Pathmate Technologies AG, a university spin-off company that creates and delivers digital clinical pathways with the help of chatbots. DR started working for Pathmate Technologies in 2022. However, Pathmate Technologies had no role in the study described in this paper.

Multimedia Appendix 1
Supplementary material.
[PDF File (Adobe PDF File), 1047 KB - jmir_v24i4e32630_app1.pdf]

References

1. James J, Hibbard J, Agres T, Lott R, Dentzer S. Health policy brief: Patient engagement. Health Aff 2013 Feb 14;32:1-6 [FREE Full text] [doi: 10.1377/hpb2013.4]
2. Anderson CL, Agarwal R. The digitization of healthcare: Boundary risks, emotion, and consumer willingness to disclose personal health information. Inf Syst Res 2011 Sep;22(3):469-490. [doi: 10.1287/isre.1100.0335]
3. Hutton C, Gunn J. Do longer consultations improve the management of psychological problems in general practice? A systematic literature review. BMC Health Serv Res 2007 May 17;7:71 [FREE Full text] [doi: 10.1186/1472-6963-7-71] [Medline: 17506904]
4. Go E, Sundar SS. Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. Comput Human Behav 2019 Aug;97:304-316. [doi: 10.1016/j.chb.2019.01.020]
5. Tudor Car L, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng Y, et al. Conversational agents in health care: Scoping review and conceptual analysis. J Med Internet Res 2020 Aug 07;22(8):e17158 [FREE Full text] [doi: 10.2196/17158] [Medline: 32763886]
6. Schuetzler RM, Grimes GM, Scott Giboney J. The impact of chatbot conversational skill on engagement and perceived humanness. J Inf Manage Syst 2020 Nov 18;37(3):875-900. [doi: 10.1080/07421222.2020.1790204]
7. Fleisch E, Franz C, Herrmann A. Digital therapeutics. In: The Digital Pill: What Everyone Should Know About the Future of Our Healthcare System. Bingley, UK: Emerald Publishing Limited; 2021:81-98.
8. Bjaaland Skjuve M, Bae Brandtzæg P. Chatbots as a new user interface for providing health information to young people. In: Andersson Y, Dahlquist U, Ohlsson J, editors. Youth and News in a Digital Media Environment: Nordic-Baltic Perspectives. Göteborg, Sweden: Nordicom, University of Gothenburg; 2018:59-66.
9. Fadhil A, Gabrielli S. Addressing challenges in promoting healthy lifestyles: The AI-chatbot approach. In: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare. New York, NY: Association for Computing Machinery; 2017 Presented at: 11th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 23-26, 2017; Barcelona, Spain p. 261-265. [doi: 10.1145/3154862.3154914]
10. Riccardi G. Towards healthcare personal agents. In: Proceedings of the Workshop on Roadmapping the Future of Multimodal Interaction Research. 2014 Presented at: The Workshop on Roadmapping the Future of Multimodal Interaction Research; November 16, 2014; Istanbul, Turkey p. 53-56. [doi: 10.1145/2666253.2666266]
11. Horvath AO. Research on the alliance: Knowledge in search of a theory. Psychother Res 2018 Jul;28(4):499-516. [doi: 10.1080/10503307.2017.1373204] [Medline: 28899230]
12. Castonguay LG, Constantino MJ, Holtforth MG. The working alliance: Where are we and where should we go? Psychotherapy (Chic) 2006;43(3):271-279. [doi: 10.1037/0033-3204.43.3.271] [Medline: 22122096]
13. Sharf J, Primavera LH, Diener MJ. Dropout and therapeutic alliance: A meta-analysis of adult individual psychotherapy. Psychotherapy (Chic) 2010 Dec;47(4):637-645. [doi: 10.1037/a0021175] [Medline: 21198249]
14. Flückiger C, Del Re AC, Wampold BE, Horvath AO. The alliance in adult psychotherapy: A meta-analytic synthesis. Psychotherapy (Chic) 2018 Dec;55(4):316-340. [doi: 10.1037/pst0000172] [Medline: 29792475]
15. Flückiger C, Rubel J, Del Re AC, Horvath AO, Wampold BE, Crits-Christoph P, et al. The reciprocal relationship between alliance and early treatment symptoms: A two-stage individual participant data meta-analysis. J Consult Clin Psychol 2020 Sep;88(9):829-843. [doi: 10.1037/ccp0000594] [Medline: 32757587]
16. Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. Psychotherapy 1979;16(3):252-260. [doi: 10.1037/h0085885]
17. Constantino M, Castonguay L, Schut A. The working alliance: A flagship for the 'scientist-practitioner' model in psychotherapy. In: Tryon GS, editor. Counseling Based On Process Research: Applying What We Know. Boston, MA: Allyn and Bacon; 2002:81-131.
18. Aron A, Aron EN, Smollan D. Inclusion of Other in the Self Scale and the structure of interpersonal closeness. J Pers Soc Psychol 1992;63(4):596-612. [doi: 10.1037/0022-3514.63.4.596]

19. Berger T. The therapeutic alliance in internet interventions: A narrative review and suggestions for future research. Psychother Res 2017 Sep;27(5):511-524. [doi: 10.1080/10503307.2015.1119908] [Medline: 26732852]
20. Probst G, Berger T, Flückiger C. The alliance-outcome relation in internet-based interventions for psychological disorders: A correlational meta-analysis. Verhaltenstherapie 2019 Oct 10:1-12. [doi: 10.1159/000503432]
21. Ronis SD, McConnochie KM, Wang H, Wood NE. Urban telemedicine enables equity in access to acute illness care. Telemed J E Health 2017 Feb;23(2):105-112. [doi: 10.1089/tmj.2016.0098] [Medline: 27383822]
22. Darcy A, Daniels J, Salinger D, Wicks P, Robinson A. Evidence of human-level bonds established with a digital conversational agent: Cross-sectional, retrospective observational study. JMIR Form Res 2021 May 11;5(5):e27868 [FREE Full text] [doi: 10.2196/27868] [Medline: 33973854]
23. ter Stal S, Kramer LL, Tabak M, op den Akker H, Hermens H. Design features of embodied conversational agents in eHealth: A literature review. Int J Hum Comput Stud 2020 Jun;138:102409. [doi: 10.1016/j.ijhcs.2020.102409]
24. Rapp A, Curti L, Boldi A. The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. Int J Hum Comput Stud 2021 Jul;151:102630. [doi: 10.1016/j.ijhcs.2021.102630]
25. Nass C, Moon Y. Machines and mindlessness: Social responses to computers. J Soc Issues 2000 Jan;56(1):81-103. [doi: 10.1111/0022-4537.00153]
26. Ho A, Hancock J, Miner A. Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. J Commun 2018 Aug;68(4):712-733 [FREE Full text] [doi: 10.1093/joc/jqy026] [Medline: 30100620]
27. Fogg BJ. Computers as persuasive social actors. In: Persuasive Technology: Using Computers to Change What We Think and Do. San Francisco, CA: Morgan Kaufmann Publishers; 2003:89-120.
28. Nass C, Steuer J, Tauber E, Reeder H. Anthropomorphism, agency, and ethopoeia: Computers as social actors. In: Proceedings of the INTERACT '93 and CHI '93 Conference Companion on Human Factors in Computing Systems. 1993 Presented at: INTERACT '93 and CHI '93 Conference Companion on Human Factors in Computing Systems; April 24-29, 1993; Amsterdam, the Netherlands p. 110-112. [doi: 10.1145/259964.260137]
29. Chaves A, Egbert J, Hocking T, Doerry E, Gerosa M. Chatbots language design: The influence of language variation on user experience with tourist assistant chatbots. ACM Trans Comput Hum Interact 2022 Apr 30;29(2):1-38. [doi: 10.1145/3487193]
30. Biddle BJ. Recent developments in role theory. Annu Rev Sociol 1986 Aug;12(1):67-92. [doi: 10.1146/annurev.so.12.080186.000435]
31. Stewart M. Reflections on the doctor-patient relationship: From evidence and experience. Br J Gen Pract 2005 Oct;55(519):793-801 [FREE Full text] [Medline: 16212856]
32. Szasz TS, Hollender MH. A contribution to the philosophy of medicine: The basic models of the doctor-patient relationship. AMA Arch Intern Med 1956 May;97(5):585-592. [doi: 10.1001/archinte.1956.00250230079008] [Medline: 13312700]
33. Stewart M. Effective physician-patient communication and health outcomes: A review. CMAJ 1995 May 01;152(9):1423-1433 [FREE Full text] [Medline: 7728691]
34. Stiles WB, Putnam SM, Wolf MH, James SA. Interaction exchange structure and patient satisfaction with medical interviews. Med Care 1979 Jun;17(6):667-681. [doi: 10.1097/00005650-197906000-00010] [Medline: 449436]
35. Stiles WB, Putnam SM, James SA, Wolf MH. Dimensions of patient and physician roles in medical screening interviews. Soc Sci Med 1979 May;13A(3):335-341. [Medline: 462213]
36. Buller MK, Buller DB. Physicians' communication style and patient satisfaction. J Health Soc Behav 1987 Dec;28(4):375-388. [Medline: 3429807]
37. Spiegel JP. The social roles of doctor and patient in psychoanalysis and psychotherapy. Psychiatry 1954 Nov;17(4):369-376. [doi: 10.1080/00332747.1954.11022982] [Medline: 13215666]
38. Westerman D, Edwards AP, Edwards C, Luo Z, Spence PR. I-It, I-Thou, I-Robot: The perceived humanness of AI in human-machine communication. Commun Stud 2020 Apr 26;71(3):393-408. [doi: 10.1080/10510974.2020.1749683]
39. Gnewuch U, Morana S, Maedche A. Towards designing cooperative and social conversational agents for customer service. In: Proceedings of the 38th International Conference on Information Systems. 2017 Presented at: The 38th International Conference on Information Systems; December 10-13, 2017; Seoul, South Korea.
40. Janssen A, Rodríguez CD, Breitner M. More than FAQ! Chatbot taxonomy for business-to-business customer services. In: Proceedings of the 4th International Workshop on Chatbot Research and Design: CONVERSATIONS 2020. 2020 Presented at: The 4th International Workshop on Chatbot Research and Design: CONVERSATIONS 2020; November 23-24, 2020; Virtual p. 175-189. [doi: 10.1007/978-3-030-68288-0_12]
41. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: A systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
42. Følstad A, Skjuve M, Brandtzæg P. Different chatbots for different purposes: Towards a typology of chatbots to understand interaction design. In: Proceedings of the 5th International Conference on Internet Science. Cham, Switzerland: Springer International Publishing; 2019 Presented at: The 5th International Conference on Internet Science; October 24-26, 2018; St. Petersburg, Russia p. 145-156. [doi: 10.1007/978-3-030-17705-8_13]

43. Scarpellini I, Lim Y. Role-based design of conversational agents: Approach and tools. In: Proceedings of the International Conference on Human-Computer Interaction. Cham, Switzerland: Springer International Publishing; 2020 Presented at: The International Conference on Human-Computer Interaction; July 19-24, 2019; Copenhagen, Denmark p. 366-375. [doi: 10.1007/978-3-030-60700-5_47]
44. Bittner E, Oeste-Reiß S, Leimeister J. Where is the bot in our team? Toward a taxonomy of design option combinations for conversational agents in collaborative work. In: Proceedings of the 52nd Hawaii International Conference on System Sciences. 2019 Presented at: 52nd Hawaii International Conference on System Sciences; January 8-11, 2019; Maui, Hawaii p. 284-293. [doi: 10.24251/hicss.2019.035]
45. Seering J, Flores JP, Savage S, Hammer J. The social roles of bots. Proc ACM Hum Comput Interact 2018 Nov;2(CSCW):1-29. [doi: 10.1145/3274426]
46. Nißen M, Selimi D, Janssen A, Cardona DR, Breitner MH, Kowatsch T, et al. See you soon again, chatbot? A design taxonomy to characterize user-chatbot relationships with different time horizons. Comput Human Behav 2022 Feb;127:107043. [doi: 10.1016/j.chb.2021.107043]
47. McGoldrick PJ, Keeling KA, Beatty SF. A typology of roles for avatars in online retailing. J Mark Manage 2008 Apr 28;24(3-4):433-461. [doi: 10.1362/026725708x306176]
48. Rhee CE, Choi J. Effects of personalization and social role in voice shopping: An experimental study on product recommendation by a conversational voice agent. Comput Human Behav 2020 Aug;109:106359. [doi: 10.1016/j.chb.2020.106359]
49. Lee SK, Kavya P, Lasser SC. Social interactions and relationships with an intelligent virtual agent. Int J Hum Comput Stud 2021 Jun;150:102608. [doi: 10.1016/j.ijhcs.2021.102608]
50. Social role. APA Dictionary of Psychology. Washington, DC: American Psychological Association URL: https://dictionary.apa.org/social-role [accessed 2021-07-15]
51. Solomon MR, Surprenant C, Czepiel JA, Gutman EG. A role theory perspective on dyadic interactions: The service encounter. J Mark 2018 Nov 02;49(1):99-111. [doi: 10.1177/002224298504900110]
52. Oleś PK, Brinthaupt TM, Dier R, Polak D. Types of inner dialogues and functions of self-talk: Comparisons and implications. Front Psychol 2020;11:227. [doi: 10.3389/fpsyg.2020.00227] [Medline: 32210864]
53. Feine J, Gnewuch U, Morana S, Maedche A. A taxonomy of social cues for conversational agents. Int J Hum Comput Stud 2019 Dec;132:138-161. [doi: 10.1016/j.ijhcs.2019.07.009]
54. Forlizzi J, Zimmerman J, Mancuso V, Kwak S. How interface agents affect interaction between humans and computers. In: Proceedings of the Conference on Designing Pleasurable Products and Interfaces. New York, NY: ACM; 2007 Presented at: The Conference on Designing Pleasurable Products and Interfaces; August 22-25, 2007; Helsinki, Finland p. 209-221. [doi: 10.1145/1314161.1314180]
55. Parmar D, Olafsson S, Utami D, Bickmore T. Looking the part: The effect of attire and setting on perceptions of a virtual health counselor. In: Proceedings of the 18th International Conference on Intelligent Virtual Agents. New York, NY: ACM; 2018 Presented at: The 18th International Conference on Intelligent Virtual Agents; November 5-8, 2018; Sydney, Australia p. 301-306. [doi: 10.1145/3267851.3267915]
56. Olafsson S, Kimani E, Asadi R, Bickmore T. That's a Rap: Increasing engagement with rap music performance by virtual agents. In: Proceedings of the 17th International Conference on Intelligent Virtual Agents. 2017 Presented at: The 17th International Conference on Intelligent Virtual Agents; August 27-30, 2017; Stockholm, Sweden p. 325-334. [doi: 10.1007/978-3-319-67401-8_41]
57. Alsharbi B, Richards D. Using virtual reality technology to improve reality for young people with chronic health conditions. In: Proceedings of the 9th International Conference on Computer and Automation Engineering. New York, NY: ACM Press; 2017 Presented at: The 9th International Conference on Computer and Automation Engineering; February 18-21, 2017; Sydney, Australia p. 11-15. [doi: 10.1145/3057039.3057080]
58. Silverman BG, 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]
59. Creed C, Beale R, Cowan B. The impact of an embodied agent's emotional expressions over multiple interactions. Interact Comput 2014 Jan 29;27(2):172-188. [doi: 10.1093/iwc/iwt064]
60. Beattie A, Edwards AP, Edwards C. A bot and a smile: Interpersonal impressions of chatbots and humans using emoji in computer-mediated communication. Commun Stud 2020 Feb 16;71(3):409-427. [doi: 10.1080/10510974.2020.1725082]
61. Fadhil A, Schiavo G, Wang Y, Yilma B. The effect of emojis when interacting with conversational interface assisted health coaching system. In: Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare. New York, NY: ACM; 2018 Presented at: The 12th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 21-24, 2018; New York, NY p. 378-383. [doi: 10.1145/3240925.3240965]
62. Ho SM, Hancock JT. Context in a bottle: Language-action cues in spontaneous computer-mediated deception. Comput Human Behav 2019 Feb;91:33-41. [doi: 10.1016/j.chb.2018.09.008]
63. Kang S, Gratch J. People like virtual counselors that highly-disclose about themselves. In: Wiederhold BK, Riva G, Bouchard S, editors. Annual Review of Cybertherapy and Telemedicine 2011. Amsterdam, the Netherlands: IOS Press; 2011:143-148.

64. Liu B, Sundar SS. Should machines express sympathy and empathy? Experiments with a health advice chatbot. Cyberpsychol Behav Soc Netw 2018 Oct;21(10):625-636. [doi: 10.1089/cyber.2018.0110] [Medline: 30334655]
65. Bickmore TW, Picard RW. Establishing and maintaining long-term human-computer relationships. ACM Trans Comput Hum Interact 2005 Jun 01;12(2):293-327. [doi: 10.1145/1067860.1067867]
66. Ollier J, Nißen M, von Wangenheim F. The terms of "you(s)": How the term of address used by conversational agents influences user evaluations in French and German linguaculture. Front Public Health 2021 Jan 5;9:691595. [doi: 10.3389/fpubh.2021.691595] [Medline: 35071147]
67. Lee S, Lee N, Sah YJ. Perceiving a mind in a chatbot: Effect of mind perception and social cues on co-presence, closeness, and intention to use. Int J Hum Comput Interact 2019 Dec 06;36(10):930-940. [doi: 10.1080/10447318.2019.1699748]
68. Benbasat I, Dimoka A, Pavlou PA, Qiu L. The role of demographic similarity in people's decision to interact with online anthropomorphic recommendation agents: Evidence from a functional magnetic resonance imaging (fMRI) study. Int J Hum Comput Interact 2020 Jan;133:56-70. [doi: 10.1016/j.ijhcs.2019.09.001]
69. Seeger A, Pfeiffer J, Heinzl A. Texting with humanlike conversational agents: designing for anthropomorphism. J Assoc Inf Syst 2021;22(4):931-967. [doi: 10.17705/1jais.00685]
70. Calvaresi D, Calbimonte J, Siboni E, Eggenschwiler S, Manzo G, Hilfiker R, et al. EREBOTS: Privacy-compliant agent-based platform for multi-scenario personalized health-assistant chatbots. Electronics 2021 Mar 12;10(6):666. [doi: 10.3390/electronics10060666]
71. Dellaert BGC, Dabholkar PA. Increasing the attractiveness of mass customization: The role of complementary on-line services and range of options. Int J Electron Commer 2014 Dec 08;13(3):43-70. [doi: 10.2753/jec1086-4415130302]
72. Arora N, Dreze X, Ghose A, Hess JD, Iyengar R, Jing B, et al. Putting one-to-one marketing to work: Personalization, customization, and choice. Mark Lett 2008 Sep 16;19(3-4):305-321. [doi: 10.1007/s11002-008-9056-z]
73. Benlian A. Web personalization cues and their differential effects on user assessments of website value. J Manag Inf Syst 2015 Jul 06;32(1):225-260. [doi: 10.1080/07421222.2015.1029394]
74. Lisetti C, Amini R, Yasavur U, Rishe N. I can help you change! An empathic virtual agent delivers behavior change health interventions. ACM Trans Manag Inf Syst 2013 Dec 01;4(4):1-28. [doi: 10.1145/2544103]
75. Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, et al. The personalization of conversational agents in health care: Systematic review. J Med Internet Res 2019 Nov 07;21(11):e15360 [FREE Full text] [doi: 10.2196/15360] [Medline: 31697237]
76. Rampioni M, Stara V, Felici E, Rossi L, Paolini S. Embodied conversational agents for patients with dementia: Thematic literature analysis. JMIR Mhealth Uhealth 2021 Jul 16;9(7):e25381 [FREE Full text] [doi: 10.2196/25381] [Medline: 34269686]
77. Brown C, Feinberg F. How does choice affect evaluations? In: Proceedings of the Association for Consumer Research Conference: NA - Advances in Consumer Research. 2002 Presented at: The Association for Consumer Research Conference: NA - Advances in Consumer Research; October 2001; Austin, TX p. 331-332 URL: https://www.acrwebsite.org/volumes/8669
78. Coelho PS, Henseler J. Creating customer loyalty through service customization. Eur J Mark 2012 Mar 30;46(3/4):331-356. [doi: 10.1108/03090561211202503]
79. Kowatsch T, Nißen M, Rüegger D, Stieger M, Flückiger C, Allemand M, et al. The impact of interpersonal closeness cues in text-based healthcare chatbots on attachment bond and the desire to continue interacting: An experimental design. In: Proceedings of the 26th European Conference on Information Systems. 2018 Presented at: The 26th European Conference on Information Systems; June 23-28, 2018; Portsmouth, UK URL: https://aisel.aisnet.org/ecis2018_rip/46/
80. Obegi JH. The development of the client-therapist bond through the lens of attachment theory. Psychotherapy (Chic) 2008 Dec;45(4):431-446. [doi: 10.1037/a0014330] [Medline: 22122532]
81. Cooper A, Reimann R, Cronin D, Noessel C. About Face: The Essentials of Interaction Design. 4th edition. Indianapolis, IN: John Wiley & Sons, Inc; 2014.
82. Nowak KL, Rauh C. The influence of the avatar on online perceptions of anthropomorphism, androgyny, credibility, homophily, and attraction. J Comput Mediat Commun 2005 Nov;11(1):153-178. [doi: 10.1111/j.1083-6101.2006.tb00308.x]
83. Sah YJ, Peng W. Effects of visual and linguistic anthropomorphic cues on social perception, self-awareness, and information disclosure in a health website. Comput Human Behav 2015 Apr;45:392-401. [doi: 10.1016/j.chb.2014.12.055]
84. Laver J. Linguistic routines and politeness in greeting and parting. In: Rasmus Rask Studies in Pragmatic Linguistics, Volume 2: Conversational Routine. The Hague, Denmark: De Gruyter Mouton; 1981:298-304.
85. Levinson SC. Pragmatics and social deixis: Reclaiming the notion of conventional implicature. In: Proceedings of the 5th Annual Meeting of the Berkeley Linguistics Society. 1979 Presented at: 5th Annual Meeting of the Berkeley Linguistics Society; February 17-19, 1979; Berkely, CA p. 206-223. [doi: 10.3765/bls.v5i0.2162]
86. Moon Y. Intimate exchanges: Using computers to elicit self‐disclosure from consumers. J Consum Res 2000 Mar;26(4):323-339. [doi: 10.1086/209566]
87. Klein J, Moon Y, Picard R. This computer responds to user frustration: Theory, design, and results. Interact Comput 2002 Feb;14(2):119-140. [doi: 10.1016/S0953-5438(01)00053-4]

88. Morkes J, Kernal HK, Nass C. Effects of humor in task-oriented human-computer interaction and computer-mediated communication: A direct test of SRCT theory. Hum Comput Interact 2009 Nov 12;14(4):395-435. [doi: 10.1207/S15327051HCI1404_2]
89. Gilbertson J, Dindia K, Allen M. Relational continuity constructional units and the maintenance of relationships. J Soc Pers Relat 2016 Jun 29;15(6):774-790. [doi: 10.1177/0265407598156004]
90. Freeth M, Foulsham T, Kingstone A. What affects social attention? Social presence, eye contact and autistic traits. PLoS One 2013;8(1):e53286 [FREE Full text] [doi: 10.1371/journal.pone.0053286] [Medline: 23326407]
91. Utz S. Der aufbau von interpersonalen beziehungen in MUDs: Die rolle von motiven und kommunikationsstrategien. Gruppendynamik 2001 Jun 1;32(2):145-160. [doi: 10.1007/s11612-001-0015-x]
92. Bickmore T, Cassell J. Social dialogue with embodied conversational agents. In: Advances in Natural Multimodal Dialogue Systems. Dordrecht, the Netherlands: Springer; 2005:23-54.
93. Stafford L. Measuring relationship maintenance behaviors: Critique and development of the revised relationship maintenance behavior scale. J Soc Pers Relat 2010 Sep 14;28(2):278-303. [doi: 10.1177/0265407510378125]
94. Canary DJ, Stafford L, Hause KS, Wallace LA. An inductive analysis of relational maintenance strategies: Comparisons among lovers, relatives, friends, and others. Commun Res Rep 2009 Jun 06;10(1):3-14. [doi: 10.1080/08824099309359913]
95. Altman I, Taylor DA. Social Penetration: The Development of Interpersonal Relationships. New York, NY: Holt, Rinehart & Winston; 1973.
96. Collins NL, Miller LC. Self-disclosure and liking: A meta-analytic review. Psychol Bull 1994 Nov;116(3):457-475. [doi: 10.1037/0033-2909.116.3.457] [Medline: 7809308]
97. Hojat M, Gonnella JS, Mangione S, Nasca TJ, Magee M. Physician empathy in medical education and practice: Experience with the Jefferson scale of physician empathy. Semin Integr Med 2003 Mar;1(1):25-41. [doi: 10.1016/s1543-1150(03)00002-4]
98. Davis MH, Oathout HA. Maintenance of satisfaction in romantic relationships: Empathy and relational competence. J Pers Soc Psychol 1987;53(2):397-410. [doi: 10.1037/0022-3514.53.2.397]
99. Treger S, Sprecher S, Erber R. Laughing and liking: Exploring the interpersonal effects of humor use in initial social interactions. Eur J Soc Psychol 2013 Jul 02;43:532-543. [doi: 10.1002/ejsp.1962]
100. Albert S, Kessler S. Ending social encouters. J Exp Soc Psychol 1978 Nov;14(6):541-553. [doi: 10.1016/0022-1031(78)90048-3]
101. Eysenbach G. Improving the quality of web surveys: The Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
102. Stieger M, Nißen M, Rüegger D, Kowatsch T, Flückiger C, Allemand M. PEACH, a smartphone- and conversational agent-based coaching intervention for intentional personality change: Study protocol of a randomized, wait-list controlled trial. BMC Psychol 2018 Sep 04;6(1):43 [FREE Full text] [doi: 10.1186/s40359-018-0257-9] [Medline: 30180880]
103. Lin J, Reddy RM. Teaching, mentorship, and coaching in surgical education. Thorac Surg Clin 2019 Aug;29(3):311-320. [doi: 10.1016/j.thorsurg.2019.03.008] [Medline: 31235300]
104. Kowatsch T, Volland D, Shih I, Rüegger D, Künzler F, Barata F, et al. Design and evaluation of a mobile chat app for the open source behavioral health intervention platform MobileCoach. In: Proceedings of the 12th International Conference on Design Science Research in Information Systems and Technology. 2017 Presented at: The 12th International Conference on Design Science Research in Information Systems and Technology; May 30-June 1, 2017; Karlsruhe, Germany p. 485-489. [doi: 10.1007/978-3-319-59144-5_36]
105. Gächter S, Starmer C, Tufano F. Measuring the closeness of relationships: A comprehensive evaluation of the 'Inclusion of the Other in the Self' scale. PLoS One 2015 Jun 12;10(6):e0129478 [FREE Full text] [doi: 10.1371/journal.pone.0129478] [Medline: 26068873]
106. Kiluk BD, Serafini K, Frankforter T, Nich C, Carroll KM. Only connect: The working alliance in computer-based cognitive behavioral therapy. Behav Res Ther 2014 Dec;63:139-146 [FREE Full text] [doi: 10.1016/j.brat.2014.10.003] [Medline: 25461789]
107. Wixom BH, Todd PA. A theoretical integration of user satisfaction and technology acceptance. Inf Syst Res 2005 Mar;16(1):85-102. [doi: 10.1287/isre.1050.0042]
108. Field A. Discovering Statistics Using SPSS. 3rd edition. London, UK: SAGE Publications Ltd; 2009.
109. Field A, Miles J, Field Z. Discovering Statistics Using R. London, UK: SAGE Publications Ltd; 2012.
110. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd edition. New York, NY: Routledge; 1988.
111. Human subject research. Compliance Guide. English version. Zurich, Switzerland: ETH Zurich; 2019. URL: https://ethz.ch/content/dam/ethz/main/research/Ethische%20Fragen/Forschung%20an%20und%20mit%20Menschen/Compliance_Guide_RSETHZ133en.pdf [accessed 2022-01-06]
112. Baltes PB, Staudinger UM. Wisdom: A metaheuristic (pragmatic) to orchestrate mind and virtue toward excellence. Am Psychol 2000;55(1):122-136. [doi: 10.1037/0003-066x.55.1.122]
113. Thompson Prout H, Brown DT, editors. Counseling and Psychotherapy with Children and Adolescents: Theory and Practice for School and Clinical Settings. 4th edition. Hoboken, NJ: John Wiley & Sons, Inc; 2007.

114. McCall HC, Hadjistavropoulos HD, Sundström CRF. 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]
115. Bendig E, Erb B, Schulze-Thuesing L, Baumeister H. The next generation: Chatbots in clinical psychology and psychotherapy to foster mental health – A scoping review. Verhaltenstherapie 2019 Aug 20:1-13. [doi: 10.1159/000501812]
116. Stasinaki A, Büchter D, Shih CI, Heldt K, Güsewell S, Brogle B, et al. Effects of a novel mobile health intervention compared to a multi-component behaviour changing program on body mass index, physical capacities and stress parameters in adolescents with obesity: A randomized controlled trial. BMC Pediatr 2021 Jul 09;21(1):308 [FREE Full text] [doi: 10.1186/s12887-021-02781-2] [Medline: 34243738]
117. Sheeran P, Webb TL. The intention-behavior gap. Soc Personal Psychol Compass 2016 Sep 05;10(9):503-518. [doi: 10.1111/spc3.12265]
118. Marsh HW, Nagengast B, Morin AJS. Measurement invariance of big-five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and la dolce vita effects. Dev Psychol 2013 Jun;49(6):1194-1218. [doi: 10.1037/a0026913] [Medline: 22250996]
119. Ben Mimoun MS, Poncin I, Garnier M. Animated conversational agents and e-consumer productivity: The roles of agents and individual characteristics. Inf Manag 2017 Jul;54(5):545-559. [doi: 10.1016/j.im.2016.11.008]

Abbreviations

ANOVA: analysis of variance
CDHI: Centre for Digital Health Interventions
ECIS 2018: European Conference on Information Systems 2018
MANOVA: multivariate analysis of variance
T/V: *tu/vos*

Edited by R Kukafka; submitted 23.08.21; peer-reviewed by JM Leimeister, K Denecke, M Rampioni; comments to author 10.09.21; revised version received 21.01.22; accepted 17.02.22; published 27.04.22.

Please cite as:
Nißen M, Rüegger D, Stieger M, Flückiger C, Allemand M, v Wangenheim F, Kowatsch T
The Effects of Health Care Chatbot Personas With Different Social Roles on the Client-Chatbot Bond and Usage Intentions: Development of a Design Codebook and Web-Based Study
J Med Internet Res 2022;24(4):e32630
URL: https://www.jmir.org/2022/4/e32630
doi:10.2196/32630
PMID:35475761

Original Paper

Effectiveness of Web-Based Personalized Nutrition Advice for Adults Using the eNutri Web App: Evidence From the EatWellUK Randomized Controlled Trial

Rodrigo Zenun Franco[1], PhD; Rosalind Fallaize[2,3], PhD; Michelle Weech[2], PhD; Faustina Hwang[1], PhD; Julie A Lovegrove[2], PhD

[1]Biomedical Engineering, School of Biological Sciences, University of Reading, Reading, United Kingdom

[2]Hugh Sinclair Unit of Human Nutrition and Institute for Cardiovascular and Metabolic Research, University of Reading, Reading, United Kingdom

[3]School of Life and Medical Sciences, University of Hertfordshire, Hatfield, United Kingdom

Corresponding Author:
Julie A Lovegrove, PhD
Hugh Sinclair Unit of Human Nutrition and Institute for Cardiovascular and Metabolic Research
University of Reading
Whiteknights
Reading, RG6 6DZ
United Kingdom
Phone: 44 0118 378 6418
Email: j.a.lovegrove@reading.ac.uk

Abstract

Background: Evidence suggests that eating behaviors and adherence to dietary guidelines can be improved using nutrition-related apps. Apps delivering personalized nutrition (PN) advice to users can provide individual support at scale with relatively low cost.

Objective: This study aims to investigate the effectiveness of a mobile web app (eNutri) that delivers automated PN advice for improving diet quality, relative to general population food-based dietary guidelines.

Methods: Nondiseased UK adults (aged >18 years) were randomized to PN advice or control advice (population-based healthy eating guidelines) in a 12-week controlled, parallel, single-blinded dietary intervention, which was delivered on the web. Dietary intake was assessed using the eNutri Food Frequency Questionnaire (FFQ). An 11-item US modified Alternative Healthy Eating Index (m-AHEI), which aligned with UK dietary and nutritional recommendations, was used to derive the automated PN advice. The primary outcome was a change in diet quality (m-AHEI) at 12 weeks. Participant surveys evaluated the PN report (week 12) and longer-term impact of the PN advice (mean 5.9, SD 0.65 months, after completion of the study).

Results: Following the baseline FFQ, 210 participants completed at least 1 additional FFQ, and 23 outliers were excluded for unfeasible dietary intakes. The mean interval between FFQs was 10.8 weeks. A total of 96 participants were included in the PN group (mean age 43.5, SD 15.9 years; mean BMI 24.8, SD 4.4 kg/m^2) and 91 in the control group (mean age 42.8, SD 14.0 years; mean BMI 24.2, SD 4.4 kg/m^2). Compared with that in the control group, the overall m-AHEI score increased by 3.5 out of 100 (95% CI 1.19-5.78) in the PN group, which was equivalent to an increase of 6.1% (P=.003). Specifically, the m-AHEI components *nuts and legumes* and *red and processed meat* showed significant improvements in the PN group (P=.04). At follow-up, 64% (27/42) of PN participants agreed that, compared with baseline, they were still following some (*any*) of the advice received and 31% (13/42) were still motivated to improve their diet.

Conclusions: These findings suggest that the eNutri app is an effective web-based tool for the automated delivery of PN advice. Furthermore, eNutri was demonstrated to improve short-term diet quality and increase engagement in healthy eating behaviors in UK adults, as compared with population-based healthy eating guidelines. This work represents an important landmark in the field of automatically delivered web-based personalized dietary interventions.

Trial Registration: ClinicalTrials.gov NCT03250858; https://clinicaltrials.gov/ct2/show/NCT03250858

(J Med Internet Res 2022;24(4):e29088) doi:10.2196/29088

KEYWORDS

personalized nutrition; web-based; nutrition app; app; dietary intervention; eNutri; precision nutrition; mHealth; healthy eating index; diet quality scores; FFQ; food frequency questionnaire; EatWellUK

Introduction

Background

Noncommunicable diseases account for almost two-thirds of deaths globally [1]. The main recommendations for addressing noncommunicable diseases are related to lifestyle changes, such as the encouragement of healthier diets, physical activity (PA), and the reduction of tobacco use and alcohol consumption [1]. Current public health strategies that aim to address this challenge, for example, the United Kingdom's *5-a-day* campaign that encourages the consumption of at least 5 portions of a variety of fruits and vegetables daily [2] and the Eatwell Guide [3], are not personalized to individuals. Although the *5-a-day* campaign was associated with modest increases, particularly in fruit consumption, immediately after its launch [4], these were not maintained and currently only a third of UK adults meet the recommendations of this campaign [5]. These and other data have motivated investigations into the efficacy of personalized nutrition (PN) advice on dietary behavior change [6]. The promise of PN may lie in having a greater capacity to motivate individuals to change their dietary habits or the delivery of more suitable and thus more effective advice. For example, tailored health information is perceived as more personally relevant [7], and often superior [8], by consumers, and has also been shown to stimulate greater cognitive activity (eg, being read and remembered) [9]. Equally, there are known interindividual and intraindividual variations in response to diet [8], and PN attempts to account for these, depending on the level of personalization (eg, sex, metabolic requirements, physiological difference, genome, and microbiome).

The internet has considerable potential to improve health-related food choices at low cost, via apps, for example. However, a review showed that none of the popular nutrition-related mobile apps reviewed had a decision engine capable of providing personalized dietary advice to the user [10]. Evidence from the Food4Me study indicated that web-based PN advice based on dietary intake (assessed using a validated Food Frequency Questionnaire (FFQ) with portion-size photographs [11]) was more effective in improving adherence to dietary advice and diet quality than standard population guidance [12]. Their decision tree for tailoring the advice was executed manually by the researchers and automated after the completion of the randomized controlled trial (RCT) [13]; however, this automated decision tree is not currently publicly available. The authors of this study are not aware of any similar web-based PN RCT delivered automatically [14].

To address this need, our research team developed a mobile web app capable of delivering automated PN advice (eNutri v1.0 [15-17]), which, to our knowledge, is the only app to deliver PN advice automatically. Dietary advice was delivered immediately after completion of a web-based FFQ. The advice was based on and derived according to adherence to an 11-item modified Alternative Healthy Eating Index (m-AHEI). This measure of diet quality was a UK-adapted version of the US 2010 Alternative Healthy Eating Index (AHEI), which was selected for its strong association with cardiovascular disease (CVD) and health [18-20].

Objectives

The aim of this RCT (the EatWellUK study) was to evaluate the impact of the web-based PN advice tool, eNutri, on increasing diet quality in UK adults compared with generalized population dietary advice delivered on the web. This study tested the hypothesis that personalized dietary advice is more effective at eliciting beneficial dietary change than general dietary public health guidance.

Methods

Overview

The EatWellUK study was a randomized, controlled, parallel, single-blinded dietary intervention, which was delivered on the web and conducted by the Hugh Sinclair Unit of Human Nutrition (University of Reading, United Kingdom) between August 2017 and January 2018. It was designed to compare the impact of eNutri's automated personalized food-based dietary advice with generalized population dietary advice (control) delivered on the web on change in diet quality (assessed by the m-AHEI score and the scores of its individual components; see Multimedia Appendix 1 [3,18,21-28] for details).

Ethics Approval

The study was approved by the University of Reading (School of Chemistry, Food and Pharmacy) Research Ethics Committee (reference 13/17) and conformed with the Declaration of Helsinki. It was registered at ClinicalTrials.gov (NCT03250858).

Recruitment and Consent

Participants were recruited from the Hugh Sinclair Unit of Human Nutrition's volunteer database, University of Reading, mailing lists, social media (Facebook and Twitter), a university press release, web-based advertisements, and word of mouth. Interested parties received information with links to the consent form and participant information sheet hosted on the study website, where these documents were available on the home page for reading and downloading. The web-based account creation, using email and password, and the consent agreement were completed directly on the study website. It was made clear that participation was voluntary and that they were free to withdraw at any time without giving reason and without detriment. Participants were informed that they would need to complete web-based questionnaires at baseline, week 6, and week 12. There was no payment associated with participation, but to improve participant retention, all participants who completed the first set of questionnaires received an email regarding a prize draw (4 prizes of £50 [US $67.65] shopping vouchers were available) subject to the completion of the final

questionnaire. All contact with participants was via the website or email.

Screening and Randomization

Only participants aged ≥18 years were eligible to participate in the study. Screening was semiautomated in the eNutri web app, where a minimal set of exclusion criteria were applied automatically (not living in the United Kingdom, pregnant, lactating, receiving face-to-face nutrition services, lactose intolerance, food allergies, or diabetes). Other indications of potential exclusion were assessed by the researchers manually (self-report of health conditions, metabolic disorders, illness, medication, and specific dietary requirements), and in these cases, participants received an email to inform them of their eligibility.

As part of the screening form, participants were asked to report their age, sex, and highest level of education and how they heard about the study, selecting from the following options: email, Facebook, Instagram, Twitter, word of mouth, or other. Emails and social media links were created with customized URLs so that the app could also track the click source automatically [29,30]. Eligible participants were randomized automatically by the app using a random function [31] that generated a random number between 0 and 1. Depending on the value (lower or upper half of the interval), the participant was allocated to one of the two groups (PN or control). Allocation was concealed from the participants (single-blinded) who received advice at the same time points throughout the intervention.

Study Protocol

The EatWellUK study protocol is summarized in Textbox 1. Following automatic randomization, participants were asked to complete the web-based FFQ [15] and Baecke PA questionnaire [32] and to provide their self-reported weight using the eNutri web app. These measures were repeated at weeks 6 and 12 of the intervention. General (control group; Multimedia Appendix 2) or personalized (PN group) advice was displayed immediately after completion of the FFQ at baseline (week 0) and week 6. All participants received the personalized recommendations at week 12 (upon completion of the study). The eNutri FFQ and advice are described more fully in the study by Zenun Franco et al [15] and Fallaize et al [16]; see also the *Intervention Groups* section.

Textbox 1. EatWellUK study procedure.

Procedure

1. Web-based recruitment, providing the participant information sheet and consent form
2. Account creation via the study website
3. Web-based consent
4. Semiautomated screening (manual screening where analysis of text descriptions was required)
5. Participant's characteristics (sex, age, height, and level of education)
6. Group allocation (randomization)
7. Weight, physical activity questionnaire, and Food Frequency Questionnaire
8. System Usability Scale questionnaire
9. Presentation of web-based advice
10. Personalized web-based advice evaluation (optional)
11. Follow-up questionnaire (optional)

Although participants were encouraged to complete the FFQ in 1 session, it was important to offer the possibility to save the FFQ in case of interruption or temporary internet disconnection. Hence, each food selection was saved individually (after the portion-size selection), and participants could return to the last saved food item when they logged in to the system again. Incomplete FFQs expired after 24 hours, after which the participant was required to start again to maintain the validity and accuracy of the FFQs.

The interval between FFQ completions was also managed by the app. The second FFQ was made available only after 41 days (1 day before the participant reached 6 weeks), and the third (and final) FFQ only after 83 days (12 weeks) to prevent completion of the FFQs too early. Reminders were sent by email a few days before the FFQs were due. If the participant logged in to the system during the intervals, a message was shown indicating the date when their next FFQ would be available.

Using eNutri, steps 1 to 9 (Textbox 1) were completed at baseline (week 0; ~20 minutes in total) [15]. This first completion of step 7 (weight, PA, and FFQ) served as baseline data. Steps 7 and 9 were presented again by eNutri in weeks 6 and 12. Step 8 (System Usability Scale questionnaire) was presented at baseline only; detailed methods and EatWellUK study data for the System Usability Scale questionnaire have been described in the study by Zenun Franco et al [15]. The optional step 10, requesting completion of a web-based report evaluation, was presented only at the end of the study. After an interval of almost 5 months after the study ended, a further follow-up questionnaire (step 11) was conducted using a web-based survey tool.

Outcome Measures

Dietary Intake

Changes from baseline in dietary intake at end point were assessed via a graphical semiquantitative FFQ on the eNutri web app, which was based on a previously validated FFQ for a UK population [11]. The eNutri FFQ has been described previously [15]. The 2010 AHEI [18] was used as the foundation for (1) measuring the quality of the diet, (2) deriving the PN advice, and (3) quantifying changes in dietary intake. Some modifications were applied to the 2010 AHEI to adapt it to the UK dietary guidelines and to improve its use as the decision engine for the PN recommendations. The 11 food and drink components and scoring criteria for the m-AHEI can be found in Multimedia Appendix 1 [3,18,21-28]. The maximum component score was changed from 10 to 100, to facilitate data visualization and progress monitoring for the participant (details of the PN report are presented in the study by Fallaize et al [16]). Dietary intakes between the minimum (0 point) and maximum (100 points) criteria were scored using linear interpolation, with a positive slope for *healthy* components and a negative slope for *unhealthy* components, such that higher scores represented greater diet quality for every component. All the 11 individual components were weighted equally, and the overall score was presented to the participants as a percentage (ranging from 0 to 100) for ease of interpretation.

Weight, BMI, and PA Levels

Secondary outcome measures recorded via the eNutri web app included weight, BMI, and PA. Changes from baseline were measured for self-reported weight (kg) at end point. Change in weight was combined with height (collected in step 5) and reported as change in BMI (kg/m^2). For PA levels, change was measured from baseline in self-reported PA (Baecke questionnaire) at end point.

As participants could be advised to either gain or lose weight to reach their ideal BMI range, an analysis of the change in BMI without considering the direction of the change (ie, increase or decrease) would not capture the effectiveness of the recommendation (ie, opposite variations across participants would cancel one another). Thus, the absolute difference from the current BMI to the ideal BMI was analyzed to determine if the personalized advice decreased this difference significantly, in comparison with the control group.

Evaluation of Personalized Advice

Immediately after completion of the study at end point, the PN report was evaluated by the participants via 9 optional questions, also delivered via the eNutri web app, regarding the users' perceived system effectiveness [33] and perceptions of its design. The first 6 questions were Likert questions, and the final 3 questions offered the possibility to write comments. As both groups received the PN report after completing the study, their responses were combined for this evaluation.

Follow-up Questionnaire

To assess the long-term impact of the PN advice, a web-based follow-up questionnaire was administered via Online Surveys (Jisc Online Surveys) 4.6 months after the study ended, which invited all participants to provide feedback about eNutri and the advice they received. Those who responded and consented to participating were asked 32 questions, including Likert and multiple-choice questions. The primary purpose of these questions was to identify to what degree the PN advice had encouraged those in the PN group to improve their diets during the study and whether they were still following any aspects of their advice. Where they did not follow the advice, participants were also asked to identify their reasons for not doing so. In addition, the follow-up questionnaire included free-text boxes for the participants to write a short review of eNutri and comment on the advice they received. Although data for the control group were also obtained, these data are not presented here, as this group also received a PN report at the end of the study and, as such, their responses will likely be confounded.

Intervention Groups

PN Intervention

The PN report received by this group via the eNutri web app [16] consisted of (1) the participant's overall m-AHEI score (out of 100), (2) 3 diet messages, (3) feedback on BMI (including their ideal weight range), and (4) feedback on PA. The diet messages were tailored for each participant based on the 3 lowest m-AHEI component scores assessed with data from the FFQ, following a protocol published previously [15,16]. The components were presented as food-based recommendations; for example, if one of the lowest m-AHEI scores was *red or processed meat*, then the advice would use the FFQ data to highlight which individual foods in their diet were the highest contributors. Participants in the PN group were able to see a progress report after each subsequent FFQ (weeks 6 and 12). In the software version deployed in this study (eNutri v1.0), the inputs to the decision engine generating the PN feedback were a participant's dietary data and sex.

The ideal weight range of the participants was based on their BMI calculation. A healthy BMI ranges from 18.5 to 25.0 kg/m^2; hence, an ideal weight for a participant was presented as the midpoint at 21.75 kg/m^2. Textual messages and visual representations in the app were also tailored according to BMI (eg, colored bars on the scale to represent the ideal weight range) [15,16].

The PA feedback was based on the Baecke questionnaire [32]. Participants were presented with their overall PA scores, followed by scores for the three categories (sports, leisure, and work), as defined by Baecke et al [32]. Scores were on a scale of 0 to 100, with higher scores representing greater levels of PA. Advice messages related to the sports and leisure categories were provided according to the participant's score in each category. As it was deemed unlikely for participants to have much control over the nature of their activities at work, no personalized message regarding the work category was provided [15].

Control

The control group received generic healthy eating advice at baseline and week 6 via the eNutri web app (Multimedia Appendix 2). The report included 3 generalized healthy eating

messages that were based on the m-AHEI components (baseline: *vegetables*, *free sugars*, and *polyunsaturated fatty acids*; week 6: *fruit*, *wholegrain products*, and *red or processed meat*). General advice was also provided on the importance of maintaining a healthy weight and attaining adequate PA (Multimedia Appendix 2); however, this was not tailored according to the participants' BMI or reported PA levels. Following the final FFQ at week 12, the control group were provided with PN (intervention) advice. The UK Government's Healthy Eating Recommendations were used as a basis for the component messages [34].

Data Handling

As not every participant completed the eNutri questionnaires on the target dates of 6 and 12 weeks (some took longer to respond), only 2 questionnaires per participant were considered in the outcome analysis: baseline and the date closest to the target date of 12 weeks, referred to as end point. The effectiveness of the decision engine was captured in terms of users' actual dietary change between baseline and end point, using the m-AHEI as the primary outcome measure.

Participants were excluded from the analysis if (1) their ratio of energy intake to basal metabolic rate (estimated using the equation of Henry [35]) exceeded standard cutoffs (men: <0.49, >2.79; women: <0.56, >3.21; n=14) [36]; (2) there was a large difference in energy intake between FFQs (>8000 kJ) without corresponding weight change (n=3), or (3) reported intakes of food groups were considered unfeasible in relation to maximum adult intakes reported in the National Diet and Nutrition Survey [37], such as 10 eggs or 1.2 kg of porridge daily (n=6).

Statistical Analysis

Statistical analysis was conducted using Python StatsModel (version 0.11) [38]. Normality was tested using the D'Agostino-Pearson and Shapiro-Wilk tests [39], and, where necessary, m-AHEI component data were square root transformed. Treatment effects were determined based on the change from baseline between groups, where $P \leq .05$ was considered significant. To account for mean lower-energy intake reported at the end of the study compared with baseline (PN: –1316, SD 2315 kJ/day; control: –726, SD 2549 kJ/day) in the absence of weight change, baseline energy intakes were included as a covariate for the m-AHEI analysis only. Secondary outcomes (PA and BMI) were adjusted for baseline values and presented as adjusted means [40]. Unless specified, data are presented as means (SDs).

Participants in the PN group received advice on specific m-AHEI components (based on their 3 lowest m-AHEI components). For analysis of change in individual m-AHEI components (eg, *fruit*), a smaller treatment effect is expected if the participant in the PN group did not receive advice for changing that specific component. Furthermore, the subgroup of participants in the control group with low scores for a specific component have greater room to improve their score (ie, greater distance to the maximum score) for that component than the group as a whole. To account for these factors, in addition to the treatment effects for the whole group, a treatment effect was also calculated for the subgroup of participants in the PN group who received personalized messages for a specific component, in comparison with participants in the control group who were matched in the sense that, based on their 3 lowest m-AHEI scores, they would also have received advice on that component had they been in the intervention group.

This RCT was powered based on the outcomes of the similar Food4Me study [12], comparing participants who received control advice with PN advice based on dietary intake only, expecting an increase of 6.5% (mean 49.58%, SD 9.51%; α=.05; power=0.8) in the m-AHEI (Food4Me consortium, unpublished data, October 2014). With these variables, the recruitment target was 274 participants, increasing to 330 participants when factoring a 20% dropout rate.

Results

Participants

A total of 438 participants created accounts in the eNutri web app. Table 1 presents which recruitment sources were reported by the participants and the results of the URL automatic tracking. Although sources were identified (self-reported) by 91.6% (401/438) of participants, the automatic URL tracking identified sources for just 61.6% (270/438) participants. The most frequently self-reported recruitment sources were email (164/438, 37.4%), Facebook (59/438, 13.5%), and Twitter (43/438, 9.8%).

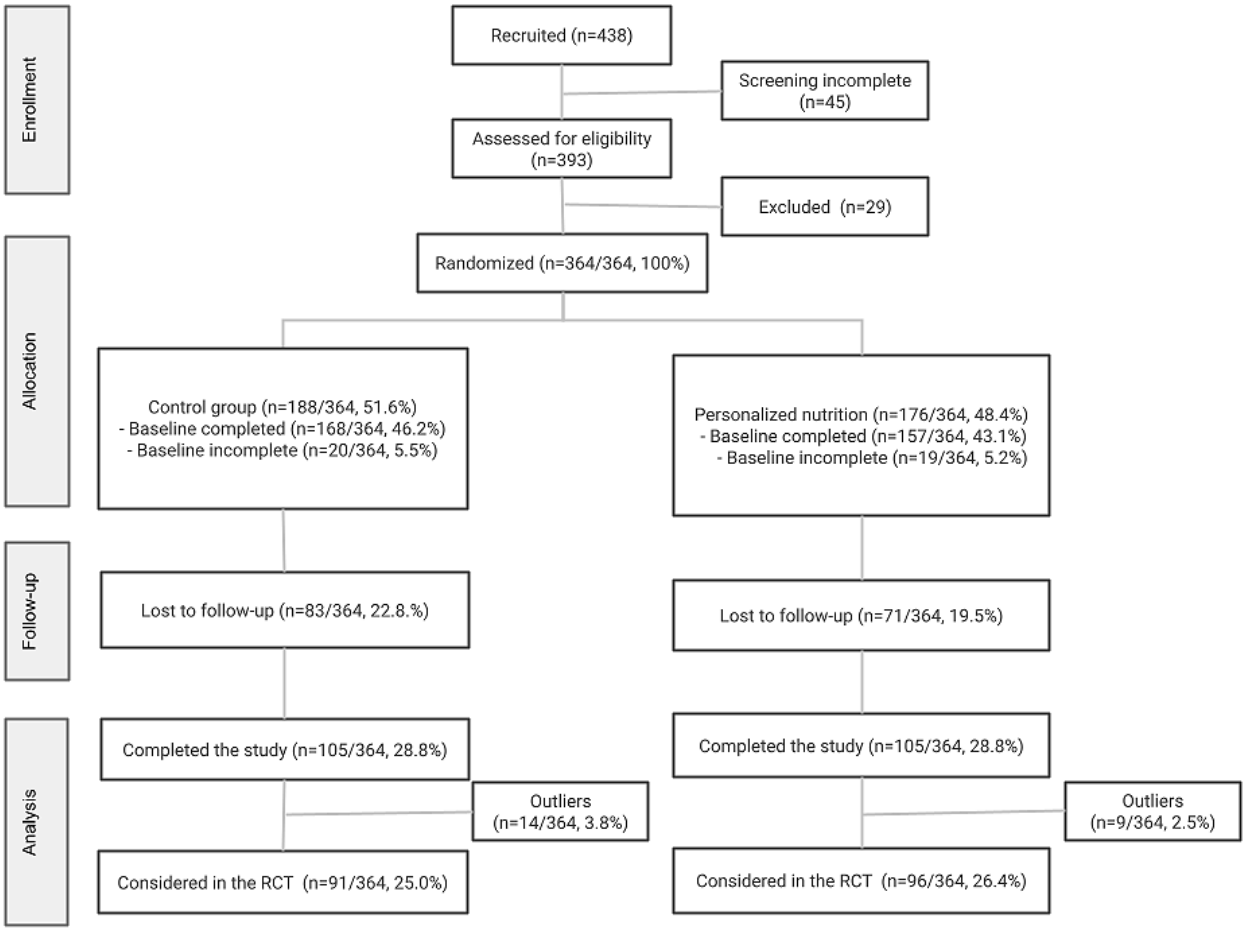
Table 1. Recruitment sources as self-reported by all participants creating an account (N=438) and from automatic detection by the app.

Recruitment source	Self-report, n (%)	Automatic URL track, n (%)
Email	164 (37.4)	199 (45.4)
Facebook	59 (13.5)	26 (5.9)
Twitter	43 (9.8)	11 (2.5)
Instagram	0 (0)	0 (0)
Word of mouth	63 (14.4)	0 (0)
Other	72 (16.4)	34 (7.8)
Not available	37 (8.4)	168 (38.4)

Of the 438 accounts, 393 (89.7%) completed the screening questionnaire. Of these 393, 29 (6.6%) participants were excluded owing to country of residence (n=6), medication use (n=8), or dietary restrictions, such as lactose intolerance (n=10)

or food allergies (n=7). The remaining 83.1% (364/438) of the participants were automatically randomized by the app to either the control or PN group, although 10.7% (39/364) of these participants did not complete the baseline questionnaires (Figure 1).

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram for the EatWellUK study. n values are expressed as percentages of the number of participants who were randomized (N=364). RCT: randomized controlled trial.



Of the 325 participants who completed the baseline FFQ, 115 (35.4%) were lost to follow-up (ie, they did not complete the next FFQ), and 210 (64.6%) completed at least 1 additional FFQ, and these were considered in the RCT. Of these 210, 23 (11%) were removed as outliers for unfeasible dietary intakes. A total of 114 participants from the control (n=54) and PN (n=60) groups completed all 3 FFQs. At the end of the study, the participants were presented with an optional personalized web-based report evaluation questionnaire to provide feedback on the PN report. Of the 111 feedback forms received, 50 (45%) were from the control group and 61 (55%) from the PN group. These feedback responses were combined because all participants were given personalized reports at the end of the study, and no significant differences were found between the groups (data not shown). The baseline (week 0) characteristics of the participants included in the analysis after removal of outliers are presented in Table 2; no significant differences in sex, age, or educational attainment were observed between the intervention groups.

Table 2. Baseline characteristics of the EatWellUK study participants (N=187).

Characteristics	Total sample	Control group	Personalized nutrition group	*P* value
Participants, n (%)	187 (100)	91 (48.7)	96 (51.3)	
Sex, n (%)				.07
Female	157 (84)	81 (43.3)	76 (40.6)	
Male	30 (16)	10 (5.3)	20 (10.6)	
Age (years), mean (SD; range)	43.2 (15.0; 18-85)	42.8 (14.0; 20-82)	43.5 (15.9; 18-85)	.76
Highest level of education, n (%)				.19
Less than secondary	0 (0)	0 (0)	0 (0)	
Secondary	20 (10.6)	13 (6.9)	7 (3.7)	
College	21 (11.2)	11 (5.8)	10 (5.3)	
Undergraduate	64 (34.2)	25 (13.4)	39 (20.9)	
Postgraduate	82 (43.9)	42 (22.5)	40 (21.4)	

Primary Outcomes Evaluation

Considering the protocol for selecting the end point FFQ (ie, the one closest to week 12), the trial resulted in an average interval between FFQs of 10.8 weeks. The analysis by group confirmed that the intervals were equivalent across the control (10.7 weeks) and PN (10.8 weeks) groups. The results for the overall changes in m-AHEI scores from baseline to end point (10.8 weeks) are presented in Table 3.

Table 3. Effects of the EatWellUK intervention on the m-AHEI[a] component scores, considering all the participants in the control (n=91) and PN[b] (n=96) groups.[c]

m-AHEI variables	Baseline, mean (SD)		Adjusted Δ,[d] mean (SD)		Treatment effect, Δ PN–Δcontrol (95% CI)[d]	*P* value
	Control	PN	Control	PN		
Overall m-AHEI score	58.9 (12.3)	56.3 (11.5)	–0.4 (2.3)	3.1 (2.1)	3.5 (1.19 to 5.78)	.003
Positive components						
Vegetable score	68.0 (26.0)	61.2 (28.3)	–4.0 (8.4)	–7.1 (9.4)	–3.2 (–9.31 to 3.01)	.32
Fruit score	67.0 (31.0)	60.6 (34.8)	–5.7 (6.8)	–3.7 (7.5)	2.0 (–4.32 to 8.22)	.54
Whole grain score[e]	43.5 (35.4)	35.5 (33.9)	–0.6 (1.2)	–0.3 (1.1)	0.3 (–0.45 to 0.95)	.48
Dairy product score	87.7 (27.8)	95.0 (18.1)	–0.9 (13.6)	0.4 (8.9)	1.3 (–4.96 to 7.52)	.69
Nuts and legume score[e]	47.3 (38.3)	26.7 (33.1)	0.3 (1.4)	1.2 (1.2)	0.9 (0.03 to 1.76)	.04
Healthy fats score	53.4 (17.7)	50.8 (16.7)	0.7 (7.5)	–1.2 (7.1)	–1.9 (–6.36 to 2.55)	.40
Oily fish score	63.0 (41.7)	69.2 (38.0)	–2.1 (17.3)	3.7 (16.0)	5.8 (–3.72 to 15.3)	.23
Negative components						
Free sugars score	44.3 (27.7)	50.5 (26.8)	–2.3 (8.9)	3.9 (8.5)	6.1 (–0.33 to 12.6)	.06
Red and processed meat score[e]	27.5 (36.6)	24.4 (36.1)	0.4 (0.9)	1.2 (0.9)	0.8 (0.05 to 1.58)	.04
Salt score	55.9 (34.4)	57.0 (30.7)	7.8 (17.1)	14.1 (15.3)	6.3 (–0.90 to 13.5)	.09
Alcohol score	90.2 (27.2)	88.7 (27.8)	2.8 (15.6)	3.5 (15.4)	0.6 (–5.15 to 6.41)	.83

[a]m-AHEI: modified Alternative Healthy Eating Index.

[b]PN: personalized nutrition.

[c]Scores are reported on a scale between 0 and 100, where higher scores reflect greater diet quality.

[d]Change from baseline at end point. Data are presented as adjusted means with the baseline energy intakes as a covariate [40].

[e]Square root transformation.

Compared with that for the control group, the treatment effect observed in the overall m-AHEI score for the PN group was 3.5 out of 100 (95% CI 1.19-5.78), which was reached statistical significance (*P*=.003). A statistically significant improvement in *nuts and legumes*, and *red and processed meat* scores were also observed in the PN group compared with the control group

during the intervention period (*P*=.04), reflecting an increased intake of nuts and legumes and reduced intake of red and processed meat.

All the participants in the PN group (n=96) received feedback regarding their overall m-AHEI score and were able to see the progress report with all the individual m-AHEI scores; however, the focus of the advice was on just 3 components [15].

The outcomes presented in Table 3 do not consider whether a specific participant received a personalized message for that component but rather how the individual m-AHEI components were affected by the intervention as a whole. Table 4 shows the component messages presented to the PN group during the intervention (ie, the 3 m-AHEI components with the lowest scores), and the *matched* participants from the control group who would have received those messages had they been in the PN group. The distribution presented in the final column of Table 4 (total messages) gives an indication of the components for which the EatWellUK cohort were most in need of improvement. *Red and processed meat*, *nuts and legumes*, and *whole grains* were the components most frequently presented by the eNutri app algorithm across both groups for having the lowest m-AHEI scores, whereas *dairy products*, *alcohol*, and *vegetables* were presented in <10% of cases for both groups, suggesting these components required least improvement in the participant's diets.

Table 4. Frequency of healthy eating messages presented to the PN[a] group (n=96) and matched messages in the control group (n=91) at baseline when only messages for the three components with the lowest scores were considered.[b]

	Matched control messages, n (%)	PN messages, n (%)	Total messages, n (%)
Red and processed meat	60 (32.1)	66 (35.3)	126 (22.5)
Nuts and legumes	42 (22.5)	70 (37.4)	112 (19.9)
Whole grains	39 (20.9)	48 (25.7)	87 (15.5)
Salt	27 (14.4)	23 (12.3)	50 (8.9)
Free sugars	30 (16)	17 (9.1)	47 (8.4)
Oily fish	26 (13.9)	18 (9.6)	44 (7.8)
Fruits	14 (7.5)	15 (8)	29 (5.2)
Healthy fats	13 (7)	10 (5.3)	23 (4.1)
Vegetables	7 (3.7)	9 (4.8)	16 (2.9)
Alcohol	7 (3.7)	8 (4.3)	15 (2.7)
Dairy products	8 (4.3)	4 (2.1)	12 (2.1)
Total messages	273 (48.7)	288 (51.3)	561 (100)

[a]PN: personalized nutrition.

[b]Components are ordered by the total number of healthy eating messages that were (personalized nutrition group) or would have been (control group) presented to participants. The personalized nutrition and control group data are presented as a contribution to the total sample of messages produced by eNutri. Because each participant (n=187) received or would have received 3 messages from eNutri, the total number of messages is 561.

The treatment effect on participants in the PN group who received personalized messages for a specific component was also calculated in comparison with the matched participants in the control group, as shown in Table 5. Although participants in the PN group displayed greater score improvements across all m-AHEI components, except *vegetables*, significant treatment effects were only observed for *salt* (+18.3; *P*=.04) and *alcohol* (+51.4; *P*=.03) in the PN group compared with that in the control group, reflecting a significantly greater reduction in intake of these components following PN intervention.

Table 5. Changes in the m-AHEI[a] component scores from baseline to end point for participants in the PN[b] group (n=96) who received these specific component messages and the matched participants in the control group (n=91).

m-AHEI component	Matched control			PN group			Treatment effect, Δ PN–Δcontrol (95% CI)[c]	*P* value
	Value, n (%)[d]	Baseline, mean (SD)	Δ (SD)[c]	Value, n (%)[d]	Baseline, mean (SD)	Δ (SD)[c]		
Positively scored components								
Nuts and legumes[e]	42 (46.2)	19.1 (17.7)	1.5 (0.8)	70 (72.9)	13.9 (19.2)	2.3 (0.8)	0.8 (–0.41 to 1.98)	.20
Whole grains[e]	39 (42.9)	16.8 (16.3)	0.2 (0.3)	48 (50)	11.1 (13.9)	1.0 (0.3)	0.8 (–0.27 to 1.78)	.15
Oily fish	26 (28.6)	7.4 (14.6)	23.8 (7.7)	18 (18.8)	10.3 (15.6)	31.7 (8.3)	7.9 (–15.2 to 31.0)	.49
Fruits	14 (15.4)	26.1 (18.8)	1.1 (2.5)	15 (15.6)	18.3 (12.2)	5.0 (3.1)	3.9 (–10.0 to 17.9)	.57
Healthy fats	13 (14.3)	46.8 (15.7)	8.0 (8.1)	10 (10.4)	38.1 (14.9)	8.2 (8.7)	0.2 (–13.3 to 13.6)	.98
Vegetables	7 (7.7)	40.0 (7.5)	10.5 (4.0)	9 (9.4)	26.0 (15.2)	3.9 (7.1)	–6.7 (–36.1 to 22.8)	.63
Dairy products	8 (8.8)	20.1 (13.8)	8.2 (21.0)	4 (4.2)	22.8 (14.1)	25.7 (17.5)	17.5 (–47.0 to 82.0)	.55
Negatively scored components								
Red and processed meat[e]	60 (65.9)	4.3 (8.9)	0.9 (0.6)	66 (68.8)	3.3 (8.3)	1.9 (0.6)	1.0 (–0.06 to 2.00)	.06
Salt	27 (29.7)	14.1 (18.3)	31.5 (16.6)	23 (24)	25.7 (24.5)	49.8 (22.8)	18.3 (1.18 to 35.5)	.04
Free sugars	30 (33)	21.6 (24.8)	7.0 (9.2)	17 (17.7)	9.9 (14.5)	19.3 (5.4)	12.3 (–3.40 to 28.0)	.12
Alcohol	7 (7.7)	8.4 (17.6)	–1.5 (36.8)	8 (8.3)	11.4 (20.7)	49.9 (38.4)	51.4 (4.93 to 97.8)	.03

[a]m-AHEI: modified Alternative Healthy Eating Index.

[b]PN: personalized nutrition.

[c]Change from baseline at end point. Data are presented as adjusted means with baseline energy intakes as a covariate.

[d]Values represent percentage of intervention group (control and intervention) who received component messages.

[e]Square root transformation.

Secondary Outcomes Evaluation

As both the PN and control participants received advice on weight and PA (albeit in different formats), analysis of matched participants was not required. Absolute BMI was not affected by the treatment (Table 6). The mean distances to the ideal BMI decreased slightly (ie, BMI improved) in the PN group (–0.09 kg/m^2) with no change in the control group, but this improvement was not statistically significant from the control group (*P*=.37). A number of participants in the control (n=13) and PN (n=21) groups reported the same weight at the end point and baseline. Although a significant improvement in Work score was reported in the PN group (*P*=.02) compared with the control group, there were no significant differences in change in overall Baecke score between the groups (*P*=.70).

Table 6. Changes in BMI and PA[a] level (Baecke) score from baseline to end point for participants in the control (n=91) and PN[b] (n=96) groups.[c]

	Baseline, mean (SD)		Adjusted, Δ (SD)		Treatment effect, Δ PN–Δcontrol (95% CI)	*P* value
	Control (n=91)	PN (n=96)	ΔControl (n=91)	ΔPN (n=96)		
BMI (kg/m^2)[d]						
Absolute BMI	24.2 (4.4)	24.8 (4.4)	–0.1 (0.1)	–0.1 (0.1)	0.0 (–0.23 to 0.18)	.79
Ideal BMI distance	3.5 (3.6)	3.7 (3.9)	0.0 (0.2)	–0.1 (0.2)	–0.1 (–0.29 to 0.11)	.37
PA (Baecke) score[e]						
Overall score	53.7 (9.6)	51.3 (9.5)	0.3 (1.7)	0.6 (1.6)	0.3 (–1.26 to 1.87)	.70
Leisure score	60.1 (13.1)	59.5 (13.3)	–2.1 (3.3)	–1.0 (3.5)	1.2 (–1.26 to 3.62)	.34
Sports score	55.8 (20.6)	49.5 (18.9)	2.9 (4.3)	0.4 (3.9)	–2.5 (–5.83 to 0.83)	.14
Work score	45.3 (11.6)	45.2 (10.4)	0.4 (1.6)	2.3 (1.3)	2.0 (0.29 to 3.63)	.02

[a]PA: physical activity.

[b]PN: personalized nutrition.

[c]Values presented as adjusted means.

[d]Presented as absolute variation and distance to the ideal BMI (21.75 kg/m^2).

[e]Values are reported on a scale between 0 and 100.

Personalized Web-Based Report Evaluation

Both the control and PN groups were presented with the PN report following the final FFQ, after which 108 provided complete feedback on the PN report. Most participants (95/108, 88%) did not report difficulty in understanding the report (question 1; Table 7). Of those who did, comments largely related to the stages before the PN advice itself (n=8), including minor issues related to the FFQ (n=3), Baecke questionnaire (n=3), and difficulties finding the link to the web-based report (n=2). Comments relating to the PN advice covered the desire to see the scientific evidence for the recommendations (ie, details of the m-AHEI score calculations) (n=1) and disagreement with the personalized advice received (n=4).

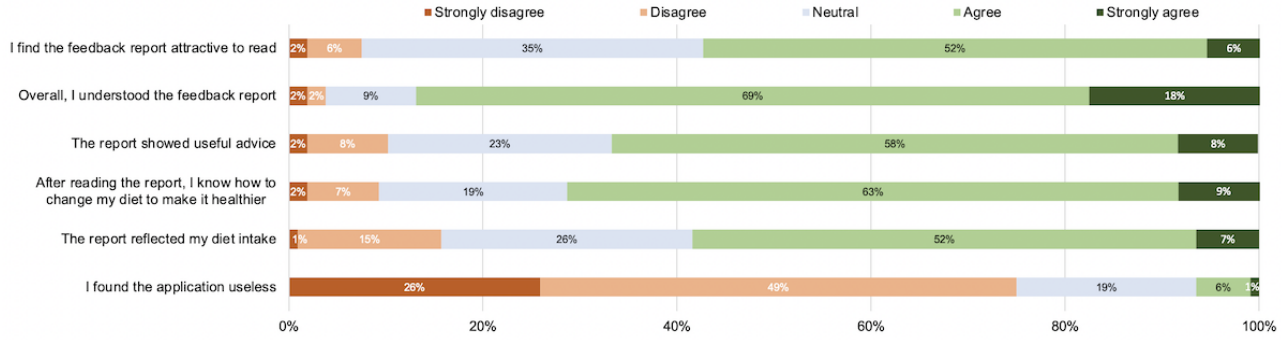
Table 7. Qualitative user feedback for the open questions related to the personalized report (N=108).

Question	Yes, n (%)	No, n (%)
Question 1: Was there anything in the report that you found particularly difficult to understand?	13 (12)	95 (88)
Question 2: Do you need additional information to help you make changes to your diet at this moment?	11 (10.2)	97 (89.8)
Question 3: Do you have any further comments regarding the feedback you received?	16 (14.8)	92 (85.2)

In response to question 2 (Table 7), 83% (5/6) of the comments provided were about barriers to healthy eating (eg, “more time to prepare meals”), and 1 participant requested more explanation of the advice (“If you want me to follow advice, I would like to understand the basis”). Of the 14 comments received in response to the third question, 3 (21%) related to the FFQ, 5 (36%) were about the limitations of the PA feedback (eg, “I do not think the report is a reflection on my sporting activity”), and the other 6 (43%) were about the diet recommendations, most of which (4/6, 67%) mentioned their partial disagreement with some of the diet advice (eg, “I do not agree with the advice to increase dairy foods. This is a very narrow view of the full picture,” and “I have too much salt and meat, but I don’t think I do”).

The results of the questions related to the quality of the report design and the perceived effectiveness of the recommendations [33] are shown in Figure 2 using a Likert scale. Although most participants (94/108, 87%) agreed to understanding the feedback, less reported (*agree* and *strongly agree*) knowing how to change their diet following feedback (77/108, 71.2%) and finding the advice useful (72/108, 66.7%). In addition, 15.7% (17/108) of participants felt that the report did not reflect their dietary intake, and 7% (7/108) reported finding the eNutri app useless.

Figure 2. User evaluation of the web-based personalized nutrition report using a Likert scale (N=108). Inconsistencies in the sum of percentages is due to the rounding of the percentages.

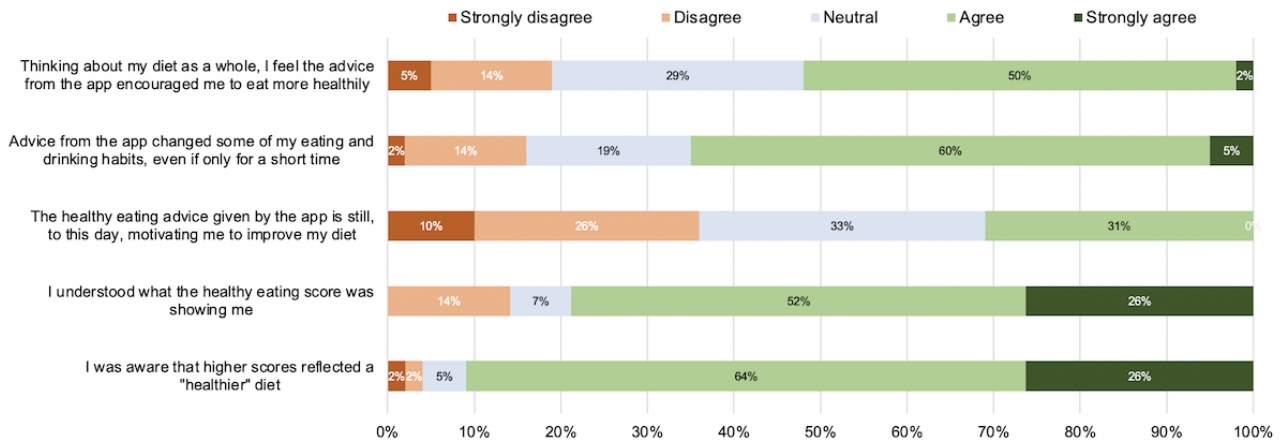


Follow-up Questionnaire

In total, 82 participants returned to complete the follow-up questionnaire that was administered 4.6 months after the main study ended; 42 (51%) of these were in the PN group (n=96) and included in the analysis. The mean follow-up time was 5.9 months (range 4.6-7.6 months, SD 0.65).

Over half of the PN participants agreed or strongly agreed that the advice encouraged them to eat more healthily (22/42, 52%) and that it changed some of their eating or drinking habits (27/42, 65%), with almost one-third (13/42, 31%) claiming that the advice was still motivating them to improve their diet 5.9 months (mean follow-up) after the study had ended (Figure 3). However, when asked, "Compared to before you started the study, are you still following ANY of the healthy eating advice in your diet now, no matter how small the changes?" this value was much greater, with 64% (27/42) responding "Yes" compared with 7% (3/42) "No" and 29% (12/42) responding "I did not make any changes to my diet whilst using the app." It should be noted that during the follow-up questionnaire, 74% (31/42) of the participants indicated they were "already motivated" to make changes to their diet before starting the study. Furthermore, the inclusion of their m-AHEI (*healthy eating*) score in the personalized advice was generally understood (33/42, 79% agreed or strongly agreed).

Figure 3. Follow-up questionnaire responses in the personalized nutrition group using a Likert scale (N=42). Inconsistencies in the sum of percentages is due to the rounding of the percentages.



The three highest rated reasons why participants did not follow the advice given were "The recommended foods didn't fit into my usual meal plans/recipes" (20/42, 47% agreed or strongly agreed), "The advice was too general and unrelated to my diet" (16/42, 38%), and "The advice was not suited to my lifestyle" (13/42, 31%). The reasons most participants disagreed or strongly disagreed with were "I was concerned my weight would change" (31/42, 74%), "I wasn't willing to try new foods" (31/42, 74%), and "I won't change certain aspects of my diet, regardless of the advice" (30/42, 71%). When asked to rate which features would be helpful additions to eNutri, the top three were "More information about the most positive aspects of my diet" (25/42, 60% rated this *extremely* or *very* helpful), "Advice on the vitamins and minerals to improve in my diet" (27/42, 64%), and "Advice on my overall energy (calorie) intake" (23/42, 55%).

Participants were also asked to write a review of the app, in which some positive comments were provided, such as "The dietary advice was relevant and easy to implement into my diet, and was evidence based so felt trustworthy. From just a short questionnaire, it was able to give personalised recommendations and practical ways to implement them into my diet. A great resource for anyone wanting to adopt a healthier dietary pattern" and "A few of the results I got from the app were surprising and encouraged me to eat differently—particularly eating more

whole grains and eating less processed meat—neither of which were diet modifications that had ever occurred to me."

Discussion

Principal Findings

This RCT was designed primarily to test whether eNutri's personalized food-based dietary web-based advice, using the m-AHEI as the foundation of the decision engine, was more effective than generalized population advice for motivating beneficial dietary change. The significant treatment effect (3.5 points in the m-AHEI scale) represented an increase of 6.1% in the mean baseline m-AHEI score of 57.5 points. This result supports the hypothesis that the eNutri app is an effective web-based tool for PN advice for UK adults.

A total of 2 m-AHEI component scores, *nuts and legumes* and *red and processed meat*, significantly increased, indicating beneficial dietary change. All component scores increased more in the PN group than in the control group, except for *vegetables*, indicating that the personalization could potentially have reached significance with more participants. The reduced population consumption of red and processed meats has become a priority for many health organizations (eg, the World Health Organization and World Cancer Research Fund) owing to the observed reduction in colorectal cancer risk following reduced intake (particularly in high consumers) and the significant impact of these items on food-related greenhouse gas emissions impact [41].

The most frequently presented components, representing dietary intakes that deviated most from the dietary recommendations in the m-AHEI, were *red and processed meat*, *nuts and legumes*, *whole grains*, *salt*, and *sugars*. Targeted analysis of m-AHEI components, based upon those presented to participants or matched controls, revealed significant decreases in *salt* and *alcohol* intake (ie, greater m-AHEI scores) following PN advice as compared with control advice. However, *alcohol* was presented in only 2.7% (8/187) of cases; thus, further data are required to confirm these findings with larger sample sizes.

The observed change in diet quality between the 2 groups was 3.5 points out of 100 (or 3.9 points out of 110 before scaling down). Although relatively small, long-term dietary changes of this magnitude are likely to have a positive impact on health if continued over time. For example, the UK Whitehall II study demonstrated that an increase of 1 SD (equivalent to 9.8 points out of 110) in the 2010 AHEI (on which the m-AHEI was based) was associated with a reduced risk of all-cause (–22%) and CVD mortality (–20%) after a 22-year follow-up [42]. Therefore, an increase in diet quality by 3.9 points maintained long term could potentially reduce the risk of CVD mortality to some degree, although further investigation of the m-AHEI score is warranted. Similarly, findings from the UK Caerphilly Prospective study reported that the risk of stroke was 17% lower for every 1 SD increase in the 2010 AHEI (equivalent to 11 points) in middle-aged men after a mean follow-up of 16.6 years [43]. There is an absence of data examining the impact of changes in diet quality on risk factors of disease in the shorter term, and this is an area worthy of further work.

Before the EatWellUK study, the most closely related and important work was the Food4Me study [44], in which 1269 participants from 7 European countries completed a 6-month PN study. The Food4Me study used the Healthy Eating Index (HEI) [45], which was the basis for the AHEI [18], as a secondary outcome measure of diet quality. Their treatment effect on the overall HEI was 1.27 points out of 100 (95% CI 0.30-2.25; P=.01) at 6 months, suggesting a significant improvement in diet quality following PN advice. Participants randomized to receive PN advice were reported to consume less red meat, salt, saturated fat, and energy and also increased their folate intake [12]. Although statistically significant, the increase in HEI in Food4Me was relatively small, confirming the challenge to encourage healthier diets and the need of similar studies.

A systematic review of remotely delivered dietary interventions (n=26) using self-monitoring or tailored feedback also concluded a significant, but small (and at risk of bias), positive effect on dietary change [14]. A total of 51 dietary outcomes were analyzed in the 23 interventions considered in the meta-analysis, resulting in an average of 2.2 dietary outcomes per intervention. The most popular ones were *fruits*, *vegetables*, and *fat*, and only 3 interventions targeted >4 dietary outcomes. This review also considered interventions delivered over the phone or offline media (eg, printed reports and CD-ROMs). Only 7 interventions used modern web-based methods, such as websites or apps. The differences in the dietary outcomes make the comparisons more difficult, especially because the changes in some dietary outcomes may affect other components not measured during the intervention; for example, a decrease in alcohol consumption may be associated with an increased consumption of sugar-sweetened beverages, owing to the dynamic aspect of diets.

A more recent systematic review (n=6) on the use of mobile apps to improve nutrition behaviors reported significant improvements in some objectives (eg, weight status) for 67% (4/6) the trials included [46]. Similar to eNutri, most apps (83%) in the review used self-assessment for feedback and monitoring. Social support in the form of personalized coaching calls [47] and a phone call and social networking feature [48] were used in 33% (2/6) of the studies. However, as noted by the review's authors, this requires significant financial input [49]. Regarding deployment costs, this version of eNutri can be deployed in Google Firebase (database and hosting) using the free plan, allowing reproducibility and scalability.

It should be noted that this study was powered to measure the overall m-AHEI treatment effect in all the participants but not the individual components. The fact that individual m-AHEI scores started from different baselines and were presented to the participants more or less frequently according to the participant's dietary intake makes it more difficult to reach statistical significance. For example, some m-AHEI components, such as *dairy products* and *alcoho*l started with mean baseline values close to the best possible score (≥88 out of 100) and were thus presented to small numbers of participants. To test the significance of the personalization of these diet messages, a much larger RCT would be necessary, which is viable over the internet.

At follow-up (mean 5.9 months after intervention), 64% (27/42) of PN participants agreed that the advice from the app had prompted them to change their eating or drinking habits and that they were still following some (*any*) of the advice. It is important to note that this sample represented only 44% (42/96) of the PN completers; therefore, these data should be interpreted with caution. However, maintenance of lifestyle changes, including dietary change, following intervention is arguably one of the greatest challenges that researchers and clinicians face; thus, it is encouraging to see reported benefits at follow-up in this study. A key consideration for personalized or precision nutrition, particularly using apps or wearables, is long-term user engagement and maintenance [50]. In a cross-sectional survey (n=217) designed to identify the impact of diet- and nutrition-related apps of health behavior change, both app engagement (P<.001) and theory-related constructs (P<.001) were positively associated with diet-related behavior change [51], with the authors recommending integration of appropriate theoretical constructs for health behavior change into app development [51]. Extensive user feedback, such as that collected in this study, is an important step to understanding how to drive future development.

The evaluation of the PN report showed that participants largely understood the report and were confident about the changes they were advised to make in their diets. Although users and nutrition professionals were consulted in the design and composition of the PN feedback, a small number of participants reported disagreeing with the advice provided (n=4) [16]. However, there is scope to enable user interaction within the app allowing the user to manipulate the advice to focus on the aspects they feel more capable of addressing, such as the choice to select which of the *top* components recommended for them they are prepared to change (ie, goal setting). There was also good acceptance of the content and design of the report, although the data suggested that the visual appearance (attractiveness) of the report could be improved, suggesting that further improvements in its design may be necessary. Of importance, most of the participants reported that they found the advice helpful.

In our study, a minimal decrease in the distance to the ideal BMI of 0.09 kg/m^2 was seen in the PN group at 12 weeks; however, this did not significantly differ from the change observed in the control group. A minimal decrease was expected, as the intervention was primarily targeted to healthy eating and relatively short in duration. In Food4Me, the researchers also reported no significant effect of personalized advice on BMI (–0.24 kg/m^2) relative to a control group [12], but it is difficult to compare the effectiveness on BMI because the authors did not report the distance to the ideal BMI, as proposed by this research. There was also no benefit of PA feedback based on the Baecke questionnaire, compared with the control advice in this study. It may confirm that more robust and personalized PA trackers, such as accelerometers, GPS, or pedometers, may be necessary for delivering effective interventions to increase PA.

Limitations

The power calculation for this study was based on the expected increase in the overall m-AHEI score. Because of recruitment challenges and high dropout rate (35.4%), the planned number of participants (n=274) was not reached, with 187 participants providing at least two valid FFQs. However, it is worthy of note that statistically significant differences in our primary outcome (m-AHEI score) were identified between the intervention and control groups. Additional studies with more participants, considering the baseline values (eg, by randomizing according to baseline diet quality) and distribution of messages may be necessary if the individual m-AHEI components are to be analyzed. Therefore, where advice on a particular component was delivered to only relatively few participants, such as *dairy products*, the effect of the advice on the component should be read cautiously considering the large CIs described. Although high, our dropout rate is within the range reported in a systematic review of engagement with remote measurement technology for managing health outcomes (0%-44%) [52]. Higher dropout rates appear synonymous with remote interventions, but there is scope to mitigate these by exploring user experiences (eg, barriers to use). The review notes the following as key drivers of engagement and re-engagement: health status, usability, convenience, accessibility, perceived utility, and motivation [52]. In the short term, more realistic dropout estimates should be included in sample size calculations for future studies.

This study used a modified US-based diet quality score for derivation of PN recommendations. However, the development of a diet quality score that aligns with the UK Eatwell Guide is warranted for future delivery and evaluations of dietary interventions.

Although the design of the diet messages followed the same protocol for each component [16], some messages were presented to only a few participants (eg, for *dairy products*); thus, the reported understanding of the report should not be generalized to all the textual diet messages. The fact that diet and weight were self-reported on the web may have impacted on the results, although dietary assessment was based on a previously validated FFQ [11]. Some participants may not have had weighing scales at home or were not able to weigh themselves for the subsequent app visits, and participants may have re-entered the original value without taking a new measurement. This may explain why 24 participants reported no change in weight during the study, which may have impacted the lack of statistical significance for changes in BMI. A face-to-face validation study (n=140) performed after the Food4Me study found that BMI was significantly lower (–0.29 kg/m^2; P<.001) for self-reported data, although BMI was still classified correctly in 93% of cases [53]. Underreporting bias for weight and BMI has also been reported previously [54,55].

A further limitation is that the FFQ was not repeated at the follow-up, which would have allowed the evaluation of longer-term dietary changes more accurately. This was to reduce the participant burden and address the low response rates in the main study. Instead, adherence to the dietary advice was self-reported in a questionnaire that may be subject to reporting

bias. Moreover, because both groups received PN dietary advice at the end of the study, it was not possible to quantitatively compare the 2 groups at the follow-up.

Conclusions

This novel study presented the treatment effects of a 12-week web-based RCT with 187 participants, which, to our knowledge, is the second largest web-based dietary intervention in the UK and the only one to deliver PN advice automatically [14]. This study aims to measure the effectiveness of a novel web-based PN advice tool (eNutri), using a modified version of the AHEI (a measure of diet quality) as the foundation of the decision engine to deliver web-based personalized food-based dietary advice. The results show that the design and protocol followed by the PN group in this study motivated greater change to follow a healthier diet relative to generalized dietary advice, as evaluated by an increase in diet quality. It is anticipated that the use of the eNutri app could contribute to improved diet quality if used more widely within the United Kingdom. The user evaluation of this study, via the web-based report evaluation and follow-up questionnaires, is important to improve eNutri and similar apps to motivate people to use them and follow the personalized advice given. The design principles and algorithms of eNutri can be used and improved by other researchers and institutions interested in web-based PN advice; the eNutri 1.0 web app was made publicly available under a permissive open-source license [56]. Larger studies evaluating the longer-term impact of automated PN interventions which include objective assessment of dietary intake and health outcomes, are recommended. This work represents an important landmark in the field of automatically delivered web-based dietary interventions, particularly those that are personalized to individual users.

Acknowledgments

The authors would like to thank Luke White for his assistance in conducting the study. RZF was sponsored by the National Council of Technological and Scientific Development of the Brazilian government. Participant retention incentives were funded by the British Nutrition Foundation under the Drummond Pump Priming Award 2017 (RF). The follow-up feedback questionnaire was funded by European Institute of Innovation and Technology Food, the innovation community on Food of the European Institute of Innovation and Technology, a body of the European Union (EU), under Horizon 2020, the EU Framework Programme for Research and Innovation (project ID: Quisper 2018 ID 18064).

Authors' Contributions

RZF was responsible for software development, data collection and analysis, and the drafting of the manuscript. RF, RZF, FH, and JAL were involved in the study design and app content. RF conducted the randomized controlled trial, and MW conducted and analyzed the follow-up feedback questionnaire. All authors reviewed and edited the manuscript and approved the final version for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Modified Alternative Healthy Eating Index components and scoring criteria.
[DOCX File , 25 KB - jmir_v24i4e29088_app1.docx]

Multimedia Appendix 2
Nutrition, weight, and physical activity advice received by the control group participants.
[DOCX File , 718 KB - jmir_v24i4e29088_app2.docx]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1220 KB - jmir_v24i4e29088_app3.pdf]

References

1. Global status report on noncommunicable diseases 2010. World Health Organization. 2010. URL: https://www.who.int/nmh/publications/ncd_report_full_en.pdf [accessed 2018-02-22]
2. Eat Well - fruit and vegetables: are you getting your 5 a day? National Health Service (NHS). 2019. URL: https://www.nhs.uk/live-well/ [accessed 2018-02-22]
3. Guidance: the eatwell guide. United Kingdom Government. 2016. URL: https://www.gov.uk/government/publications/the-eatwell-guide [accessed 2016-04-05]
4. Capacci S, Mazzocchi M. Five-a-day, a price to pay: an evaluation of the UK program impact accounting for market forces. J Health Econ 2011;30(1):87-98. [doi: 10.1016/j.jhealeco.2010.10.006] [Medline: 21129797]

5. Official statistics: NDNS: results from years 9 to 11 (combined) – statistical summary. United Kingdom Government. 2020. URL: https://www.gov.uk/government/statistics/ndns-results-from-years-9-to-11-2016-to-2017-and-2018-to-2019/ndns-results-from-years-9-to-11-combined-statistical-summary [accessed 2021-05-13]
6. Celis-Morales C, Lara J, Mathers JC. Personalising nutritional guidance for more effective behaviour change. Proc Nutr Soc 2015;74(2):130-138. [doi: 10.1017/S0029665114001633] [Medline: 25497396]
7. Brug J, Campbell M, van Assema P. The application and impact of computer-generated personalized nutrition education: a review of the literature. Patient Educ Couns 1999;36(2):145-156. [doi: 10.1016/s0738-3991(98)00131-1] [Medline: 10223019]
8. Betts JA, Gonzalez JT. Personalised nutrition: what makes you so special? Nutr Bull 2016;41(4):353-359. [doi: 10.1111/nbu.12238]
9. Noar SM, Grant Harrington N, Van Stee SK, Shemanski Aldrich R. Tailored health communication to change lifestyle behaviors. Am J Lifestyle Med 2011;5(2):112-122 [FREE Full text] [doi: 10.1177/1559827610387255]
10. Franco RZ, Fallaize R, Lovegrove JA, Hwang F. Popular nutrition-related mobile apps: a feature assessment. JMIR Mhealth Uhealth 2016;4(3):e85 [FREE Full text] [doi: 10.2196/mhealth.5846] [Medline: 27480144]
11. Fallaize R, Forster H, Macready AL, Walsh MC, Mathers JC, Brennan L, et al. Online dietary intake estimation: reproducibility and validity of the Food4Me food frequency questionnaire against a 4-day weighed food record. J Med Internet Res 2014;16(8):e190 [FREE Full text] [doi: 10.2196/jmir.3355] [Medline: 25113936]
12. Celis-Morales C, Livingstone KM, Marsaux CF, Macready AL, Fallaize R, O'Donovan CB, Food4Me Study. Effect of personalized nutrition on health-related behaviour change: evidence from the Food4Me European randomized controlled trial. Int J Epidemiol 2017;46(2):578-588. [doi: 10.1093/ije/dyw186] [Medline: 27524815]
13. Forster H, Walsh MC, Drevon CA, Manios Y, Traczyk I, Martinez JA, Food4Me study. Development and automation of a dietary feedback system for the delivery of personalised dietary advice. Proc Nutr Soc 2015;74(OCE4):E275. [doi: 10.1017/s0029665115003171]
14. Teasdale N, Elhussein A, Butcher F, Piernas C, Cowburn G, Hartmann-Boyce J, et al. Systematic review and meta-analysis of remotely delivered interventions using self-monitoring or tailored feedback to change dietary behavior. Am J Clin Nutr 2018;107(2):247-256 [FREE Full text] [doi: 10.1093/ajcn/nqx048] [Medline: 29529158]
15. Zenun Franco R, Fallaize R, Lovegrove JA, Hwang F. Online dietary intake assessment using a graphical food frequency app (eNutri): usability metrics from the EatWellUK study. PLoS One 2018;13(8):e0202006 [FREE Full text] [doi: 10.1371/journal.pone.0202006] [Medline: 30096211]
16. Fallaize R, Franco RZ, Hwang F, Lovegrove JA. Evaluation of the eNutri automated personalised nutrition advice by users and nutrition professionals in the UK. PLoS One 2019;14(4):e0214931 [FREE Full text] [doi: 10.1371/journal.pone.0214931] [Medline: 30943252]
17. Fairbrother lecture 2018: would you let your phone tell you what to eat? University of Reading. 2018. URL: https://www.youtube.com/watch?v=nIiB-w80hhM [accessed 2021-05-13]
18. Chiuve SE, Fung TT, Rimm EB, Hu FB, McCullough ML, Wang M, et al. Alternative dietary indices both strongly predict risk of chronic disease. J Nutr 2012;142(6):1009-1018 [FREE Full text] [doi: 10.3945/jn.111.157222] [Medline: 22513989]
19. McCullough ML, Willett WC. Evaluating adherence to recommended diets in adults: the alternate healthy eating index. Public Health Nutr 2006;9(1A):152-157. [doi: 10.1079/phn2005938] [Medline: 16512963]
20. Fallaize R, Livingstone KM, Celis-Morales C, Macready AL, San-Cristobal R, Navas-Carretero S, et al. Association between diet-quality scores, adiposity, total cholesterol and markers of nutritional status in European adults: findings from the Food4Me study. Nutrients 2018;10(1):49 [FREE Full text] [doi: 10.3390/nu10010049] [Medline: 29316612]
21. Kromhout D, Spaaij CJ, de Goede J, Weggemans RM. The 2015 Dutch food-based dietary guidelines. Eur J Clin Nutr 2016;70(8):869-878 [FREE Full text] [doi: 10.1038/ejcn.2016.52] [Medline: 27049034]
22. Food-based dietary guidelines - France. Food and Agriculture Organization of the United Nations. URL: http://www.fao.org/nutrition/education/food-based-dietary-guidelines/regions/countries/france/en/ [accessed 2018-04-05]
23. Carbohydrates and health. Scientific Advisory Committee on Nutrition. 2015. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/445503/SACN_Carbohydrates_and_Health.pdf [accessed 2018-02-22]
24. Swan GE, Powell NA, Knowles BL, Bush MT, Levy LB. A definition of free sugars for the UK. Public Health Nutr 2018;21(9):1636-1638 [FREE Full text] [doi: 10.1017/S136898001800085X] [Medline: 29587886]
25. Meat, fish and dairy products and the risk of cancer. World Cancer Research Fund/American Institute for Cancer Research. 2018. URL: https://www.wcrf.org/dietandcancer/meat-fish-and-dairy/ [accessed 2021-05-13]
26. French agency ANSES slashes recommended meat intake in new guidelines. Food Navigator. 2017. URL: https://www.foodnavigator.com/Article/2017/01/30/French-agency-ANSES-slashes-recommended-meat-intake-in-new-guidelines [accessed 2021-02-01]
27. Actualisation des repères du PNNS: révision des repères de consommations alimentaires. Anses. 2016. URL: http://www.caloris.fr/projets/17-01-ed-rapport-pnns-revision-reperes-1.pdf [accessed 2021-03-15]
28. Rehm J, Mathers C, Popova S, Thavorncharoensap M, Teerawattananon Y, Patra J. Global burden of disease and injury and economic cost attributable to alcohol use and alcohol-use disorders. Lancet 2009;373(9682):2223-2233. [doi: 10.1016/S0140-6736(09)60746-7] [Medline: 19560604]

29. Analytics help: URL builders: generate custom campaign parameters for your advertising URLs. Google. URL: https://support.google.com/analytics/answer/1033867?hl=en [accessed 2018-01-29]
30. Analytics for angularJS applications. Github: angulartics. 2017. URL: https://github.com/angulartics/angulartics [accessed 2021-05-13]
31. Math.random() - JavaScript. Mozilla Developer Network Web Docs. 2017. URL: https://developer.mozilla.org/en-US/docs/Web/JavaScript/Reference/Global_Objects/Math/random [accessed 2021-05-13]
32. Baecke JA, Burema J, Frijters JE. A short questionnaire for the measurement of habitual physical activity in epidemiological studies. Am J Clin Nutr 1982;36(5):936-942. [doi: 10.1093/ajcn/36.5.936] [Medline: 7137077]
33. Knijnenburg BP, Willemsen MC, Gantner Z, Soncu H, Newell C. Explaining the user experience of recommender systems. User Model User-Adap Inter 2012;22(4-5):441-504. [doi: 10.1007/s11257-011-9118-4]
34. A quick guide to the government's healthy eating recommendations. Public Health England. 2018. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/742746/A_quick_guide_to_govt_healthy_eating_update.pdf [accessed 2021-05-13]
35. Henry CJ. Basal metabolic rate studies in humans: measurement and development of new equations. Public Health Nutr 2005;8(7A):1133-1152. [doi: 10.1079/phn2005801] [Medline: 16277825]
36. Welch AA, Luben R, Khaw KT, Bingham SA. The CAFE computer program for nutritional analysis of the EPIC-Norfolk food frequency questionnaire and identification of extreme nutrient values. J Hum Nutr Diet 2005;18(2):99-116. [doi: 10.1111/j.1365-277X.2005.00593.x] [Medline: 15788019]
37. Bates B, Cox L, Page SN, Prentice A, Steer T, Swan G. National Diet and Nutrition Survey: results from years 5 and 6 (combined) of the Rolling Programme (2012/2013 – 2013/2014): a survey carried out on behalf of Public Health England and the Food Standards Agency. Public Health England. 2016. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/551352/NDNS_Y5_6_UK_Main_Text.pdf [accessed 2021-05-13]
38. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with Python. In: Proceedings of the 9th Python in Science Conference. 2010 Presented at: SciPy '10; June 28-July 3, 2010; Austin, TX. [doi: 10.25080/majora-92bf1922-011]
39. Pearson ES, D'"Agostino RB, Bowman KO. Tests for departure from normality: comparison of powers. Biometrika 1977;64(2):231-246. [doi: 10.1093/biomet/64.2.231]
40. Vickers AJ, Altman DG. Statistics notes: analysing controlled trials with baseline and follow up measurements. BMJ 2001;323(7321):1123-1124 [FREE Full text] [doi: 10.1136/bmj.323.7321.1123] [Medline: 11701584]
41. Aston LM, Smith JN, Powles JW. Impact of a reduced red and processed meat dietary pattern on disease risks and greenhouse gas emissions in the UK: a modelling study. BMJ Open 2012;2(5):e001072 [FREE Full text] [doi: 10.1136/bmjopen-2012-001072] [Medline: 22964113]
42. Shivappa N, Hebert JR, Kivimaki M, Akbaraly T. Alternative Healthy Eating Index 2010, Dietary Inflammatory Index and risk of mortality: results from the Whitehall II cohort study and meta-analysis of previous Dietary Inflammatory Index and mortality studies. Br J Nutr 2017;118(3):210-221. [doi: 10.1017/S0007114517001908] [Medline: 28831955]
43. Shan Z, Li Y, Baden MY, Bhupathiraju SN, Wang DD, Sun Q, et al. Association between healthy eating patterns and risk of cardiovascular disease. JAMA Intern Med 2020;180(8):1090-1100 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2176] [Medline: 32539102]
44. Celis-Morales C, Livingstone KM, Marsaux CF, Forster H, O'Donovan CB, Woolhead C, et al. Design and baseline characteristics of the Food4Me study: a web-based randomised controlled trial of personalised nutrition in seven European countries. Genes Nutr 2015;10(1):450 [FREE Full text] [doi: 10.1007/s12263-014-0450-2] [Medline: 25491748]
45. Guenther PM, Casavale KO, Reedy J, Kirkpatrick SI, Hiza HA, Kuczynski KJ, et al. Update of the Healthy Eating Index: HEI-2010. J Acad Nutr Diet 2013;113(4):569-580 [FREE Full text] [doi: 10.1016/j.jand.2012.12.016] [Medline: 23415502]
46. Paramastri R, Pratama SA, Ho DK, Purnamasari SD, Mohammed AZ, Galvin CJ, et al. Use of mobile applications to improve nutrition behaviour: a systematic review. Comput Methods Programs Biomed 2020;192:105459. [doi: 10.1016/j.cmpb.2020.105459] [Medline: 32234632]
47. 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;13:7 [FREE Full text] [doi: 10.1186/s12966-016-0329-2] [Medline: 26785637]
48. 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;161(10 Suppl):S5-12 [FREE Full text] [doi: 10.7326/M13-3005] [Medline: 25402403]
49. Probst Y, Morrison E, Sullivan E, Dam HK. First-stage development and validation of a web-based automated dietary modeling tool: using constraint optimization techniques to streamline food group and macronutrient focused dietary prescriptions for clinical trials. J Med Internet Res 2016;18(7):e190 [FREE Full text] [doi: 10.2196/jmir.5459] [Medline: 27471104]
50. Gambhir SS, Ge TJ, Vermesh O, Spitler R. Toward achieving precision health. Sci Transl Med 2018;10(430):eaao3612 [FREE Full text] [doi: 10.1126/scitranslmed.aao3612] [Medline: 29491186]

51. West JH, Belvedere LM, Andreasen R, Frandsen C, Hall PC, Crookston BT. Controlling your "App"etite: how diet and nutrition-related mobile apps lead to behavior change. JMIR Mhealth Uhealth 2017;5(7):e95 [FREE Full text] [doi: 10.2196/mhealth.7410] [Medline: 28694241]
52. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and facilitators of engagement with remote measurement technology for managing health: systematic review and content analysis of findings. J Med Internet Res 2018;20(7):e10480 [FREE Full text] [doi: 10.2196/10480] [Medline: 30001997]
53. Celis-Morales C, Livingstone KM, Woolhead C, Forster H, O'Donovan CB, Macready AL, et al. How reliable is internet-based self-reported identity, socio-demographic and obesity measures in European adults? Genes Nutr 2015;10(5):28 [FREE Full text] [doi: 10.1007/s12263-015-0476-0] [Medline: 26143178]
54. Lassale C, Péneau S, Touvier M, Julia C, Galan P, Hercberg S, et al. Validity of web-based self-reported weight and height: results of the Nutrinet-Santé study. J Med Internet Res 2013;15(8):e152 [FREE Full text] [doi: 10.2196/jmir.2575] [Medline: 23928492]
55. Pursey K, Burrows TL, Stanwell P, Collins CE. How accurate is web-based self-reported height, weight, and body mass index in young adults? J Med Internet Res 2014;16(1):e4 [FREE Full text] [doi: 10.2196/jmir.2909] [Medline: 24398335]
56. Zenun R. eNutri - personalised nutrition web app. Zenodo. 2018. URL: https://doi.org/10.5281/zenodo.1304929 [accessed 2021-05-13]

Abbreviations

AHEI: Alternative Healthy Eating Index
CVD: cardiovascular disease
FFQ: Food Frequency Questionnaire
HEI: Healthy Eating Index
m-AHEI: modified Alternative Healthy Eating Index
PA: physical activity
PN: personalized nutrition
RCT: randomized controlled trial

Edited by R Kukafka; submitted 20.05.21; peer-reviewed by J Woodside, S Ashraf; comments to author 10.09.21; revised version received 12.11.21; accepted 16.12.21; published 25.04.22.

Please cite as:
Zenun Franco R, Fallaize R, Weech M, Hwang F, Lovegrove JA
Effectiveness of Web-Based Personalized Nutrition Advice for Adults Using the eNutri Web App: Evidence From the EatWellUK Randomized Controlled Trial
J Med Internet Res 2022;24(4):e29088
URL: https://www.jmir.org/2022/4/e29088
doi:10.2196/29088
PMID:35468093

Original Paper

What Injured Workers With Complex Claims Look For in Online Communities: Netnographic Analysis

Michelle Manning Hutson[1], BSc (Hons), MPH; Sarah M Hosking[1], PhD; Soula Mantalvanos[2], MA; Michael Berk[1,2,3], MBBS, PhD; Julie Pasco[1,3,4,5], PhD; Trisha Dunning[3,6], RN, MEd, PhD

[1]School of Medicine, Deakin University, Geelong, Australia

[2]Pain Train, Geelong, Australia

[3]Barwon Health, Geelong, Australia

[4]Department of Medicine – Western Health, Melbourne Medical School, The University of Melbourne, Melbourne, Australia

[5]Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Australia

[6]Centre for Quality and Patient Safety Research, Barwon Health Partnership, School of Nursing and Midwifery, Deakin University, Geelong, Australia

Corresponding Author:
Michelle Manning Hutson, BSc (Hons), MPH
School of Medicine
Deakin University
PO Box 281
Geelong, 3220
Australia
Phone: 61 3 4215 3320
Email: manningm@deakin.edu.au

Abstract

Background: Improved understanding of social constructs around injury may help insurance case managers to understand how best to support people after injury.

Objective: This study sought to explore what people who sustain work-related injuries may seek from online communities. The study highlights potential opportunities for improved engagement with insurance case management practice.

Methods: An observational netnographic analysis was undertaken on anonymous, publicly available messages posted on Australian message boards. All research data were drawn from anonymous, online communities. A person (author SM) with experience of making a claim through an Australian workers' compensation system and online engagement was involved in study conception, design, and analysis. Data were analyzed using NVivo12 in an iterative, multistage process including coding, journaling, and member checking. A total of 141 people were engaged in discussion across 47 threads housed on 4 Australian forums.

Results: In this qualitative study, themes emerged from the data, describing how injured workers use online communities to help make decisions, get support, and solve problems. The key motivators for action and engagement were seeking information, connection, or justice. Establishment of relationships was a key mediator of each of these parameters.

Conclusions: Some work-related injuries may involve medical and medicolegal complexity as well as changed lifestyle and routine during convalescence and recovery. The mechanism used by some injured workers to seek information and problem solve suggests a capacity for self-management and self-care after work-related injury. Netnography provides information on a community that may not regularly engage with research because of the complexity of their situation and their vulnerability.

(J Med Internet Res 2022;24(4):e17180) doi:10.2196/17180

KEYWORDS

work-related injury; care coordination; case management; netnography

Introduction

Recovery from work-related injuries involves more than a clear clinical path. Injured workers can also be impacted by psychological and lifestyle factors, which moderate both recovery and return to activities of daily living [1]. Biopsychosocial factors are known to influence recovery outcomes, overlaying other more obvious aspects of recovery

such as biomechanics and access to care [2]. Recovery responses are mediated not only by compliance with treatment, but also by the expectations of the treatment team, friends and family, work colleagues, and society at large [3]. Work-related injuries can be costly for organizations, insurers, and individuals, so understanding the experiences of injured workers may help improve care outcomes and reduce costs.

Many individuals making a claim for compensable injuries describe feeling stigmatized [4]. Previous qualitative research studies with injured workers describe some positive experiences, mitigated by a variety of negative experiences [5,6]. Calvey and Jansz [6] highlight opportunities for practice improvement, which largely center around communication and information quality, and maintaining a sense of connection with work teams and treatment teams.

Research with potentially vulnerable injured workers must take into account the impact of power differentials, and digital ethnographic research methods provide a useful way to undertake sensitive observational research addressing experiences and perceptions in a public arena [7]. Qualitative, human-centered web research techniques such as netnography offer a valid way to describe social phenomena in a less invasive, cost-effective manner. Qualitative studies highlight the need to ensure quality communication with workers to promote improved outcomes following work-related injury [8].

This study sought to explore why injured workers participate in online communities, and to understand what injured workers may be seeking to promote or expedite their recovery. This research sought to discern areas for practice improvement for claim management, with a view to improve worker experience and health outcomes. The explorative nature of this work seeks to inform future research directions.

Methods

Overview

A person with experience of making a claim through an Australian workers' compensation system and online engagement (author SM) was involved in study conception, design, and analysis. Material was sourced from a variety of publicly available forums or message boards, all of which had an explicit position statement indicating that information shared on the board or website was publicly available, and all with a policy requiring users to maintain anonymity for themselves and others. Forums were selected as the most appropriate medium on the internet (compared with websites or social media) because of their interactivity and public nature. Forums were identified through a Google search for "work", "claim", and the truncated wildcard search stems "injur*" and "comp*". Forum selection was based upon the search terms; in addition, only forums with a minimum of 5 posts per thread and minimum of 5 threads that discussed work-related injury were selected. Privately held websites that did not have policies on poster anonymity or were not designed as discussion forums were excluded because of the potential risk of reidentification of posters. Anonymous posts from Australian websites were included, based upon either a variant of "Australia" in the forum title or the use of Australia's URL domain extension (.au). Data were collected from March to June 2019.

Data were analyzed using NVivo12 (QSR International). Code development was an iterative, multistage process, including attribute coding, descriptive coding, and modified versus coding, following a process described by Saldana [9]. In order to become familiar with the data corpus, the complete data corpus was reviewed 4 times before coding began, and a codebook acted as a reference for a priori codes. Analytic memos were used to record research ideas, as a research log and as part of the audit trail. To promote rigor and reliability of results, reflexive journaling was undertaken and member checking occurred with members of the research team and individuals with lived experience of compensable injury, as described by Liamputtong [10]. We used the Standards for Reporting Qualitative Research Checklist when writing our report [11].

Ethical Considerations

The study design was observational, not participatory, and used only publicly available archival data. The researchers did not elicit information from participants or participate in discussion. We considered the position of Eysenbach and Till [12] in the development of our methodology, aiming to reduce intrusiveness through passive analysis of anonymous, publicly available data. Our intention was to maintain privacy, dignity, and integrity for the posters who participated in the forums, and we excluded forums with fewer than 4 threads addressing the topic of work-related injury to minimize risk of reidentifiability.

Results

Overview

Community members who posted in threads were considered "produsers" as they both consumed and generated content in their engagement with the community [13]. Although it is known that active posters, people who visit such forums and do not post (lurkers), and nonusers of online forums may differ significantly, data on the latter two groups was unavailable [14]. The primary sources were 4 individual forums, yielding a total of 47 threads discussing work-related injury. In total, there were 70,206 words in the data corpus. There were 141 individuals interacting within the data corpus, which spanned from 2006-2019. No analysis of changes in the use of social media over time was undertaken. As the study was observational rather than interventional, assessment of change of psychological states over time was not possible. A total of 26 codes (6 a priori, 20 emerging, 19 in vivo) were described, as seen in Table 1.

Table 1. Themes emerging from the data corpus.

Coding, themes, and subthemes
A priori codes
Advice
Claim management
Management and treatment
Physical to psychological injury
Support
Venting
Emergent codes
Legal
Entitled
Legalities
Operational
Decision-making
Financial stressors
Practical loss
Physical, emotional
Experiences of pain
Grief and loss
Justice and injustice
Mindfulness and acceptance
Suicide risk
Trust and mistrust
Validation
Relationship development
Humor and black humor
Mirroring
Relationships emerging
Within and between systems, navigation
Expected responses
Health literacy
System design and navigation

Characteristics of Websites With Injured Worker Communities

The study sought to examine Australian forums; however, the forums were anonymous and thus geographic limitations could not be applied as this would require verification of identity. Forums provided less restrictive character or word limits than other digital spaces. There was an apparent preference for sites where weblog-like dialogue could be shared, rather than performative communities like Instagram, TikTok, or YouTube, as evaluated by number of posts available and amount of interaction achieved by the poster. Private conversations as may be found in private Facebook groups or WhatsApp, Weebo, or Messenger groups were excluded because of their private nature. Online participants largely communicated with words rather than imagery. Although members of the community often used humor and black humor, they did not appear to use memes or gifs as seen in other online communities. Many within the group used emojis or evocative profile pictures to supplement their text; however, these were not included in the analysis. The lack of use of enriched images such as memes, gifs, or video may indicate the intense desire for anonymity or a potential lack of confidence or comfort with technology within this cohort.

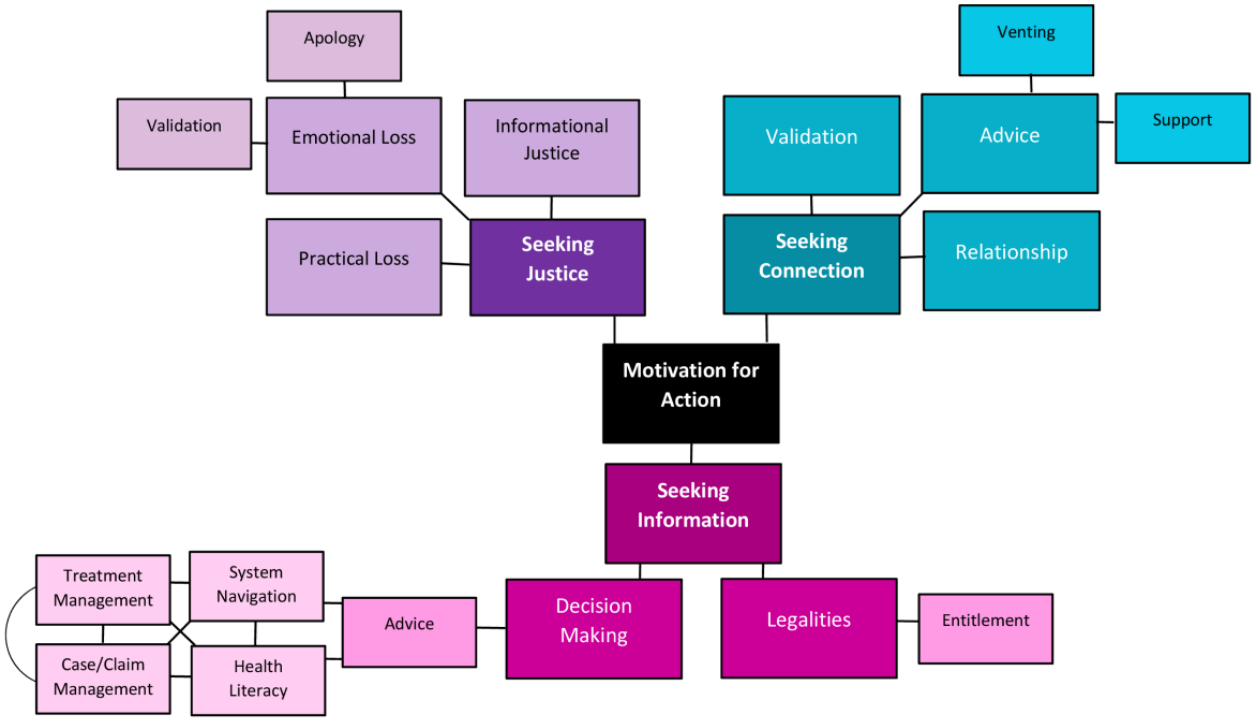
At times, posts were created in threads that appeared to function as a diary, with factual information and little engagement with others; these may provide a historical record, which may be externally validated by providing timestamps or other metadata.

An individual worker may be able to identify their own content; however, it would be difficult for employers, investigators, or lawyers to reidentify posters because of the anonymity policy of the boards. This feature allows forums to provide a record or history, acting as a form of documentary evidence for isolated workers. Further to this, Hinton and Hjorth [15] notes some digital content is produced not to elicit or provoke response, but remind the community that an individual is still there, which has particular relevance given the social isolation that can be endemic in this community.

A large proportion of individuals described complicated or complex injuries; people with straightforward claims appeared less likely to engage or participate in the communities identified. Those who engaged in these communities were following a model of interaction (Figure 1) that appeared to be underpinned by a need for validation; however, they were consistently seeking justice, validation, or connection, or combinations of justice, validation, and connection.

Others in the community seldom provided advice without also providing validation or empathy to the original poster. This interplay of seeking advice while receiving validation or empathy supported the development and emergence of relationships within groups.

Figure 1. Conceptual framework describing motivation for injured worker community interactions.



Seeking Justice

It appeared that individuals desired validation from members of their community—they described acknowledgment of their situation or difficulty by another as validation. Elbers [16] explains that perceptions of fairness may moderate health outcomes for individuals with compensable injuries. Individuals talked generally about the importance of acknowledgment or an apology following a work-related injury ("I'm still hurt and angry because no one will listen to me and try to acknowledge that what I'm going through is real and valid" [B19]).

The individual or organization providing the apology was less important than providing an acknowledgment of the worker's situation ("I just want to feel validated for my experiences and injury… most injured workers just want an apology… so you can move forward" [Z18]).

Although some sought financial recompense for their injury, many others noted that an acknowledgment or validation of their suffering would go some way toward supporting their recovery. Previous work has noted the importance of empathy in increasing perceptions of interpersonal and informational justice in organizations [17].

Information

Individuals often noted difficulty with navigating complex medicolegal systems, or alluded to having difficulty with interpreting information provided by insurance companies or the workplace Human Resources department. Some workers displayed frustration directed toward insurers or their agents: "It's like they [insurance companies] want us to die to avoid paying out" [M17]. Individuals also described experiences of how they are treated as influencing their perceptions of outcomes or mental states: "I feel like dealing with them is going to break me" [K19].

Some posters indicated that they required formalized assistance with navigating complex systems. Although some community members noted problems interacting with lawyers or understanding legal complexities, others recommended engaging a lawyer as an advocate in addition to the treatment and claim management teams:

> *I have 2 [Psychiatric Independent Medical Examiners], my [Psych] and my doctors reports all saying I'll never work again. I also have a no win no fee lawyer who is handling everything.* [G17]

Individuals often described complex circumstances and empathized with others who were trying to navigate the system without support. They often engaged with a discussion of difficulty navigating systems by listening and empathizing rather than by giving advice.

Connection

Many appeared to be motivated by connection and relationship with other members of the community in an altruistic way ("Sitting with you" [B19]), although others described emotional needs that had been unmet by the system or other stakeholders within the formal system ("the word frustration is inadequate compared to the suffering endured" [O13]). As individuals developed trust and rapport within the community, they turned to other members of the community for justice, information, connection, or a combination of these.

Discussion

Principal Findings

In this qualitative study of online communities, the broad themes of justice, information, and connection emerged. Many individuals described how acknowledgement and validation may contribute to or form part of their recovery. For some injured workers, online communities act as a place to find support, to aid decision-making, or to problem solve. The establishment of relationships was a key mediator of each of these parameters.

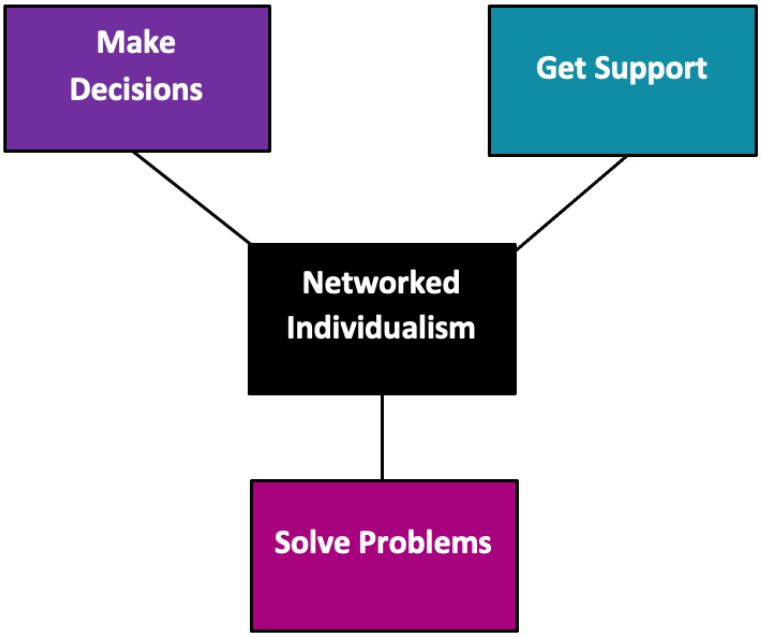
Oldenburg describes a "third place" outside of home and the workplace, which serves to meet a distinct social and emotional need [18]. Soukup [18] describes online communities as an essential modern "third place," with online communities acting as a social and emotional surrogate for real-world interaction. Calvey and Jansz [6] describe the consumer-focused service delivery model for provision of care for injured workers, the heart of which is being treated as a person and a partner in activities, and giving and receiving empowerment, trust, and advocacy. The online injured worker community demonstrates features of this model, creating community, seeking and sharing information within their community, and creating a space for occasional levity. The phrase "I feel like" is used often, which has a less declarative tone and invites discourse and new perspectives. Indeed, "I feel like" occurred with twice the frequency within the data corpus as the less emotionally driven "I feel that." Often individuals prefaced a discussion of intangible concepts by using "I feel like" as a signal, such as "I feel like a forgotten soldier," when compared with more concrete concepts such as "I feel that my right knee is…" Members of the communities often engaged in relationship building, empathizing, and advising, as evidenced by repetition of terms, mirroring of concepts, and reflection on an aspect of a post before providing further information. This suggests that there is less desire for performative self-expression and more desire for engagement with other members of the group.

In addition to online message boards, other communities exist online. Personal websites such as WorkCover Victim's Diary or Pudendal Nerve [19] are not anonymous and serve as a diary or record, as well as being a source of information; they are not strictly designed to promote dialogue within the community, although they permit commenting on published articles. In contrast, message boards meet "third place" criteria such as having a sense of community, being accessible, being based on conversation, and including a sense of playfulness.

From a system navigation perspective, there are some health literacy implications: injured workers are actively seeking information and assistance to improve their ability to navigate complex systems. "Straightforward" claims are seldom represented: most describe claims which are medically, operationally, or psychosocially complex (eg, disputed entitlement, unexpected complications, delayed recovery). Many describe having made attempts to have their informational needs met through usual channels such as discussion with stakeholders including their employer, treatment team, or insurer. Further work is required to understand informational needs of injured workers to determine what aspects of formalized information sharing may need to be improved to better facilitate decision-making in an adversarial system—whether this is improved information quality, improved availability, or better timing of information sharing.

Hinton [15] described Wellman's [20] "networked individualism" (Figure 2) as a process through which individuals participate in the online world to make decisions, solve problems, or get support. This pattern of participation and engagement is apparent within the injured worker "produser" community, where individuals are seeking something beyond even "person-centered care" or "circles of support"—they are meaningfully engaged with a broader community of virtual strangers who provide information, collaboration, and validation. "Produsers" who act as networked individuals may be demonstrating their capacity for self-actualization and self-management of their recovery. Previous case management research has shown that in a consumer-directed model of care, the locus of power shifts from the case manager toward the individual [21]. As individuals move toward self-management, case managers provide information, enable self-empowerment, build capacity, and provide support to individuals to self-manage their care needs [21]. This research elicits the possibility that some injured workers may benefit from a style of case management that centers the locus of control away from the organization and supports proactive decision-making by injured workers.

Figure 2. Networked individualism in practice (adapted from Wellman [20]).



Although netnographic analysis permits a rapid and broad analysis of the online milieu, there are limitations inherent in using an observational technique. We did not actively engage participants in research; rather, we observed what was currently available online. Accordingly, not all workers were represented and it is not possible to determine whether there were underrepresented groups within the cohort. This study only investigates posters, not lurkers or nonusers, and the findings cannot be generalized. In addition, due to the specific nature of the Australian system, the results cannot necessarily be generalized to other systems and countries. Therefore, the findings must be interpreted and applied with great caution. This research has a role in informing practice improvement opportunities for case managers but should not be taken into account in actuarial models without further analysis.

Conclusions

An observational netnography of injured workers revealed how a sample of injured workers with complex care needs use networked individualism online to make decisions, get support, and solve problems. The study revealed broad themes of seeking justice, information, and connection, all of which were underpinned by a desire to find validation. This mechanism for information seeking and problem solving demonstrates a meaningful engagement with community and a capacity for self-management of their own care, as distinct from traditional models of insurance case management. The actuarial implications of these findings have not been explored, and accordingly the findings should be interpreted with cautious optimism in any scheme translation.

Acknowledgments

MMH holds a PhD scholarship provided by the Australian Government Research Training Program. SMH is supported by a Deakin University Dean's Research Postdoctoral Fellowship. MB is supported by a National Health and Medical Research Council (NHMRC) Senior Principal Research Fellowship (1059660 and 1156072).

Data Availability

The data that support the findings of this study are available from the corresponding author, MMH (ORCID: 0000-0002-3312-462X), upon reasonable request.

Authors' Contributions

MMH and SMH made substantial contributions to the conception and design of the work; the acquisition, analysis, and interpretation of the data for the work; and the drafting of the manuscript. SM made substantial contributions to the conception and design of the work, as well as the analysis and interpretation of data for the work. MB and JP made substantial contributions to the design of the work and interpretation of data for the work. TD made substantial contributions to the design of the work and the acquisition, analysis, and interpretation of data for the work. All authors reviewed the work and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Eggert S. Psychosocial factors affecting employees abilities to return to work. AAOHN J 2010 Feb;58(2):51-55. [doi: 10.3928/08910162-20100118-01] [Medline: 20128518]
2. Dean A, Matthewson M, Buultjens M, Murphy G. Scoping review of claimants' experiences within Australian workers' compensation systems. Aust Health Rev 2019 Aug;43(4):457-465. [doi: 10.1071/AH17244] [Medline: 30032739]
3. Beales D, Fried K, Nicholas M, Blyth F, Finniss D, Moseley GL. Management of musculoskeletal pain in a compensable environment: Implementation of helpful and unhelpful Models of Care in supporting recovery and return to work. Best Pract Res Clin Rheumatol 2016 Jun;30(3):445-467. [doi: 10.1016/j.berh.2016.08.011] [Medline: 27886941]
4. Kirsh B, Slack T, King CA. The nature and impact of stigma towards injured workers. J Occup Rehabil 2012 Jun;22(2):143-154. [doi: 10.1007/s10926-011-9335-z] [Medline: 22012555]
5. Beardwood BA, Kirsh B, Clark NJ. Victims twice over: perceptions and experiences of injured workers. Qual Health Res 2005 Jan;15(1):30-48. [doi: 10.1177/1049732304268716] [Medline: 15574714]
6. Calvey J, Jansz J. Women's Experience of the Workers Compensation System. Australian Journal of Social Issues 2016 Dec 22;40(2):285-311. [doi: 10.1002/j.1839-4655.2005.tb00972.x]
7. Pink S, Horst H, Postill J, Hjorth L, Lewis T, Tacchi J. Digital ethnography: principles and practice. London, UK: Sage Publications; 2016.
8. Australasian Faculty of Occupational Medicine. Compensable injuries and health outcomes.: Royal Australasian College of Physicians; 2001. URL: https://www.racp.edu.au/docs/default-source/advocacy-library/compensable-injuries-and-health-outcomes.pdf [accessed 2022-01-27]
9. Saldaña J. The coding manual for qualitative researchers. London, UK: Sage Publications; 2016.
10. Liamputtong P. Qualitative research methods. Third edition. South Melbourne, Vic: Oxford University Press; 2009.
11. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
12. Eysenbach G, Till JE. Ethical issues in qualitative research on internet communities. BMJ 2001 Nov 10;323(7321):1103-1105 [FREE Full text] [doi: 10.1136/bmj.323.7321.1103] [Medline: 11701577]
13. Bruns A. Blogs, Wikipedia, Second Life, and Beyond: From Production to Produsage. New York, NY: Peter Lang; 2008.
14. Gliddon E, Lauder S, Berk L, Cosgrove V, Grimm D, Dodd S, et al. Evaluating discussion board engagement in the MoodSwings online self-help program for bipolar disorder: protocol for an observational prospective cohort study. BMC Psychiatry 2015 Oct 14;15:243 [FREE Full text] [doi: 10.1186/s12888-015-0630-7] [Medline: 26462799]
15. Hinton S, Hjorth L. Understanding social media. First edition. London, UK: Sage Publications; 2013.
16. Elbers NA, Collie A, Hogg-Johnson S, Lippel K, Lockwood K, Cameron ID. Differences in perceived fairness and health outcomes in two injury compensation systems: a comparative study. BMC Public Health 2016 Jul 29;16:658 [FREE Full text] [doi: 10.1186/s12889-016-3331-3] [Medline: 27473747]
17. Patient DL, Skarlicki DP. Increasing Interpersonal and Informational Justice When Communicating Negative News: The Role of the Manager's Empathic Concern and Moral Development. Journal of Management 2008 Dec 22;36(2):555-578. [doi: 10.1177/0149206308328509]
18. Soukup C. Computer-mediated communication as a virtual third place: building Oldenburg's great good places on the world wide web. New Media & Society 2016 Jun 30;8(3):421-440. [doi: 10.1177/1461444806061953]
19. Pudendal Nerve. URL: https://pudendalnerve.com.au/ [accessed 2021-11-10]
20. Wellman B, Haythornthwaite C. The Internet in Everyday Life. Oxford, UK: Oxford University Press; 2002.
21. You EC, Dunt D, Doyle C. How would case managers' practice change in a consumer-directed care environment in Australia? Health Soc Care Community 2017 Jan;25(1):255-265. [doi: 10.1111/hsc.12303] [Medline: 26500161]

Edited by G Eysenbach; submitted 24.11.19; peer-reviewed by S Lauder, R Drake, Y Wang, S Li; comments to author 24.08.20; revised version received 15.06.21; accepted 27.07.21; published 07.04.22.

Please cite as:
Manning Hutson M, Hosking SM, Mantalvanos S, Berk M, Pasco J, Dunning T
What Injured Workers With Complex Claims Look For in Online Communities: Netnographic Analysis
J Med Internet Res 2022;24(4):e17180
URL: https://www.jmir.org/2022/4/e17180
doi:10.2196/17180
PMID:35389358

Original Paper

Fidelity to Program Specification of the National Health Service Digital Diabetes Prevention Program Behavior Change Technique Content and Underpinning Theory: Document Analysis

Rhiannon E Hawkes[1], BSc, MSc; Lisa M Miles[1], BSc, MMedSc, DrPH; David P French[1], BSc, MSc, PhD

Manchester Centre for Health Psychology, Division of Psychology and Mental Health, University of Manchester, Manchester, United Kingdom

Corresponding Author:
David P French, BSc, MSc, PhD
Manchester Centre for Health Psychology
Division of Psychology and Mental Health
University of Manchester
Coupland 1 Building
Oxford Road
Manchester, M13 9PL
United Kingdom
Phone: 44 0161 275 2605
Email: david.french@manchester.ac.uk

Abstract

Background: The National Health Service (NHS) Diabetes Prevention Program is a behavior change intervention for adults in England who are identified as being at high risk of developing type 2 diabetes. The face-to-face service was launched in 2016, followed by a digital service (NHS Digital Diabetes Prevention Program [NHS-DDPP]) in 2019. A total of 4 service providers were commissioned to deliver the NHS-DDPP and were required to deliver the digital service in line with a program specification detailing the key intervention content. The fidelity of the behavior change content in the digital service (ie, the extent to which the program is delivered as intended) is currently unknown. Digital interventions may allow higher fidelity as staff do not have to be trained to deliver *all* intervention content. Assessing fidelity of the intervention design is particularly important to establish the planned behavior change content in the NHS-DDPP and the extent to which this adheres to the program specification. This is the first known independent assessment of design fidelity in a large-scale digital behavior change intervention.

Objective: This study aims to assess the fidelity of the behavior change content in each of the 4 NHS-DDPP providers' intervention designs to the full program specification.

Methods: We conducted a document review of each provider's NHS-DDPP intervention design, along with interviews with program developers employed by the 4 digital providers (n=6). Providers' intervention design documents and interview transcripts were coded for behavior change techniques (BCTs; ie, the *active ingredients* of the intervention) using the Behavior Change Technique Taxonomy version 1 and underpinning theory using the Theory Coding Scheme framework. The BCTs identified in each digital provider's intervention design were compared with the 19 BCTs included in the program specification.

Results: Of the 19 BCTs specified in the program specification, the 4 providers planned to deliver 16 (84%), 17 (89%), 16 (84%), and 16 (84%) BCTs, respectively. An additional 41 unspecified BCTs were included in at least one of the 4 digital providers' intervention designs. By contrast, inconsistent use of the underpinning theory was apparent across providers, and none of the providers had produced a logic model to explain how their programs were expected to work. All providers linked some of their planned BCTs to theoretical constructs; however, justification for the inclusion of other BCTs was not described.

Conclusions: The fidelity of BCT content in the NHS-DDPP was higher than that previously documented for the face-to-face service. Thus, if service users engage with the NHS-DDPP, this should increase the effectiveness of the program. However, given that a clear theoretical underpinning supports the translation of BCTs in intervention designs to intervention delivery, the absence of a logic model describing the constructs to be targeted by specific BCTs is potentially problematic.

(J Med Internet Res 2022;24(4):e34253) doi:10.2196/34253

KEYWORDS

diabetes prevention; digital interventions; behavior change; fidelity

Introduction

Background

Type 2 diabetes mellitus (T2DM) is an international public health concern, with >4.9 million people having diabetes and 13.6 million people now at an increased risk of T2DM in the United Kingdom [1]. T2DM is preventable by lifestyle modifications in diet, physical activity, and weight loss, although family history, ethnicity, and age are risk factors. Treatment for T2DM and its associated complications currently costs the National Health Service (NHS) £10 billion (US $13 billion) each year; thus, its prevention is a public health priority [2].

In response to this, the NHS Diabetes Prevention Program (NHS-DPP) was launched in 2016, a behavior change intervention for adults in England who have been identified as having elevated blood glucose levels and are thus at a higher risk of developing T2DM. The main aims of the program are to improve diet, increase physical activity, and achieve weight loss, contributing to a reduction in T2DM risk [3]. Early evaluations suggest that the program is successful in preventing the progression to T2DM [4]. However, there is a weakness in reach and retention in the face-to-face program (eg, younger adults and those from more deprived communities) [5], which a digital program may overcome. Digital interventions have the potential to address logistical challenges such as scheduling, travel, work, and childcare [6], and they have the capacity to offer wide-reaching and tailored support at a lower cost, both of which are important for large-scale behavior change [7]. For example, technology can offer more choice and convenience to patients who may not be able to engage face to face with health professionals, and interventions can be designed to address the specific needs of disadvantaged groups by tailoring those interventions to the target group [8], especially when these groups are included in the design of digital interventions [9].

In 2017, NHS England launched a pilot program of the NHS Digital DPP (NHS-DDPP) [10]. Following this pilot, a digital pathway was introduced to the program in 2019 [11], with 4 independent service providers commissioned to deliver the digital service. NHS England produced a service specification detailing the key features that should be present in the NHS-DPP [11], based on the currently available evidence [12,13]. This specification indicated the key behavior change techniques (BCTs) that should be present in the intervention [11]. BCTs are defined as the *active ingredients* of an intervention that promotes behavioral change (eg, setting goals, monitoring behavior, and social support) [14]. The use of BCTs to self-regulate behaviors (eg, goal setting and self-monitoring) was particularly emphasized in the specification, as the wider literature suggests that these are key to behavioral change [13].

Intervention fidelity is used to describe whether an intervention is *delivered as intended* [15,16]. Without an assessment of fidelity, it cannot be ascertained whether intervention effectiveness, or a lack thereof, is because of the intervention content or other factors added or omitted in the intervention delivery. A recent evaluation of the face-to-face delivery of the NHS-DPP produced a thorough fidelity assessment of the behavior change content in the program, including the assessment of intervention design [17], staff training [18], intervention delivery [19-21], and service user receipt [22]. The results of the face-to-face evaluation revealed an underdelivery of BCTs to self-regulate behaviors (eg, problem solving and goal setting) [19,21]. This suggests a gap between what the evidence base indicated as the most effective in changing health behaviors and actual program delivery, which is largely because of failures to translate the BCTs specified in the evidence base into providers' program designs [19].

An assessment of the fidelity of the design of interventions is an important first stage in fidelity evaluation to establish what key intervention features are planned and whether these are in line with the evidence base from the outset. If a program is not designed with fidelity to the evidence base, it will have a downstream influence on its delivery and receipt. The findings from the assessment of fidelity of design for the face-to-face service indicated 19 BCTs in the NHS-DPP full program specification that should be included in the program; face-to-face providers described 74% of those BCTs in their intervention designs [17]. However, we are yet to establish the fidelity of the behavior change content in the digital offering of the NHS-DPP; that is, the extent to which the digital providers include the 19 specified BCTs in their intervention designs.

The underpinning theory of an intervention refers to the theories that inform the intervention, such as identifying theoretical constructs to be targeted [23]. Having identified these constructs allows for appropriate BCTs to be selected. Despite recommendations for clear descriptions of the underpinning theory of behavior change interventions [24], there is still a lack of literature that reports how interventions work (ie, the theoretical basis of the intervention and the rationale for selecting the BCTs designed for the intervention). To facilitate the description of different aspects of theory use, the Theory Coding Scheme (TCS) was developed [25], which is a 19-item framework providing a fine-grained assessment of the use of theory, including the extent to which all mentioned BCTs are linked to relevant constructs.

A concise way of visually presenting the theoretical basis for interventions that are informed by multiple theories could be via a logic model that uses simple diagrams to demonstrate the causal pathways between the intervention components and desired outcomes [26]. Without an explicit underpinning theory, it is not clear how the program is expected to produce the desired behavioral changes; thus, the rationale for the design of the intervention is not clear, and there is a lack of justification for why particular BCTs have been chosen for the intervention [25]. The evaluation of the face-to-face NHS-DPP found variation in the underpinning theory across face-to-face providers, with a lack of justification for the inclusion of BCTs in providers' intervention designs [27].

Digital interventions may allow higher fidelity as staff do not have to be trained to deliver *all* the intervention content [7]. To date, research has primarily focused on patient engagement with digital health interventions (eg, via methods such as system usage data, qualitative measures, and self-report questionnaires) [28] and adherence to digital interventions (ie, the extent to which patients keep using the digital intervention in the desired

way) [29]. Some studies have summarized the development process of digital interventions (eg, describing the theory and existing evidence to inform the development of a digital self-management intervention for people with T2DM) [30] and compared web-based intervention content against the intervention description [31]. However, there is a lack of literature examining the extent to which the planned intervention content of digital health interventions adheres to the evidence base for what works in changing health behaviors.

Objectives

Assessing the fidelity of nationally implemented digital behavior change interventions is rare, and to the best of our knowledge, none to date have assessed the fidelity of design to behavior change content. This analysis aims to provide the first known independent evaluation of fidelity of design of a nationally implemented digital intervention and thus builds on previous methods set out to measure fidelity [16] by applying them in a novel way to digital intervention designs using the NHS-DDPP as an exemplar. Therefore, the objectives of this study are to (1) describe the planned BCT content and theoretical foundations of the NHS-DDPP, (2) examine variations in planned BCT content and theoretical foundations between the digital providers, and (3) evaluate the fidelity of the planned BCT content in each provider's NHS-DDPP intervention design to the full program specification.

Methods

Design

We conducted a document analysis of the NHS-DDPP providers' intervention design documentation, along with qualitative interviews with program developers of each of the 4 digital providers. There were no single documents for any digital provider that offered a full and complete summary of their intervention designs; thus, interviews were used to supplement the documentation to allow researchers to gain in-depth information about the development of the NHS-DDPP and the implementation journey of the digital program. Digital providers' intervention designs were compared with the NHS-DPP program specification.

Document Review

Program Specification Documentation

The program specification documentation indicated the key intervention content that should be included in the NHS-DDPP. The documents that comprised the program specification are described in the following sections.

NHS-DPP Service Specification

This document [11] was specific to the commissioning of the NHS-DPP and was based on an evidence review of lifestyle interventions for the prevention of type 2 diabetes [12]. It specified what NHS England required to be included in the NHS-DPP (including face-to-face and digital offerings), drawing on recommendations from the National Institute for Health and Care Excellence (NICE) PH38 guidance [13].

The NICE PH38 Guideline, Type 2 Diabetes: Prevention in People at High Risk

This document [13] provided additional information regarding behavior change content to be included in diabetes prevention programs and was referred to in the NHS Service Specification [11].

The research team also consulted the recently published NICE NG183 guideline, *Behavior change: digital and mobile health interventions* [32], providing guidance on digital intervention development targeting health behaviors, including diet and physical activity [32]. However, there was no additional information regarding the specific behavior change content that was not already included in the NICE PH38 guidance [13]. Thus, the NG183 guideline [32] has not been included as a specification document in the current research. Multimedia Appendix 1 provides further justification.

Digital Providers' Intervention Design Documentation

The following documents from digital providers described providers' intervention designs obtained between June 2020 and February 2021:

- Providers' framework response bids describing digital providers' proposed service delivery submitted to NHS England during service procurement; 1 per digital provider (n=4)
- Additional documentation supplied by 75% (3/4) of digital providers, which detailed further information about the planned behavior change content designed in their digital programs (Table 1)
- Email correspondence with 25% (1/4) of digital providers containing further details of the BCTs and the underpinning theory of their digital programs; this was obtained when program developers were interviewed and sought further information from other colleagues who were not available to be interviewed but could provide further details to some of the interview questions via email (Table 1)

Table 1. Data obtained from each digital provider.

Digital provider	Number of interviews	Further information obtained from participant via email?	Further documentation obtained from provider?
A	2	Yes	Yes
B	1	No	Yes
C	2	No	No
D	1	No	Yes

Participants

For the purpose of this manuscript, the term *digital providers* indicates the 4 commercial companies commissioned to deliver the NHS-DDPP. The 4 digital providers were private service organizations, each of whom secured contracts to deliver the NHS-DDPP from 2019 to 2021. Of the 4 digital providers, 3 (75%; Oviva, Second Nature, and Liva) were in partnership with one of the face-to-face providers of the NHS-DPP to deliver the digital pathway of the program. One of the providers (WW) delivered both face-to-face and digital pathways of the NHS-DPP. In this report, providers are labeled A to D to preserve their anonymity.

To supplement the digital providers' intervention design documentation, interviews were conducted between September 2020 and December 2020 with program developers employed by each of the 4 digital providers. Program developers were involved in the design and development of the NHS-DDPP. Participants had job titles including Head of NHS Partnerships, Head of Clinical Services, and Head of Coaching. Participants were involved in either adapting their current digital programs to be in line with the NHS Service Specification [11] or were a key contact at the digital provider best placed to describe how the behavior change content was developed for the program.

Ethics Approval

The wider program of research of which this study is a part of was reviewed and approved by the North West Greater Manchester East NHS Research Ethics Committee (reference 17/NW/0426; August 1, 2017).

Procedures and Recruitment

The authors were in contact with the management staff of each of the 4 digital providers to obtain all relevant documentation detailing the providers' intervention designs. All documents were emailed to the research team.

The authors contacted the management staff of each of the 4 digital providers to help identify the appropriate staff members to be interviewed. The management staff contacted the relevant individuals directly, and those individuals proactively contacted the research team, where they expressed an interest in participating in an interview. We aimed to interview professionals from different backgrounds who had different roles in program development (eg, specialists with backgrounds in behavioral science, nutrition, or physical activity, and provider leads overseeing program development) to gather a range of views and provide a comprehensive understanding of the processes involved in the design and development of each NHS-DDPP intervention. Before scheduling the interviews, the participants were sent an information sheet containing details about the research. Participants were given the opportunity to ask any questions about the research, and fully recorded verbal consent was obtained from each participant before commencing the interview. Interviews were conducted via a video call platform—Zoom (Zoom Video Communications) videoconferencing—with one of the two researchers (REH and LMM) who had previous training in qualitative data collection.

Interviews with 2 professionals took place with digital providers A and C, and interviews with 1 professional took place with digital providers B and D (n=6 interviews in total). The interviews covered the following topics:

- Participants' professional backgrounds and roles in the development of the NHS-DDPP
- The theoretical underpinning of their programs, including targeted constructs and why specific theories or models were chosen
- Planned BCTs in the program, including behaviors targeted by these techniques and how these techniques were expected to work in producing behavioral changes
- The extent to which the programs were adapted from pre-existing digital programs for the NHS-DDPP
- The extent to which the digital offering of the NHS-DPP changed from the face-to-face offering of the program
- Strategies in place to support service users throughout the program (eg, at first contact and continued engagement)
- Content of the digital intervention and format of the different intervention features included in the program

The interviews were semistructured; the researchers initially asked open questions about the general topic, followed by a more detailed probing on specific issues. It was noted that participants were not expected to be able to answer all the interview questions, depending on their specific role and expertise in intervention development. Interviews were transcribed verbatim. Taken together, the documentation from each digital provider and interviews with program developers described the behavior change content that digital providers intended to deliver in each of their programs (providers' intervention designs).

Data Analysis

Coding Frameworks

A content analysis of BCTs and the underpinning theory contained within all documentation and interview transcripts was conducted using the following coding frameworks: the Behavior Change Technique Taxonomy version 1 (BCTTv1) [14] to document the BCTs described in providers' digital intervention designs and the TCS [25] to document the underpinning theory and theoretical constructs described in providers' digital intervention designs.

The BCTTv1 [14] lists 93 distinct techniques, each with its own labels and definitions, categorized into 16 groups. The TCS [25] is a 19-item framework that captures how theory is directly used in the intervention (eg, links between theoretical constructs and BCTs). Both frameworks have been widely used for reporting interventions; the BCTTv1 has demonstrated good interrater reliability (IRR), test-retest reliability, and good validity [33], and the authors of the TCS have reported satisfactory IRR for each item of the TCS [25].

The semistructured interviews were analyzed using deductive content analysis. Information relating to the behavior change content of the program, including BCTs and the underpinning theory, was extracted from all documentation and interview transcripts using these coding frameworks.

Coding Procedures

Providers' intervention design documentation and interview transcripts were coded separately using the BCTTv1 [14] and TCS [25] by one researcher (REH). BCTs were coded using an author-developed data extraction sheet. Researchers underwent training in the use of the BCTTv1 [34], and a set of coding rules was developed through team discussions. BCTs present in the program specification documentation [11,13] were coded for a previous fidelity assessment of the face-to-face offering of the NHS-DPP conducted by the same research team [17]. The study by Hawkes et al [17], as well as Multimedia Appendix 2, provides further details on the developed BCT coding rules. To code the underpinning theory, a data extraction sheet was developed using the TCS [25]. Multimedia Appendix 2 provides further details on how the data on the underpinning theory were extracted.

The framework response bids from the digital providers' intervention design documentation were double coded for BCTs and the underpinning theory by a second researcher (LMM). IRR was calculated using the Cohen κ coefficient [35] to determine consistency between coders. Identified coding discrepancies were discussed among REH, LMM, and DPF until consensus was reached. Cohen κ values were previously determined for the NHS Service Specification [11] and NICE PH38 guideline [13] (program specification documentation) by Hawkes et al [17]. Multimedia Appendix 3 provides all κ values.

Fidelity Analysis

The BCTs present in the NHS Service Specification and NICE guidance (program specification) were compared with those present in providers' intervention design documentation and interview transcripts (digital providers' intervention designs), and the results were tabulated. The theoretical principles detailed in digital providers' intervention designs were summarized for each provider.

Results

BCT Content

Program Specification Documentation

A previous analysis of the NHS-DPP program specification indicated 19 BCTs that should be included in the NHS-DPP intervention [17] (Table 2). The program specification emphasized the use of self-regulatory techniques, including goal setting, action planning, problem solving, and self-monitoring.

Table 2. Behavior change techniques specified in the full program specification compared with behavior change techniques specified in digital providers' intervention designs[a].

Behavior change techniques[b]	NICE[c] PH38	NHS[d] Service Specification	A	B	C	D
Credible source	✓[e]	✓	✓	✓	✓	✓
Goal setting (behavior)	✓	✓	✓	✓	✓	✓
Goal setting (outcome)	✓	✓	✓	✓	✓	✓
Graded tasks	✓	✓	✓	✓	✓	✓
Information about health consequences	✓	✓	✓	✓	✓	✓
Social support (unspecified)	✓	✓	✓	✓	✓	✓
Action planning	✓		✓	✓	✓	✓
Behavior substitution	✓		✓	✓	✓	✓
Problem solving	✓		✓	✓	✓	✓
Review outcome goals	✓		✓	✓	✓	✓
Self-monitoring of behavior	✓		✓	✓	✓	✓
Self-monitoring of outcomes of behavior	✓		✓	✓	✓	✓
Social support (emotional)[f]	✓		✓	✓	✓	✓
Feedback on behavior	✓		✓	✓	✓	
Social support (practical)[f]	✓			✓	✓	✓
Behavioral practice or rehearsal		✓	✓	✓	✓	✓
Monitoring of outcomes of behavior without feedback		✓		✓		
Pros and cons	✓		✓			✓
Pharmacological support		✓				
Commitment			✓	✓[g]	✓	✓
Demonstration of behavior			✓	✓	✓	✓
Feedback on outcomes of behavior			✓	✓	✓	✓
Habit formation			✓	✓	✓	✓
Instruction on how to perform the behavior			✓	✓	✓	✓
Reduce negative emotions			✓	✓	✓	✓
Restructuring the physical environment			✓[h]	✓	✓	✓[h]
Review behavior goals			✓	✓	✓	✓
Reward (outcome)			✓[i]	✓[i]	✓[i]	✓
Social reward			✓[i]	✓	✓[i]	✓
Framing and reframing			✓	✓[g]	✓	
Identification of self as a role model			✓	✓	✓	
Restructuring the social environment			✓[h]	✓		✓[h]
Social comparison			✓	✓		✓
Biofeedback			✓		✓	✓
Information about antecedents			✓		✓	✓
Prompts and cues			✓		✓	✓
Behavioral contract				✓	✓	✓
Focus on past success				✓	✓	✓
Self-talk				✓[g]	✓	✓

Behavior change techniques[b]	NICE[c] PH38	NHS[d] Service Specification	A	B	C	D
Increase positive emotions[j]			✓	✓		
Verbal persuasion about capability				✓[g]	✓	
Information about emotional consequences				✓		✓
Avoidance or reducing exposure to cues for the behavior				✓		
Behavioral experiments				✓[g]		
Material incentive (behavior)				✓		
Material reward (behavior)				✓		
Mental rehearsal of a successful performance				✓		
Monitoring of emotional consequences				✓		
Nonspecific incentive				✓[g]		
Remove aversive stimulus				✓		
Reward alternative behavior				✓		
Self-incentive				✓		
Self-reward				✓[g]		
Information about social and environmental consequences					✓	
Salience of behaviors[k]					✓	
Vicarious consequences					✓	
Adding objects to the environment						✓
Body changes						✓
Habit reversal						✓
Social incentive						✓

[a]The data in this table combine results from design documentation supplied by providers and interviews with program developers involved in the design of the digital diabetes prevention program.

[b]The first 19 behavior change techniques (from *Credible source* to *Pharmacological support*) are those 19 core behavior change techniques specified in the full program specification underpinning the NHS Diabetes Prevention Program.

[c]NICE: National Institute for Health and Care Excellence.

[d]NHS: National Health Service.

[e]Addition from coding of updated guidance when the document was updated in September 2017.

[f]*Social support (practical)* and *Social support (emotional)* were coded as 1 behavior change technique in the NICE guideline, as it stated that either of these forms of social support could be delivered.

[g]*Framing and reframing*, *Self-talk*, *Verbal persuasion about capability*, *Nonspecific incentive*, *Self-reward*, *Commitment*, and *Behavioral experiments* were present in optional extra sessions only for this provider.

[h]*Restructuring the physical environment* and *Restructuring the social environment* were coded as 1 behavior change technique in the framework response, as it stated only to restructure the environment without specifying whether this was the physical or social environment.

[i]*Social reward* and *Reward (outcome)* were coded as 1 behavior change technique in the framework response, as it did not state whether the reward was for the behavior or outcome.

[j]Increased positive emotions are not listed in the Behavior Change Technique Taxonomy version 1 but were noted by the authors for inclusion in the next version of the taxonomy and used in Hawkes et al [17].

[k]Salience of behaviors was not listed in the Behavior Change Technique Taxonomy version 1 but has been identified as a new behavior change technique by the authors of this paper and used in Hawkes et al [17].

Digital Providers' Intervention Designs

Cohen κ values ranged from 0.66 to 0.93 for the double coding of BCTs in digital providers' intervention design documentation, demonstrating moderate to almost perfect agreement between coders [36] before resolving discrepancies (see Multimedia Appendix 3 for all κ values). Table 2 shows the BCTs identified in each of the 4 digital providers' intervention designs based on all supplied documentation and interviews with the program developers of each digital provider.

Combining the content analysis of all documentation and interview data, providers A, B, C, and D planned to deliver 34, 48, 38, and 39 unique BCTs, respectively, in their digital programs. For provider B, 15% (7/48) of these BCTs were present only in optional extra sessions. All 4 providers planned

to deliver 23 common BCTs, including action planning, goal setting (for behaviors and outcomes), graded tasks, information about health consequences, problem solving, reviewing goals (for behaviors and outcomes), self-monitoring (for behaviors and outcomes), and social support (emotional and unspecified). Multimedia Appendix 4 provides BCT definitions according to the BCTTv1 [14].

Fidelity of BCT Content

There were 19 BCTs specified in the program specification documentation [11,13]. Of these 19 BCTs, the 4 providers included 16 (84%), 17 (89%), 16 (84%), and 16 (84%) BCTs in their intervention designs, respectively. Thus, overall, the mean proportion of BCTs in digital providers' intervention designs to the program specification was 85% (SD 2.5%). Of the 19 BCTs required by the full program specification, 14 (74%) were included in the intervention designs of all 4 digital providers (Table 2). These included self-regulatory techniques such as goal setting, action planning, problem solving, and self-monitoring.

A total of 6 BCTs specified in the program specification documentation were missing from at least one digital provider's intervention designs, indicating a lack of fidelity to the program specification by BCT omission. One of the BCTs, pharmacological support (eg, advising on medication for smoking cessation), was not provided by any of the 4 digital providers. There were 41 additional BCTs not indicated in the program specification, which were included in at least one of the 4 digital providers' intervention designs.

Underpinning Theory

Across the 4 digital providers, the mean κ value for coding using the TCS was 0.90, and the mean κ values for theories and constructs mentioned were 1.00 and 0.97, respectively. Cohen κ values demonstrated strong to an almost perfect agreement [36] between coders using TCS before resolving discrepancies (Multimedia Appendix 3 provides all κ values).

The extent to which theory was used in each of the 4 providers' intervention designs is summarized in Table 3. All digital providers mentioned theories on which their programs were based; however, none had produced an explicit logic model of how their program was expected to produce changes in behavior. Providers A, B, and D stated that their digital programs were informed by the Capability, Opportunity, Motivation–Behavior framework [37], with providers B and D describing a particular emphasis on this model. Providers A and C cited multiple theories to inform their intervention designs, including Social Cognitive Theory [38], the Transtheoretical Model [39], and Self-Determination Theory [40] (Table 4).

There was wide variation in the constructs that providers were targeting in their digital programs. All 4 digital providers mentioned the constructs of goals, motivations, and capabilities in their intervention designs. The following constructs were also mentioned in the digital designs of 75% (3/4) of providers: self-efficacy, opportunity, and social support (Table 4; constructs have been grouped based on the Theoretical Domains Framework [52]).

None of the 4 digital providers linked all of their planned BCTs (listed in Table 2) to the proposed theoretical constructs, although provider B did link all constructs and intervention functions of the Capability, Opportunity, Motivation–Behavior model (which their intervention was solely based on) to some planned BCTs, which was detailed in their intervention design documents (see item 10 in Table 3). All providers linked at least some of their planned BCTs to theoretical constructs (see item 8a in Table 3). For example, provider A described modeling, goal setting, and problem solving to drive self-efficacy, and provider C described the techniques of verbal persuasion, reviewing goals, and social support to target self-efficacy.

Table 3. Use of theory in digital providers' intervention designs[a].

Use of theory (item[b])	A	B	C	D
Theory mentioned (1a)	✓	✓	✓	✓
Construct mentioned (1b)[c]	✓	✓	✓	✓
Target construct mentioned as a predictor of behavior (2)	✓	✓	✓	✓
Intervention based on a single theory (3)		✓		✓
Theory or predictors used to select recipients for the intervention (4)				
Theory or predictors used to select or develop intervention techniques (5)	✓	✓	✓	✓
Theory or predictors used to tailor intervention techniques to recipients (6)	✓	✓	✓	
All intervention techniques are explicitly linked to at least one theory-relevant construct or predictor (7a)				
All intervention techniques are explicitly linked to an overall *theory* or *model* but not a specific construct (7b)[c]				
At least one but not all of the intervention techniques are explicitly linked to at least one theory-relevant construct or predictor (8a)	✓	✓	✓	✓
At least one but not all of the intervention techniques are explicitly linked to an overall *theory* or *model* but not a specific construct (8b)[c]				
The group of techniques is linked to a group of constructs or predictors (9a)	✓		✓	
The group of techniques is linked to an overall *theory* or *model* but not a specific construct (9b)[c]				
All theory-relevant constructs or predictors are explicitly linked to at least one intervention technique (10)		✓		
At least one but not all of the theory-relevant constructs or predictors are explicitly linked to at least one intervention technique (11)	✓[d]		✓	✓
Theory-relevant constructs or predictors are measured (12)				
Quality of measures (13)				

[a]The data in this table combine the results from the design documentation supplied by providers and interviews with program developers involved in the design of the digital diabetes prevention program.

[b]Denotes items of the Theory Coding Scheme [25]; items 14 to 19 of the Theory Coding Scheme relate to postintervention rather than protocol assessment and are therefore not included in this analysis.

[c]Additional items that the authors added to the Theory Coding Scheme for this analysis (Multimedia Appendix 2).

[d]Provider linked all constructs and intervention functions of the Capability Opportunity Motivation–Behavior model [37] to behavior change techniques but did not link all other listed theoretical constructs to behavior change techniques.

Table 4. Theories, approaches, and constructs mentioned in each digital providers' diabetes prevention program designs[a].

Models, approaches, constructs, and concepts	A	B	C	D
Model of behavior (change) mentioned				
Behavior change wheel [41]			✓	✓
Biopsychosocial model [42]		✓		
Cognitive behavioral theory [43]	✓			
COM-B[b] model [37]	✓	✓		✓
Functional Learning Theory [44]	✓			
GROW[c] model [45]			✓	
Health Action Process Approach [46]	✓			
Self-Determination Theory [40]			✓	
Social Cognitive Theory [39]	✓		✓	
Social Comparison Theory [47]	✓			
Transtheoretical model [39]	✓		✓	
Approaches to behavior change mentioned				
Acceptance and commitment therapy [48]	✓		✓	
Cognitive behavioral therapy [49]	✓	✓	✓	
Motivational interviewing [50]	✓	✓	✓	
Positive psychology [51]			✓	
Constructs[d] mentioned				
Autonomy			✓	
Behavior regulation (internal), self-control, and self-management	✓			✓
Beliefs about consequences, benefits and risks, personal beliefs, and risk perceptions	✓		✓	✓
Capability, cognitive and interpersonal skills, competence, and physical skills	✓	✓	✓	✓
Goals	✓	✓	✓	✓
Intentions				✓
Knowledge	✓			✓
Motivation	✓	✓	✓	✓
Intrinsic and extrinsic motivation			✓	
Opportunity	✓	✓		✓
Optimism				✓
Reinforcement	✓			✓
Self-efficacy, beliefs about capabilities, and observational learning	✓		✓	✓
Social context, environmental context, and environmental influences			✓	✓
Social support and relatedness	✓		✓	✓
Behavior change concepts mentioned				
Accountability				✓
Approach-avoidance language			✓	
Behavioral influences			✓	
Committed action			✓	
Connectedness			✓	
Lapses and relapses	✓		✓	

[a]The data in this table combine the results from the design documentation supplied by providers and interviews with program developers involved in

the design of the digital diabetes prevention program.

[b]COM-B: Capability, Opportunity, Motivation–Behavior.

[c]GROW: Goal, Reality, Options, and Way forward.

[d]Constructs were grouped together based on the Theoretical Domains Framework [52].

Discussion

Principal Findings

The digital providers of the NHS-DDPP demonstrated good fidelity to the BCT content of the NHS-DDPP design. In particular, self-regulatory BCTs, which may be most effective in changing health behaviors [13], were included in the designs of all 4 digital providers. However, providers appeared to have inconsistent use of theory to inform their intervention designs, and none had developed a logic model or provided an explicit description of how their digital programs were expected to achieve the desired health outcomes. Without a clear underpinning theory and identification of the mechanisms underlying particular BCTs, the justification for the inclusion of specific BCT content in digital providers' programs is not clear.

Comparison With Prior Work

Overall, digital providers planned to deliver an average of 85% of the BCT content specified in the NHS-DPP program specification. Thus, the fidelity of design in the digital offering of the NHS-DPP is, in principle, higher than that previously reported for the face-to-face offering, where the fidelity of design was 74% [17]. The planned self-regulatory BCT content is particularly encouraging, given that these BCTs are backed by evidence to suggest that such techniques are most effective in changing health behaviors such as diet and physical activity [13]. However, the inconsistent use of theory in digital programs is similar to the evaluation of face-to-face NHS-DPP, which also highlighted a lack of mapping of BCTs to theoretical constructs and a lack of a logic model produced by any of the providers [27]. Both findings could be explained by the fact that providers' digital programs were adapted from pre-existing digital interventions already implemented before the commissioning of the NHS-DDPP service. This may be a particular problem for digital interventions.

Strengths and Limitations

This study provides a detailed analysis of the planned behavior change content of each of the 4 digital providers commissioned to deliver the NHS-DDPP using standardized coding frameworks [14,25], which demonstrated high κ agreement between coders. All framework response documentation obtained from each digital provider was submitted to the commissioners during service procurement; thus, they should represent an accurate description of digital providers' intervention designs. In addition, the research team requested any other relevant documentation from each digital provider that contained information about the planned behavior change content of the NHS-DDPP. Providers were given multiple opportunities to share all relevant documentation with the research team to support the current analysis.

However, one of the providers was reluctant to share all documentation detailing the theoretical underpinning because of the commercial sensitivity of these documents. Researchers attempted to obtain as much information as possible about the theoretical foundations of providers' programs via interviews, especially when documentation could not be shared with the research team. The interviews were able to supplement the document review and were used to fill any gaps in the information that was not present in the documentation. Despite some reluctance to share documentation with the research team from one of the digital providers, analyses from this provider were similar to those of the other 3 providers, who shared all documentation. Furthermore, all digital providers confirmed that they had not developed a logic model to explicitly map how their digital programs would achieve the desired program outcomes, nor was there an alternative presentation of how their interventions were expected to work (eg, via free text).

It should also be noted that participants who were interviewed for this study were not always necessarily directly involved in the development of the BCT content and the theoretical underpinning of the programs. Some people involved in the early stages of design and development had since moved on to other roles. Nonetheless, the research team tried to identify ≥1 relevant individual from each digital provider and aimed to interview professionals from different backgrounds to gather a range of views and provide a comprehensive understanding of the processes involved in the design and development of each NHS-DDPP intervention.

Future Directions

Digital interventions have the potential to be delivered with higher fidelity than face-to-face interventions [7] as they do not rely solely on facilitator-centered delivery; instead, the BCT content can be standardized in its delivery via the functions of an app or website. Therefore, the current fidelity of design results is encouraging, as drift in delivery fidelity is less likely, thus increasing the chances of the program being effective if the BCT content is proven to be engaged with effectively.

Despite the encouraging results with regard to the planned BCTs in digital providers' program designs, there is inconsistent use of theory from providers to justify why specific BCTs have been included and how they will produce the desired behavior changes (ie, the theoretical constructs they are targeting). Given that a clear underpinning theory supports the translation of BCTs in intervention designs to intervention delivery [24-26], the absence of a clear logic model or an explicit explanation of underpinning theory describing the constructs to be targeted by specific BCTs is potentially problematic.

One of the ways in which providers could usefully describe the underpinning theory is via a logic model, which provides a concise and clear way of visually presenting interventions informed by multiple theories. However, there is no clear guidance on whether it should be the role of the providers to produce a logic model or whether it would be more appropriate for commissioners of large-scale programs to produce an initial

logic model to guide providers from the outset [27]. However, the requirement of a logic model during the commissioning process of large-scale programs could ensure a clear rationale for the BCTs included in intervention designs, which may result in fewer issues in delivery and receipt fidelity in program implementation [27]. It should be noted that the inclusion of a logic model is a recommendation that has emerged since the evaluation of the NHS-DPP commenced and following the previous round of service procurement. Commissioners of the program are now increasingly focusing on the important role of logic models in the upcoming round of service procurement for the NHS-DPP.

In addition, a further 68% of the planned BCT content in the digital service was not specified in the program specification, a finding similar to the face-to-face offering of the NHS-DPP [17]. It is possible that interventions containing more techniques to help people change their diet and physical activity behaviors may be more effective [53,54]. However, without a logic model or explicit description of the underpinning theory justifying the inclusion of additional BCT content, programs cannot be tested and evaluated effectively to establish the techniques that work in digital behavior change interventions and why.

This research provides an example of a design fidelity assessment of a large-scale digital intervention. Using the present findings as a baseline, future research will assess fidelity of the delivery of the digital service and service user engagement with the program. The current analysis found that some digital providers still plan to rely on health coaches to deliver aspects of the program to service users (eg, support via telephone); thus, the nature of this delivery will depend on the skills of the health coach in a similar way as that delivered by facilitators of face-to-face sessions. This may help to increase program engagement [55] but may also pose a risk to delivery fidelity because of variations in the background or training of health coaches. This study will be critical to assess whether the plans digital providers state they will put in place (with high fidelity) translate into program content that is actually delivered and understood as intended. Ongoing qualitative work examines how the behavior change content of the program is understood by service users, an underresearched aspect of fidelity [56]. Further analysis of the interview data with program developers from this study is also planned, with the aim of qualitatively capturing the nuances of intervention development and how each of the digital providers' programs has evolved over time.

Conclusions

When designing their programs, the digital providers of the NHS-DDPP demonstrated good fidelity to the BCT content of the program specification. Thus, if the NHS-DDPP is delivered with high fidelity and engaged with by service users, it should increase the effectiveness of the program in achieving a reduction in the incidence of T2DM. However, the lack of theory described in digital providers' intervention designs is potentially problematic if providers are not explicit in how their programs will work to achieve behavior change, given that a clear underpinning theory supports the translation of BCT content in intervention design to intervention delivery.

Acknowledgments

The authors would like to thank the NHS-DPP team and all digital provider program leads for providing all the relevant documentation required for this manuscript. The authors would also like to thank the participants who took part in the interview and the provider program leads and management staff for helping to organize the interviews. The authors would also like to thank the following researchers in the DIPLOMA (Diabetes Prevention–Long Term Multimethod Assessment) team who provided valuable feedback during manuscript preparation: Peter Bower, Sarah Cotterill, and Elizabeth Murray. This work is independent research funded by the National Institute for Health Research (Health Services and Delivery Research, 16/48/07—Evaluating the National Health Service Diabetes Prevention Program: the DIPLOMA research program). The views and opinions expressed in this manuscript are those of the authors and do not necessarily reflect those of the National Institute for Health Research or the Department of Health and Social Care.

Data Availability

The intervention design documentation from digital providers and audio recordings of interviews analyzed in this study are not publicly available because of confidentiality agreements with the provider organizations, as some information is commercially sensitive. Some data sets are available from the corresponding author on reasonable request, although the authors will require explicit permission from the relevant provider organizations.

Authors' Contributions

DPF designed the research and secured funding as part of the wider DIPLOMA (Diabetes Prevention–Long Term Multimethod Assessment) project. DPF supervised the research and helped prepare the manuscript. REH conducted the research interviews, analyzed the data, and prepared the manuscript. LMM conducted the research interviews, analyzed a proportion of the data, and helped draft the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
National Institute of Health and Care Excellence NG183 guidance (2020)—Behavior Change: Digital interventions.
[DOCX File , 13 KB - jmir_v24i4e34253_app1.docx]

Multimedia Appendix 2
Coding procedures.
[DOCX File , 14 KB - jmir_v24i4e34253_app2.docx]

Multimedia Appendix 3
Cohen κ values.
[DOCX File , 13 KB - jmir_v24i4e34253_app3.docx]

Multimedia Appendix 4
Behavior change technique definitions.
[DOCX File , 16 KB - jmir_v24i4e34253_app4.docx]

References

1. Diabetes statistics: diabetes prevalence 2019. Diabetes UK. 2020. URL: https://www.diabetes.org.uk/professionals/position-statements-reports/statistics/diabetes-prevalence-2019 [accessed 2022-04-04]
2. Hex N, Bartlett C, Wright D, Taylor M, Varley D. Estimating the current and future costs of type 1 and type 2 diabetes in the UK, including direct health costs and indirect societal and productivity costs. Diabet Med 2012 Jul;29(7):855-862 [FREE Full text] [doi: 10.1111/j.1464-5491.2012.03698.x] [Medline: 22537247]
3. NHS England Diabetes Prevention Programme (NHS DPP). NHS.UK. URL: https://www.england.nhs.uk/diabetes/diabetes-prevention/ [accessed 2022-04-04]
4. Valabhji J, Barron E, Bradley D, Bakhai C, Fagg J, O'Neill S, et al. Early outcomes from the English National Health Service Diabetes Prevention Programme. Diabetes Care 2020 Jan;43(1):152-160 [FREE Full text] [doi: 10.2337/dc19-1425] [Medline: 31719054]
5. Howarth E, Bower PJ, Kontopantelis E, Soiland-Reyes C, Meacock R, Whittaker W, et al. 'Going the distance': an independent cohort study of engagement and dropout among the first 100 000 referrals into a large-scale diabetes prevention program. BMJ Open Diabetes Res Care 2020 Dec;8(2):e001835 [FREE Full text] [doi: 10.1136/bmjdrc-2020-001835] [Medline: 33303493]
6. Arigo D, Jake-Schoffman DE, Wolin K, Beckjord E, Hekler EB, Pagoto SL. The history and future of digital health in the field of behavioral medicine. J Behav Med 2019 Feb;42(1):67-83 [FREE Full text] [doi: 10.1007/s10865-018-9966-z] [Medline: 30825090]
7. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
8. Tinder Foundation. URL: https://nhs.goodthingsfoundation.org/wp-content/uploads/2016/07/Improving_Digital_Health_Skills_Report_2016.pdf [accessed 2022-04-04]
9. Digital technology and health inequalities: a scoping review (Wales). Health Inequalities Portal. URL: https://phw.nhs.wales/publications/publications1/digital-technology-and-health-inequalities-a-scoping-review/ [accessed 2022-04-12]
10. Murray E, Daff K, Lavida A, Henley W, Irwin J, Valabhji J. Evaluation of the digital diabetes prevention programme pilot: uncontrolled mixed-methods study protocol. BMJ Open 2019 May 22;9(5):e025903 [FREE Full text] [doi: 10.1136/bmjopen-2018-025903] [Medline: 31122973]
11. Provision of behavioural interventions for people with nondiabetic hyperglycaemia. NHS England. URL: https://www.england.nhs.uk/wp-content/uploads/2016/08/nhs-dpp-service-specification-aug-2019.pdf [accessed 2022-04-04]
12. A systematic review and metaanalysis assessing the effectiveness of pragmatic lifestyle interventions for the prevention of type 2 diabetes mellitus in routine practice. Public Health England. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/456147/PHE_Evidence_Review_of_diabetes_prevention_programmes-_FINAL.pdf [accessed 2022-04-12]
13. Type 2 diabetes: prevention in people at high risk. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/guidance/ph38 [accessed 2022-04-12]
14. 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 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
15. Bellg AJ, Borrelli B, Resnick B, Hecht J, Minicucci DS, Ory M, Treatment Fidelity Workgroup of the NIH Behavior Change Consortium. Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the

NIH Behavior Change Consortium. Health Psychol 2004 Sep;23(5):443-451. [doi: 10.1037/0278-6133.23.5.443] [Medline: 15367063]
16. Borrelli B. The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials. J Public Health Dent 2011;71(s1):S52-S63 [FREE Full text] [doi: 10.1111/j.1752-7325.2011.00233.x] [Medline: 21499543]
17. Hawkes RE, Cameron E, Bower P, French DP. Does the design of the NHS Diabetes Prevention Programme intervention have fidelity to the programme specification? A document analysis. Diabet Med 2020 Aug;37(8):1357-1366 [FREE Full text] [doi: 10.1111/dme.14201] [Medline: 31808578]
18. Hawkes RE, Cameron E, Miles LM, French DP. The fidelity of training in behaviour change techniques to intervention design in a national diabetes prevention programme. Int J Behav Med 2021 Dec;28(6):671-682 [FREE Full text] [doi: 10.1007/s12529-021-09961-5] [Medline: 33559009]
19. French DP, Hawkes RE, Bower P, Cameron E. Is the NHS Diabetes Prevention Programme Intervention delivered as planned? An observational study of fidelity of intervention delivery. Ann Behav Med 2021 Oct 27;55(11):1104-1115 [FREE Full text] [doi: 10.1093/abm/kaaa108] [Medline: 33580647]
20. Hawkes RE, Cameron E, Cotterill S, Bower P, French DP. The NHS Diabetes Prevention Programme: an observational study of service delivery and patient experience. BMC Health Serv Res 2020 Nov 27;20(1):1098 [FREE Full text] [doi: 10.1186/s12913-020-05951-7] [Medline: 33246460]
21. Hawkes RE, Warren L, Cameron E, French DP. An evaluation of goal setting in the NHS England diabetes prevention programme. Psychol Health 2022 Feb;37(2):131-150. [doi: 10.1080/08870446.2021.1872790] [Medline: 33517780]
22. Miles LM, Hawkes RE, French DP. How is the Behavior Change Technique Content of the NHS Diabetes Prevention Program understood by participants? A qualitative study of Fidelity, with a focus on receipt. Ann Behav Med 2021 Nov 11:kaab093 (forthcoming). [doi: 10.1093/abm/kaab093] [Medline: 34788358]
23. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
24. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:n2061 [FREE Full text] [doi: 10.1136/bmj.n2061] [Medline: 34593508]
25. Michie S, Prestwich A. Are interventions theory-based? Development of a theory coding scheme. Health Psychol 2010 Jan;29(1):1-8. [doi: 10.1037/a0016939] [Medline: 20063930]
26. O'Cathain A, Croot L, Duncan E, Rousseau N, Sworn K, Turner KM, et al. Guidance on how to develop complex interventions to improve health and healthcare. BMJ Open 2019 Aug 15;9(8):e029954 [FREE Full text] [doi: 10.1136/bmjopen-2019-029954] [Medline: 31420394]
27. Hawkes RE, Miles LM, French DP. The theoretical basis of a nationally implemented type 2 diabetes prevention programme: how is the programme expected to produce changes in behaviour? Int J Behav Nutr Phys Act 2021 May 13;18(1):64 [FREE Full text] [doi: 10.1186/s12966-021-01134-7] [Medline: 33985524]
28. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
29. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
30. Dack C, Ross J, Stevenson F, Pal K, Gubert E, Michie S, et al. A digital self-management intervention for adults with type 2 diabetes: combining theory, data and participatory design to develop HeLP-Diabetes. Internet Interv 2019 Sep;17:100241 [FREE Full text] [doi: 10.1016/j.invent.2019.100241] [Medline: 31372349]
31. Lambert JD, Elliott LR, Taylor AH, Farrand P, Haase AM, Greaves CJ. A novel method for assessing design fidelity in web-based behavioral interventions. Health Psychol 2021 Mar;40(3):217-225. [doi: 10.1037/hea0001046] [Medline: 33630643]
32. Behaviour change: digital and mobile health interventions. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/guidance/ng183 [accessed 2022-04-12]
33. Abraham C, Wood CE, Johnston M, Francis J, Hardeman W, Richardson M, et al. Reliability of identification of behavior change techniques in intervention descriptions. Ann Behav Med 2015 Dec;49(6):885-900. [doi: 10.1007/s12160-015-9727-y] [Medline: 26290002]
34. Online training. BCT Taxonomy V1. URL: https://www.bct-taxonomy.com/ [accessed 2022-04-12]
35. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
36. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
37. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Interventions. London: Silverback Publishing; 2014.

38. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. London, UK: Pearson; 1985.
39. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
40. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
41. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
42. Engel GL. The need for a new medical model: a challenge for biomedicine. Science 1977 Apr 08;196(4286):129-136. [doi: 10.1126/science.847460] [Medline: 847460]
43. Beck A. Cognitive Therapy and the Emotional Disorders. New York, United States: Penguin Publishing Group; 1976.
44. Reading for Working A Functional Literacy Anthology. Alexandria, Virginia: Human Resources Research Organization; 1975.
45. Excellence in Coaching: The Industry Guide. London, UK: Kogan Page; 2016.
46. Schwarzer R. Modeling health behavior change: how to predict and modify the adoption and maintenance of health behaviors. Applied Psychol 2008 Jan 30;57(1):1-29 [FREE Full text] [doi: 10.1111/j.1464-0597.2007.00325.x]
47. Festinger L. A theory of social comparison processes. Human Relations 1954 May 01;7(2):117-140 [FREE Full text] [doi: 10.1177/001872675400700202]
48. Acceptance and Commitment Therapy. Washington, D.C., United States: American Psychological Association; 1999.
49. Cognitive Therapy: Basics and Beyond. New York, USA: Guildford Press; 1995.
50. Rollnick S, Miller WR. What is motivational interviewing? Behav Cognit Psychother 1995 Oct;23(4):325-334 [FREE Full text] [doi: 10.1017/s135246580001643x]
51. Seligman ME, Csikszentmihalyi M. Positive psychology: an introduction. Am Psychol 2000 Jan;55(1):5-14. [doi: 10.1037//0003-066x.55.1.5] [Medline: 11392865]
52. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implement Sci 2012 Apr 24;7:37 [FREE Full text] [doi: 10.1186/1748-5908-7-37] [Medline: 22530986]
53. Samdal GB, Eide GE, Barth T, Williams G, Meland E. Effective behaviour change techniques for physical activity and healthy eating in overweight and obese adults; systematic review and meta-regression analyses. Int J Behav Nutr Phys Act 2017 Mar 28;14(1):42 [FREE Full text] [doi: 10.1186/s12966-017-0494-y] [Medline: 28351367]
54. Van Rhoon L, Byrne M, Morrissey E, Murphy J, McSharry J. A systematic review of the behaviour change techniques and digital features in technology-driven type 2 diabetes prevention interventions. Digit Health 2020;6:2055207620914427 [FREE Full text] [doi: 10.1177/2055207620914427] [Medline: 32269830]
55. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
56. Toomey E, Hardeman W, Hankonen N, Byrne M, McSharry J, Matvienko-Sikar K, et al. Focusing on fidelity: narrative review and recommendations for improving intervention fidelity within trials of health behaviour change interventions. Health Psychol Behav Med 2020 Mar 12;8(1):132-151 [FREE Full text] [doi: 10.1080/21642850.2020.1738935] [Medline: 34040865]

Abbreviations

BCT: behavior change technique
BCTTv1: Behavior Change Technique Taxonomy version 1
DIPLOMA: Diabetes Prevention–Long Term Multimethod Assessment
IRR: interrater reliability
NHS: National Health Service
NHS-DDPP: National Health Service Digital Diabetes Prevention Program
NHS-DPP: National Health Service Diabetes Prevention Program
NICE: National Institute for Health and Care Excellence
T2DM: type 2 diabetes mellitus
TCS: Theory Coding Scheme

Edited by T Leung; submitted 13.10.21; peer-reviewed by I Wilson, E Neter; comments to author 21.11.21; revised version received 03.12.21; accepted 13.02.22; published 27.04.22.

Please cite as:
Hawkes RE, Miles LM, French DP
Fidelity to Program Specification of the National Health Service Digital Diabetes Prevention Program Behavior Change Technique Content and Underpinning Theory: Document Analysis
J Med Internet Res 2022;24(4):e34253
URL: https://www.jmir.org/2022/4/e34253
doi:10.2196/34253
PMID:35476035

Original Paper

Understanding the Research Landscape of Deep Learning in Biomedical Science: Scientometric Analysis

Seojin Nam[1], MA; Donghun Kim[1], MA; Woojin Jung[1], MA; Yongjun Zhu[2], PhD

[1]Department of Library and Information Science, Sungkyunkwan University, Seoul, Republic of Korea

[2]Department of Library and Information Science, Yonsei University, Seoul, Republic of Korea

Corresponding Author:
Yongjun Zhu, PhD
Department of Library and Information Science
Yonsei University
50 Yonsei-ro
Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 2 2123 2409
Email: zhu@yonsei.ac.kr

Abstract

Background: Advances in biomedical research using deep learning techniques have generated a large volume of related literature. However, there is a lack of scientometric studies that provide a bird's-eye view of them. This absence has led to a partial and fragmented understanding of the field and its progress.

Objective: This study aimed to gain a quantitative and qualitative understanding of the scientific domain by analyzing diverse bibliographic entities that represent the research landscape from multiple perspectives and levels of granularity.

Methods: We searched and retrieved 978 deep learning studies in biomedicine from the PubMed database. A scientometric analysis was performed by analyzing the metadata, content of influential works, and cited references.

Results: In the process, we identified the current leading fields, major research topics and techniques, knowledge diffusion, and research collaboration. There was a predominant focus on applying deep learning, especially convolutional neural networks, to radiology and medical imaging, whereas a few studies focused on protein or genome analysis. Radiology and medical imaging also appeared to be the most significant knowledge sources and an important field in knowledge diffusion, followed by computer science and electrical engineering. A coauthorship analysis revealed various collaborations among engineering-oriented and biomedicine-oriented clusters of disciplines.

Conclusions: This study investigated the landscape of deep learning research in biomedicine and confirmed its interdisciplinary nature. Although it has been successful, we believe that there is a need for diverse applications in certain areas to further boost the contributions of deep learning in addressing biomedical research problems. We expect the results of this study to help researchers and communities better align their present and future work.

(J Med Internet Res 2022;24(4):e28114) doi:10.2196/28114

KEYWORDS

deep learning; scientometric analysis; research publications; research landscape; research collaboration; knowledge diffusion

Introduction

Deep learning is a class of machine learning techniques based on neural networks with multiple processing layers that learn representations of data [1,2]. Stemming from shallow neural networks, many deep learning architectures, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have been developed for various purposes [3]. The exponentially growing amount of data in many fields and recent advances in graphics processing units have further expedited research progress in the field. Deep learning has been actively applied to tasks, such as natural language processing (NLP), speech recognition, and computer vision, in various domains [1] and has shown promising results in diverse areas of biomedicine, including radiology [4], neurology [2], cardiology [5], cancer detection and diagnosis [6,7],

radiotherapy [8], and genomics and structural biology [9-11]. Medical image analysis is a field that has actively used deep learning. For example, successful applications have been made in diagnosis [12], lesion classification or detection [13,14], organ and other substructure localization or segmentation [15,16], and image registration [17,18]. In addition, deep learning has also made an impact on predicting protein structures [19,20] and genomic sequencing [21-23] for biomarker development and drug design.

Despite the increasing number of published biomedical studies on deep learning techniques and applications, there has been a lack of scientometric studies that both qualitatively and quantitatively explore, analyze, and summarize the relevant studies to provide a bird's-eye view of them. Previous studies have mostly provided qualitative reviews [2,9,10], and the few available bibliometric analyses were limited in their scope in that the researchers focused on a subarea such as public health [24] or a particular journal [25]. The absence of a coherent lens through which we can examine the field from multiple perspectives and levels of granularity leads to a partial and fragmented understanding of the field and its progress. To fill this gap, the aim of this study is to perform a scientometric analysis of metadata, content, and citations to investigate current leading fields, research topics, and techniques, as well as research collaboration and knowledge diffusion in deep learning research in biomedicine. Specifically, we intend to examine (1) biomedical journals that had frequently published deep learning studies and their coverage of research areas, (2) diseases and other biomedical entities that have been frequently studied with deep learning and their relationships, (3) major deep learning architectures in biomedicine and their specific applications, (4) research collaborations among disciplines and organizations, and (5) knowledge diffusion among different areas of study.

Methods

Data

Data were collected from PubMed, a citation and abstract database that includes biomedical literature from MEDLINE and other life science journals indexed with Medical Subject Heading (MeSH) terms [26]. MeSH is a hierarchically structured biomedical terminology with descriptors organized into 16 categories, with subcategories [27]. In this study, *deep learning [MeSH Major Topic]* was used as the query to search and download deep learning studies from PubMed. Limiting a MeSH term as a major topic increases the precision of retrieval so that only studies that are highly relevant to the topic are found [28]. As of January 1, 2020, a total of 978 PubMed records with publication years ranging from 2016 to 2020 have been retrieved using the National Center for Biotechnology Information Entrez application programming interface. Entrez is a data retrieval system that can be programmatically accessed through its Biopython module to search and export records from the National Center for Biotechnology Information's databases, including PubMed [26,29]. The metadata of the collected bibliographic records included the PubMed identifier or PubMed ID, publication year, journal title and its electronic ISSN, MeSH descriptor terms, and author affiliations. We also downloaded the citation counts and references of each bibliographic record and considered data sources other than PubMed as well. We collected citation counts of the downloaded bibliographic records from Google Scholar (last updated on February 8, 2020) and the subject categories of their publishing journals from the Web of Science (WoS) Core Collection database using the electronic ISSN.

Detailed Methods

Metadata Analysis

Journals

Journals are an important unit of analysis in scientometrics and have been used to understand specific research areas and disciplines [30]. In this study, biomedical journals that published deep learning studies were grouped using the WoS Core Collection subject categories and analyzed to identify widely studied research areas and disciplines.

MeSH Terms

Disease-related MeSH terms were analyzed to identify major diseases that have been studied using deep learning. We mapped descriptors to their corresponding numbers in MeSH Tree Structures to identify higher level concepts for descriptors that were too specific and ensured that all the descriptors had the same level of specificity. Ultimately, all descriptors were mapped to 6-digit tree numbers (C00.000), and terms with >1 tree number were separately counted for all the categories they belonged to. In addition, we visualized the co-occurrence network of major MeSH descriptors using VOSviewer (version 1.6.15) [31,32] and its clustering technique [33] to understand the relationships among the biomedical entities, as well as the clusters they form together.

Author Affiliations

We analyzed author affiliations to understand the major organizations and academic disciplines that were active in deep learning research. The affiliations of 4908 authors extracted from PubMed records were recorded in various formats and manually standardized. We manually reviewed the affiliations to extract organizations, universities, schools, colleges, and departments. For authors with multiple affiliations, we selected the first one listed, which is usually the primary. We also analyzed coauthorships to investigate research collaboration among organizations and disciplines. All the organizations were grouped into one of the following categories: universities, hospitals, companies, or research institutes and government agencies to understand research collaboration among different sectors. We classified medical schools under hospitals as they are normally affiliated with each other. In the category of research institutes or government agencies, we included nonprofit private organizations or foundations and research centers that do not belong to a university, hospital, or company. We extracted academic disciplines from the department section or the school or college section when department information was unavailable. As the extracted disciplines were not coherent with multiple levels and combinations, data were first cleaned with OpenRefine (originally developed by Metaweb then Google), an interactive data transformation tool for profiling and cleaning messy data [34], and then manually grouped based

on WoS categories and MeSH Tree Structures according to the following rules. We treated interdisciplinary fields and fields with high occurrence as separate disciplines from their broader fields and aggregated multiple fields that frequently co-occurred under a single department name into a single discipline after reviewing their disciplinary similarities.

Content Analysis

We identified influential studies by examining their citation counts in PubMed and Google Scholar. Citation counts from Google Scholar were considered in addition to PubMed as Google Scholar's substantial citation data encompasses WoS and Scopus citations [35]. After sorting the articles in descending order of citations, the 2 sources showed a Spearman rank correlation coefficient of 0.883. From the PubMed top 150 list (ie, citation count >7) and Google Scholar top 150 list (ie, citation count >36), we selected the top 109 articles. Among these, we selected the sources that met the criteria for applying or developing deep learning models as the subjects of analysis to understand the major deep learning architectures in biomedicine and their applications. Specifically, we analyzed the research topics of the studies, the data and architectures used for those purposes, and how the *black box* problem was addressed.

Cited Reference Analysis

We collected the references from downloaded articles that had PubMed IDs. Citations represent the diffusion of knowledge from cited to citing publications; therefore, analyzing the highly cited references in deep learning studies in biomedicine allows for the investigation of disciplines and studies that have greatly influenced the field. Toward this end, we visualized networks of knowledge diffusion among WoS subjects using Gephi (v0.9.2) [36] and examined metrics such as modularity, PageRank score, and weighted outdegree using modularity for community detection [37]. PageRank indicates the importance of a node by measuring the quantity and quality of its incoming edges [38], and weighted outdegree measures the number of outgoing edges of a node. We also reviewed the contents of the 10 most highly cited influential works.

Results

Metadata Analysis

Journals

On the basis of the data set, 315 biomedical journals have published deep learning studies, and Table 1 lists the top 10 journals selected based on publication size. Different WoS categories and MeSH terms are separated using semicolons.

From a total of 978 records, 96 (9.8%) were unindexed in the WoS Core Collection and were excluded, following which, an average of 2.02 (SD 1.19) categories were assigned per record. The top ten subject categories, which mostly pertained to (1) biomedicine, with 22.2% (196/882) articles published in *Radiology, Nuclear Medicine, and Medical Imaging* (along with *Engineering, Biomedical*: 121/882, 13.7%; *Mathematical and Computational Biology*: 107/882, 12.1%; *Biochemical Research Methods*: 103/882, 11.7%; *Biotechnology and Applied Microbiology*: 76/882, 8.6%; *Neurosciences*: 74/882, 8.4%); (2) computer science and engineering (*Computer Science, Interdisciplinary Applications*: 112/882, 12.7%; *Computer Science, Artificial Intelligence*: 75/882, 8.5%; *Engineering, Electrical and Electronic*: 75/882, 8.5%); or (3) *Multidisciplinary Sciences* (82/882, 9.3%).

Table 1. Top 10 journals with the highest record counts.

Journal title	Web of Science category	National Library of Medicine catalog Medical Subject Heading term	Publisher	Record count, n
BMC[a] *Bioinformatics*	Biochemical Research Methods; Mathematical and Computational Biology; Biotechnology and Applied Microbiology	Computational Biology	BMC	38
Scientific Reports	Multidisciplinary Sciences	Natural Science Disciplines	Nature Research	37
Neural Networks	Neurosciences; Computer Science, Artificial Intelligence	Nerve Net; Nervous System	Elsevier	35
Proceedings of the Annual International Conference of the IEEE[b] *Engineering in Medicine and Biology Society*	N/A[c]	Biomedical Engineering	IEEE	31
IEEE Transactions on Medical Imaging	Imaging Science and Photographic Technology; Engineering, Electrical and Electronic; Computer Science, Interdisciplinary Applications; Radiology, Nuclear Medicine, and Medical Imaging; Engineering, Biomedical	Electronics, Medical; Radiography	IEEE	30
Sensors	Chemistry, Analytical; Electrochemistry; Instruments and Instrumentation; Engineering, Electrical and Electronic	Biosensing Techniques	Multidisciplinary Digital Publishing Institute	26
Bioinformatics	Biochemical Research Methods; Mathematical and Computational Biology; Biotechnology and Applied Microbiology	Computational Biology; Genome	Oxford University Press	22
Nature Methods	Biochemical Research Methods	Biomedical Research/methods; Research Design	Nature Research	21
Medical Physics	Radiology, Nuclear Medicine, and Medical Imaging	Biophysics	American Association of Physicists in Medicine	20
PloS one	Multidisciplinary Sciences	Medicine; Science	Public Library of Science	20

[a]BMC: BioMed Central.

[b]IEEE: Institute of Electrical and Electronics Engineers.

[c]N/A: not applicable.

MeSH Terms

For the main MeSH term or descriptor, an average of 9 (SD 4.21) terms was assigned to each record as subjects. Among them, we present in Figure 1 the diseases that were extracted from the *C* category. In the figure, the area size is proportional to the record count, and the terms are categorized by color. In addition, terms under >1 category were counted multiple times. For instance, the term *Digestive System Neoplasms* has two parents in MeSH Tree Structures, *Neoplasms* and *Digestive System Diseases*, and as such, we counted articles in this category under *Neoplasmsby Site* as well as under *Digestive System Neoplasms*. Owing to the limited space, 7 categories whose total record counts were ≤10 (eg, *Congenital, Hereditary, and Neonatal Diseases and Abnormalities*; *Nutritional and Metabolic Diseases*; and *Stomatognathic Diseases*) were combined under the *Others* category, and individual diseases that had <10 record counts were summed up with each other in the same category to show only their total count (or with one of the diseases included as an example). In the process, we identified *Neoplasms* as the most frequently studied disease type, with a total of 199 studies.

We further constructed a co-occurrence network of the complete set of major MeSH descriptors assigned to the records to understand the relationships among the biomedical entities. To enhance legibility, we filtered out terms with <5 occurrences. Figure 2 presents the visualized network of nodes (100/966, 10.4% of the total terms) with 612 edges and 7 clusters. In the figure, the sizes of the nodes and edges are proportional to the number of occurrences, and the node color indicates the assigned cluster (although the term *deep learning* was considered nonexclusive to any cluster as it appeared in all records).

Figure 1. Disease-related Medical Subject Heading descriptors studied with deep learning.

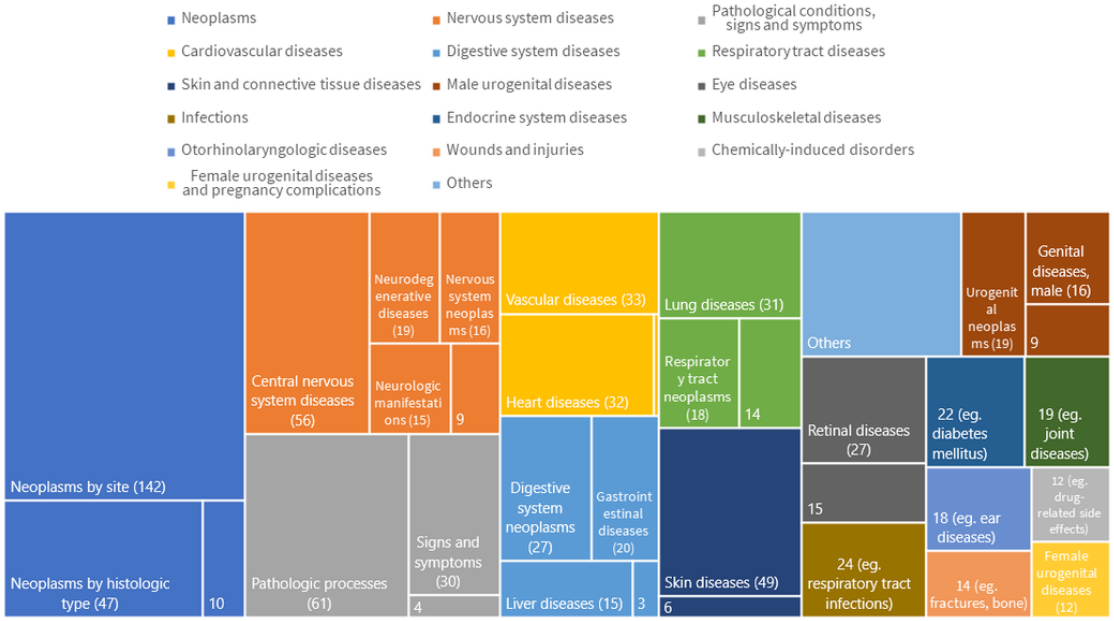
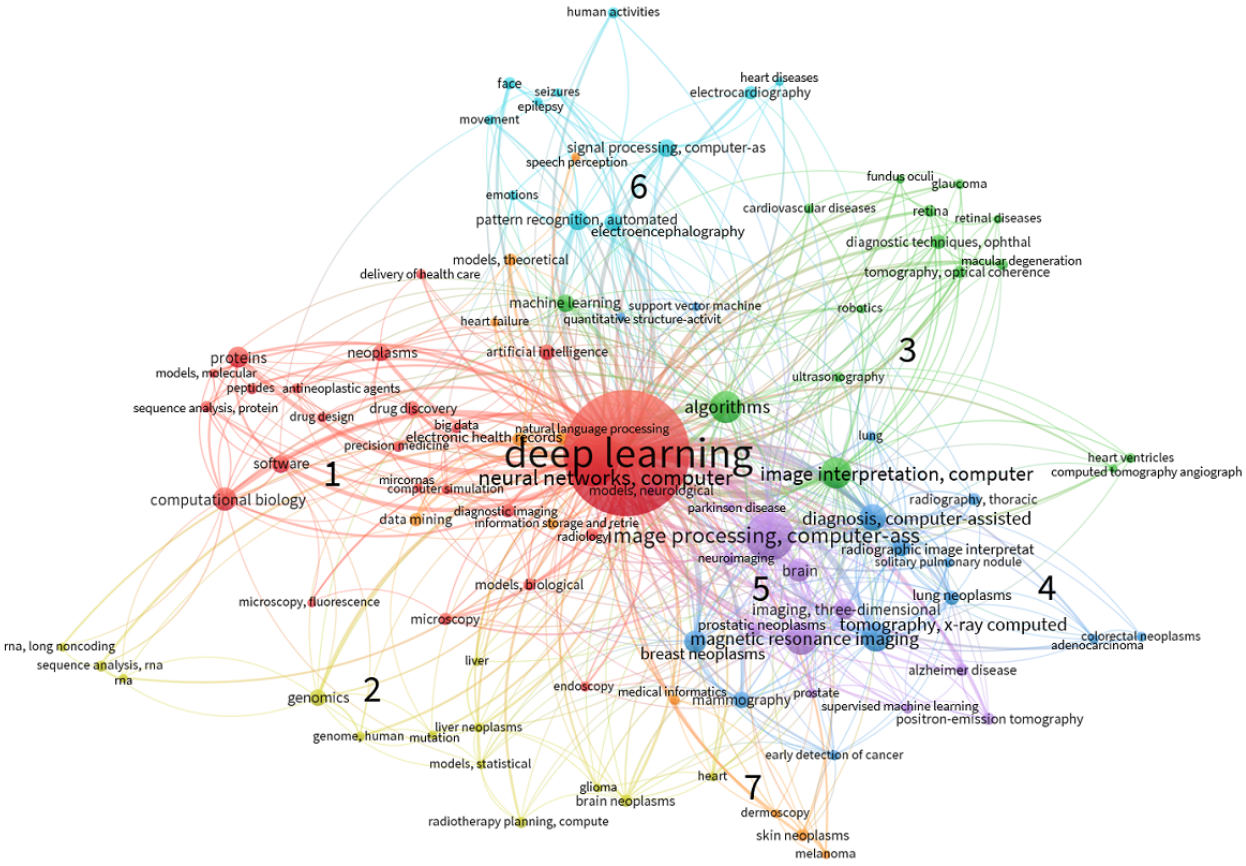


Figure 2. Co-occurrence network of the major Medical Subject Heading descriptors (number of nodes=100; number of edges=612; number of clusters=7).



As depicted in Figure 2, each cluster comprised descriptors from two groups: (1) biomedical domains that deep learning was applied to, including body regions, related diseases, diagnostic imaging methods, and theoretical models, and (2) the purposes of deep learning and techniques used for the tasks, including diagnosis, analysis, and processing of biomedical data. In the first cluster, *computer neural networks* and *software* were studied for the purposes of *computational biology*, specifically *protein sequence analysis*, *drug discovery*, and *drug design*, to achieve *precision medicine*. These were relevant to the biomedical domains of (1) *proteins*, related visualization methods (*microscopy*), and *biological models*, and (2) *neoplasms*, related drugs (*antineoplastic agents*), and *diagnostic imaging* (*radiology*). In the second cluster, deep learning and *statistical models* were used for *RNA sequence analysis* and *computer-assisted radiotherapy planning* in relation to the

domains of (1) *genomics*, *RNA*, and *mutation*, and (2) *brain neoplasms* and *liver neoplasms*. The third cluster comprised (1) heart structures (*heart ventricles*), *cardiovascular diseases*, and *ultrasonography* and (2) eye structures (*retina*), diseases (*glaucoma*), and *ophthalmological diagnostic techniques*. These had been studied for *computer-assisted image interpretation* using *machine learning* and deep learning *algorithms*. The biomedical domain group of the fourth cluster involved specific terms related to neoplasms such as type (*adenocarcinoma*), different regions (*breast neoplasms*, *lung neoplasms*, and *colorectal neoplasms*), and respective imaging methods (*mammography* and *X-ray computed tomography*) to which deep learning and *support vector machines* have been applied for the purpose of *computer-assisted radiographic image interpretation* and *computer-assisted diagnosis*. The fifth cluster included (1) *brain* disorders (*Alzheimer disease*), *neuroimaging*, and *neurological models*; (2) *prostatic neoplasms*; and (3) diagnostic *magnetic resonance imaging* and *3D imaging*. S*upervised machine learning* had been used for *computer-assisted image processing* of these data. In the sixth cluster, *automated pattern recognition* and *computer-assisted signal processing* were studied with (1) *human activities* (eg, *movement* and *face*), (2) abnormal brain activities (*epilepsy* and *seizures*) and monitoring methods (*electroencephalography*), and (3) *heart diseases* and *electrocardiography*. In the last cluster, *medical informatics*, specifically *data mining* and *NLP*, including *speech perception*, had been applied to (1) *electronic health records*, related *information storage and retrieval*, and *theoretical models* and (2) skin diseases (*skin neoplasms* and *melanoma*) and diagnostic *dermoscopy*.

Author Affiliations

To investigate research collaboration within the field, we analyzed paper-based coauthorships using author affiliations with different levels of granularity, including organization and academic disciplines. We extracted organizations from 98.7% (4844/4908) of the total affiliations and visualized the collaboration of different organization types. The top 10 organizations with the largest publication records included Harvard University (37/844, 4.4%), Chinese Academy of Sciences (21/844, 2.5%; eg, Institute of Computing Technology, Institute of Automation, and Shenzhen Institutes of Advanced Technology), Seoul National University (21/844, 2.5%), Stanford University (20/844, 2.4%), Sun Yat-sen University (14/844, 1.7%; eg, Zhongshan Ophthalmic Center and Collaborative Innovation Center of Cancer Medicine), University of California San Diego (14/844, 1.7%; eg, Institute for Genomic Medicine, Shiley Eye Institute, and Institute for Brain and Mind), University of California San Francisco (14/844, 1.7%), University of Michigan (14/844, 1.7%), Yonsei University (14/844, 1.7%), and the University of Texas Health Science Center at Houston (12/844, 1.4%). The extracted organizations were assigned to one of the following four categories according to their main purpose: universities, hospitals, companies, or research institutes and government agencies. Among these, universities participated in most papers (567/844, 67.2%), followed by hospitals (429/844, 50.8%), companies (139/844, 16.5%), and research institutes or government agencies (88/844, 10.4%). We used a co-occurrence matrix to visualize the degrees of organizational collaboration, with the co-occurrence values log normalized to compare the relative differences (Figure 3).

From Figure 3, we found that universities were the most active in collaborative research, particularly with hospitals, followed by companies and research institutes or government agencies. Hospitals also frequently collaborated with companies; however, research institutes or government agencies tended not to collaborate much as they published relatively fewer studies.

We also examined the collaborations among academic disciplines, which we could extract, as described in the *Methods* section, from 76.24% (3742/4908) of the total affiliations. Approximately half (ie, 386/756, 51.1%) of the papers were completed under disciplinary collaboration. Figure 4 depicts the network with 36 nodes (36/148, 24.3% of the total) and 267 edges after we filtered out disciplines with weighted degrees <10, representing the number of times one collaborated with the other disciplines. In the figure, the node and edge sizes are proportional to the weighted degree and link strength, respectively, and the node color indicates the assigned cluster.

As shown in the figure, the academic disciplines were assigned to 1 of 6 clusters, including 1 engineering-oriented cluster (cluster 1) and other clusters that encompassed biomedical fields. We specifically looked at the degree of collaboration between the biomedical and engineering disciplines. Figure 4 depicts that the most prominent collaboration was among *Radiology, Medical Imaging, and Nuclear Medicine*; *Computer Science*; and *Electronics and Electrical Engineering*. There were also strong links among *Computer Science* or *Electronics and Electrical Engineering* and *Biomedical Informatics*, *Biomedical Engineering*, and *Pathology and Laboratory Medicine*.

Among the top 10 disciplines in Figure 4, the following three had published the most papers and had the highest weighted degree and degree centralities: *Computer Science* (number of papers=195, weighted degree=193, and degree centrality=32); *Radiology, Medical Imaging, and Nuclear Medicine* (number of papers=168, weighted degree=166, and degree centrality=30); and *Electronics and Electrical Engineering* (number of papers=161, weighted degree=160, and degree centrality=32). Meanwhile, some disciplines had high weighted degrees compared with their publication counts, indicating their activeness in collaborative research. These included *Pathology and Laboratory Medicine* (5th in link strength vs 8th in publications) and *Public Health and Preventive Medicine* (9th in link strength vs 15th in publications). A counterexample was *Computational Biology*, which was 12th in link strength but 7th in publications.

Figure 3. Collaboration of organization types.

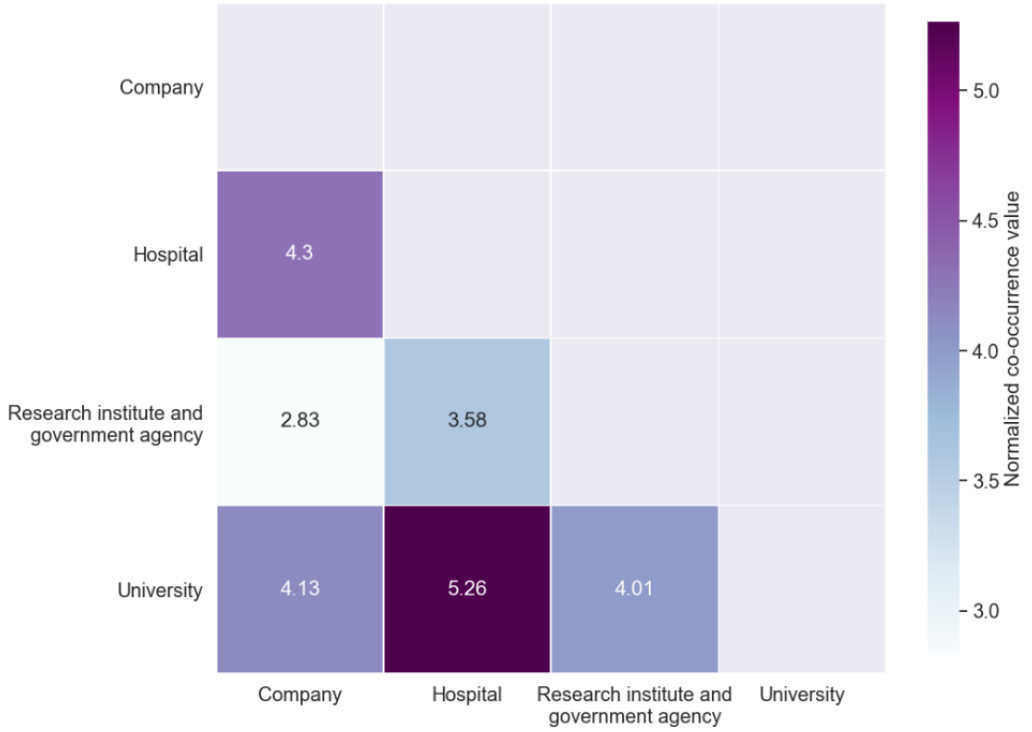
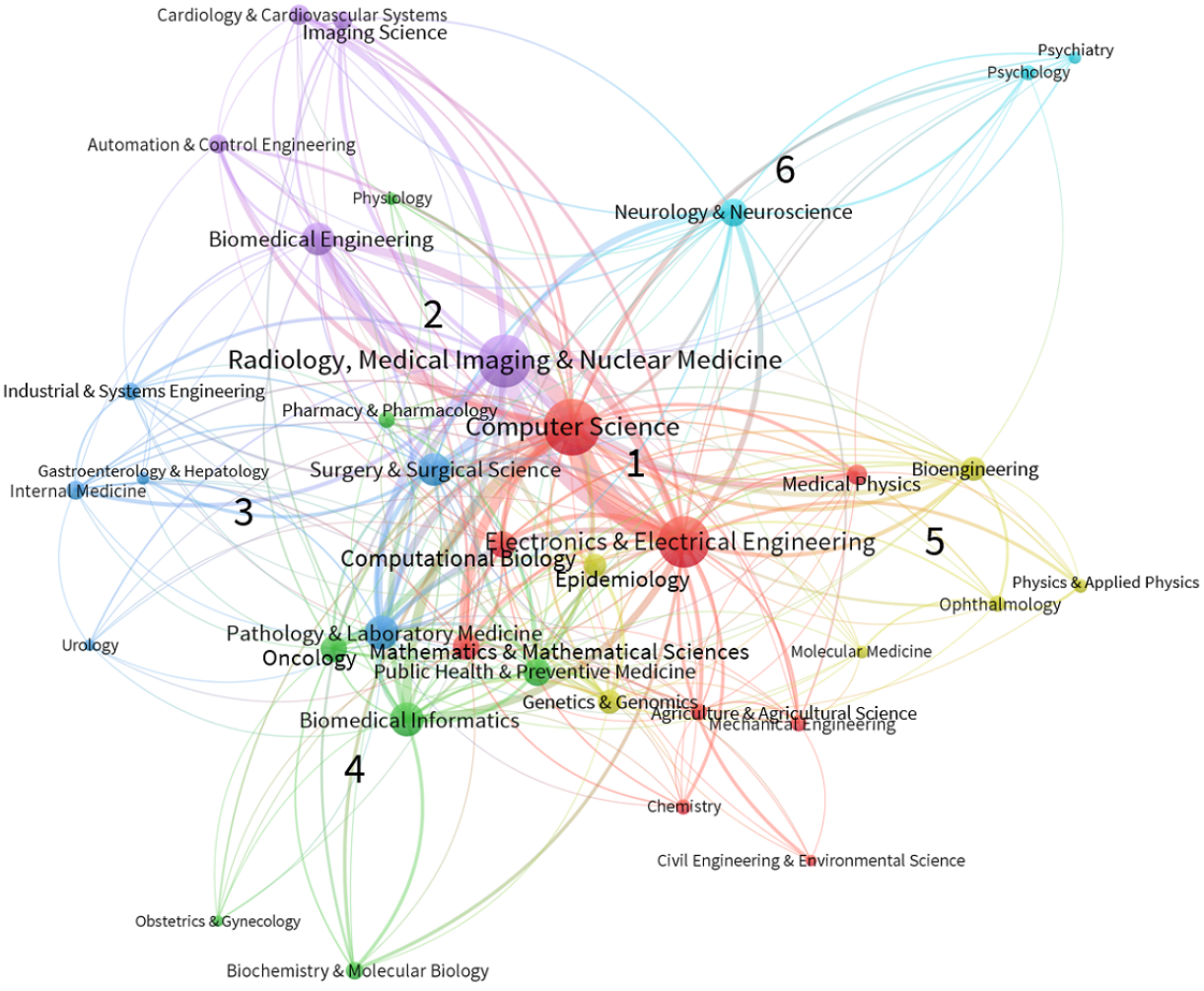


Figure 4. Collaboration network of academic disciplines (number of nodes=36; number of edges=267; number of clusters=6).



Content Analysis

Overview

We analyzed the content of influential studies that had made significant contributions to the field through the application or development of deep learning architectures. We identified these studies by examining the citation counts from PubMed and Google Scholar, assigning the 109 most-cited records to one of the following categories: (1) *review*, (2) *application* of existing deep learning architectures to certain biomedical domains (denoted by *A*), or (3) *development* of a novel deep learning model (denoted by *D*). Table 2 summarizes the 92 papers assigned to the application or development category according to their research topic in descending order of citation count.

Table 2. Top 92 studies with the highest citation count under the application or development category, according to the research topic.

Research topic and number	Task type	Data	Deep learning architectures
(Diagnostic) image analysis			
A1 [39]	Classification	Retinal disease OCT[a] and chest x-ray with pneumonia	Inception
A2 [40]	Segmentation and classification	Retinal disease OCT	U-net and CNN[b]
A3 [41]	Classification	Melanoma dermoscopic images	Inception
A4 [42]	Survival prediction	Brain glioblastoma MRI[c]	CNN_S
A6 [43]	Classification and segmentation	WSI[d] of 13 cancer types	CNN with CAE[e] and DeconvNet
D1 [44]	Segmentation	Brain MRI	ResNet[f] based
A7 [45]	Prediction	Retinal fundus images with cardiovascular disease	Inception
D2 [46]	Tracking	Video of freely behaving animal	ResNet-based DeeperCut subset
A8 [47]	Classification	Colonoscopy video of colorectal polyps	Inception
A9 [48]	Classification	Lung cancer CT[g]	CNN
A10 [49]	Classification and segmentation	Retinal OCT with macular disease	Encoder-decoder CNN
D3 [50]	Segmentation	Brain glioma MRI	CNN based
D4 [51]	Binding affinities prediction	Protein-ligand complexes as voxel	SqueezeNet based
A11 [52]	Survival classification	Brain glioma MRI, functional MRI, and DTI[h]	CNN and mCNN[i]
A12 [53]	Classification	Fundus images with glaucomatous optic neuropathy	Inception
A13 [54]	Classification	Chest radiographs with pneumonia	ResNet and CheXNet
A14 [55]	Classification and segmentation	Critical head abnormality CT	ResNet, U-net, and DeepLab
A15 [56]	Classification	Brain glioma MRI	ResNet
D6 [57]	Classification	Thoracic disease radiographs	DenseNet based
A16 [58]	Classification and segmentation	Echocardiogram video with cardiac disease	VGGNet and U-net
A17 [59]	Classification	Brain positron emission tomography with Alzheimer	Inception
D7 [60]	Classification	Breast cancer histopathological images	CNN based
A18 [61]	Classification	Skin tumor images	ResNet
A19 [62]	Classification and prediction	Chest CT with chronic obstructive pulmonary disease and acute respiratory disease	CNN
A20 [63]	Segmentation	Brain MRI with autism spectrum disorder	FCNN[j]
D8 [16]	Segmentation	Fetal MRI and brain tumor MRI	Proposal network (P-Net) based
A21 [64]	Classification, prediction, and reconstruction	Natural movies and functional MRI of watching movies	AlexNet and De-CNN
D9 [65]	Detection and classification	Facial images with a genetic syndrome	CNN based
A22 [66]	Detection and segmentation	Microscopic images of cells	U-net
A23 [67]	Classification and localization	Breast cancer mammograms	Faster region-based CNN with VGGNet
A24 [68]	Segmentation and prediction	Lung cancer CT	Mask-RCNN, CNN with GoogLeNet and RetinaNet
A26 [69]	Classification	Lung cancer CT	CNN; fully connected NN; SAE[k]
A27 [70]	Survival classification	Lung cancer CT	CNN
A29 [71]	Prediction	Polar maps of myocardial perfusion imaging with CAD[l]	CNN

Research topic and number	Task type	Data	Deep learning architectures
A30 [72]	Classification	Prostate cancer MRI	CNN
D12 [73]	Classification	Liver SWE[m] with chronic hepatitis B	CNN based
D14 [74]	Segmentation	Liver cancer CT	DenseNet with U-net based
A31 [75]	Classification	Fundus images with macular degeneration	AlexNet, GoogLeNet, VGGNet, inception, ResNet, and inception-ResNet
A32 [76]	Classification	Bladder cancer CT	cuda-convnet
A34 [77]	Classification	Prostate cancer tissue microarray images	MobileNet
D19 [78]	Classification	Holographic microscopy of *Bacillus* species	CNN based
A36 [79]	Survival classification	Chest CT	CNN
D20 [80]	Classification and localization	Malignant lung nodule radiographs	ResNet based
A37 [81]	Classification	Shoulder radiographs with proximal humerus fracture	ResNet
A39 [82]	Classification	Facial images of hetero and homosexual	VGG-Face
A41 [83]	Segmentation and classification	CAD CT angiography	CNN and CAE
A42 [84]	Classification and localization	Radiographs with fracture	U-net
A43 [85]	Binding classification	Peptide major histocompatibility complex as image-like array	CNN
A44 [86]	Detection	Lung nodule CT	CNN
A45 [87]	Classification	Confocal endomicroscopy video of oral cancer	LeNet
A46 [88]	Classification	WSI of prostate, skin, and breast cancer	MIL[n] with ResNet and RNN
D24 [89]	Tracking	Video of freely behaving animal	FCNN based
D25 [90]	Segmentation	Fundus images with glaucoma	U-net based
A47 [91]	Segmentation and classification	Cardiac disease cine MRI	U-net; M-Net; Dense U-net; SVF-Net; Grid-Net; Dilated CNN
D27 [92]	Classification	Knee abnormality MRI	AlexNet based
D28 [93]	Binding affinities prediction	Protein-ligand complexes as grid	CNN based
A50 [94]	Segmentation	Autosomal dominant polycystic kidney disease CT	FCNN with VGGNet
A51 [95]	Segmentation and classification	Knee cartilage lesion MRI	VGGNet
A52 [96]	Classification	Mammograms	ResNet
A54 [97]	Prediction	CAD CT angiography	FCNN
D31 [98]	Classification and localization	WSI of lymph nodes in metastatic breast cancer	Inception based
D35 [99]	Classification	Fluorescence microscopic images of cells	FFNN[o] based
A56 [100]	Classification	Retinal fundus images with diabetic retinopathy and breast mass mammography	ResNet; GoogLeNet
Image processing			
A25 [101]	Artifact reduction	Brain and abdomen CT and radial MR[p] data	U-net
A28 [102]	Resolution enhancement	Fluorescence microscopic images	GAN[q] with U-net and CNN
D15 [103]	Dealiasing	Compressed sensing brain lesion and cardiac MRI	GAN with U-net and VGGNet based
D16 [104]	Resolution enhancement	Superresolution localization microscopic images	GAN with U-net–based pix2pix network modified
A33 [105]	Reconstruction	Brain and pelvic MRI and CT	GAN with FCNN and CNN
D18 [106]	Artifact reduction	CT	CNN based
A38 [107]	Reconstruction	Contrast-enhanced brain MRI	Encoder-decoder CNN

Research topic and number	Task type	Data	Deep learning architectures
D22 [108]	Reconstruction	Brain MR fingerprinting data	FFNN based
D23 [109]	Resolution enhancement	Hi-C matrix of chromosomes	CNN based
A48 [110]	Resolution enhancement	Brain tumor MRI	U-net
D26 [111]	Reconstruction	Lung vessels CT	CNN based
D32 [112]	Resolution enhancement	Knee MRI	CNN based
D33 [113]	Reconstruction	CT	CNN based
D34 [18]	Registration	Cardiac cine MRI and chest CT	CNN based
Sequence analysis			
D17 [114]	Novel structures generation and property prediction	SMILES[r]	Stack-RNN[s] with GRU[t]- and LSTM[u] based
A40 [115]	Novel structures generation	SMILES	variational AE[v]; CNN- and RNN with GRU-based AAE[w]
D21 [116]	Gene expression (variant effects) prediction	Genomic sequence	CNN based
D30 [117]	Novel structures generation and classification	SMILES	GAN with differentiable neural computer and CNN based
A53 [118]	Novel structures generation	SMILES	LSTM
A57 [119]	Classification	Antimicrobial peptide sequence	CNN with LSTM
Sequence and image analysis			
D13 [120]	Contact prediction	Protein sequence to contact matrix	ResNet based
(Diagnostic) pattern analysis			
A5 [121]	Subtype identification (survival classification)	Multi-omics data from liver cancer	AE
D5 [122]	Phenotype prediction	Genotype	GoogLeNet and deeply supervised net based
D10 [123]	Survival prediction	Genomic profiles from cancer	FFNN based
D11 [124]	Drug synergies prediction	Gene expression profiles of cancer cell line and chemical descriptors of drugs	FFNN based
A35 [125]	NLP[x] (classification)	Electronic health record with pediatric disease	Attention-based BLSTM[y]
A49 [126]	Binding classification	Protein sequence as matrix and drug molecular fingerprint	SAE
D29 [127]	Classification	Electrocardiogram signal	BLSTM based

Research topic and number	Task type	Data	Deep learning architectures
A55 [128]	Classification	Polysomnogram signal	CNN

[a]OCT: optical coherence tomography.
[b]CNN: convolutional neural network.
[c]MRI: magnetic resonance imaging.
[d]WSI: whole slide image.
[e]CAE: convolutional autoencoder.
[f]ResNet: residual networks.
[g]CT: computed tomography.
[h]DTI: diffusion tensor imaging.
[i]mCNN: multicolumn convolutional neural network.
[j]FCNN: fully convolutional neural network.
[k]SAE: stacked autoencoder.
[l]CAD: coronary artery disease.
[m]SWE: shear wave elastography.
[n]MIL: multiple instance learning.
[o]FFNN: feedforward neural network.
[p]MR: magnetic resonance.
[q]GAN: generative adversarial network.
[r]SMILES: simplified molecular input line-entry system.
[s]RNN: recurrent neural network.
[t]GRU: gated recurrent unit.
[u]LSTM: long short-term memory.
[v]AE: autoencoder.
[w]AAE: adversarial autoencoder.
[x]NLP: natural language processing.
[y]BLSTM: bidirectional long short-term memory.

Research Topics

In these studies, researchers applied or developed deep learning architectures mainly for the following purposes: image analysis, especially for diagnostic purposes, including the classification or prediction of diseases or survival, and the detection, localization, or segmentation of certain areas or abnormalities. These 3 tasks, which aim to identify the location of an object of interest, are different in that detection involves a single reference point, whereas localization involves an area identified through a bounding box, saliency map, or heatmap, segmentation involves a precise area with clear outlines identified through pixel-wise analysis. Meanwhile, in some studies, models for image analysis unrelated to diagnosis were proposed, such as classifying or segmenting cells in microscopic images and tracking moving animals in videos through pose estimation. Another major objective involved image processing for reconstructing or registering medical images. This included enhancing low-resolution images to high resolution, reconstructing images with different modalities or synthesized targets, reducing artifacts, dealiasing, and aligning medical images.

Meanwhile, several researchers used deep learning architectures to analyze molecules, proteins, and genomes for various purposes. These included drug design or discovery, specifically for generating novel molecular structures through sequence analysis and for predicting binding affinities through image analysis of complexes; understanding protein structure through image analysis of contact matrix; and predicting phenotypes, cancer survival, drug synergies, and genomic variant effects from genes or genomes. Finally, in some studies, deep learning was applied to the diagnostic classification of sequential data, including electrocardiogram or polysomnogram signals and electronic health records. In summary, in the reviewed literature, we identified a predominant focus on applying or developing deep learning models for image analysis regarding localization or diagnosis and image processing, with a few studies focusing on protein or genome analysis.

Deep Learning Architectures

Regarding the main architectures, most of them were predominantly CNNs and based on ≥1 CNN architecture such as a fully CNN (FCNN) and its variants, including U-net; residual neural network (ResNet) and its variants; GoogLeNet (Inception v1) or Inception and VGGNet and its variants; and other architectures. Meanwhile, a few researchers based their models on feedforward neural networks that were not CNNs, including autoencoders (AEs) such as convolutional AE and stacked AE. Others adapted RNNs, including (bidirectional) long short-term memory and gated recurrent unit. Furthermore, models that combined RNNs or AEs with CNNs were also proposed.

Content analysis of the reviewed literature showed that different deep learning architectures were used for different research

tasks. Models for classification or prediction tasks using images were predominantly CNN based, with most being ResNet and GoogLeNet or Inception. ResNet with shortcut connections [129] and GoogLeNet or Inception with 1×1 convolutions, factorized convolutions, and regularizations [130,131] allow networks of increased depth and width by solving problems such as vanishing gradients and computational costs. These mostly analyzed medical images from magnetic resonance imaging or computed tomography, with cancer-related images often used as input data for diagnostic classification, in addition to image-like representations of protein complexes. Meanwhile, when applying these tasks to data other than images, such as genomic or gene expression profiles and protein sequence matrices, researchers used feedforward neural networks, including AEs, that enabled semi- or unsupervised learning and dimensionality reduction.

Image analysis for segmentation and image processing were achieved through CNN-based architectures as well, with most of them being FCNNs, especially U-net. FCNNs produce an input-sized pixel-wise prediction by replacing the last fully connected layers to convolution layers, making them advantageous for the abovementioned tasks [132], and U-net enhances these performances through long skip connections that concatenate feature maps from the encoder path to the decoder path [133]. In particular, for medical image processing tasks, a few researchers combined FCNNs (U-net) with other CNNs by adopting the generative adversarial network structure, which generates new instances that mimic the real data through an adversarial process between the generator and discriminator [134]. We found that images of the brain were often used as input data for these studies.

On the other hand, RNNs were applied to sequence analysis of the string representation of molecules (simplified molecular input line-entry system) and pattern analysis of sequential data such as signals. A few of these models, especially those generating novel molecular structures, combined RNNs with CNNs by adopting generative adversarial networks, including adversarial AE. In summary, the findings showed that the current deep learning models were predominantly CNN based, with most of them focusing on analyzing medical image data and different architectures that are preferred for the specific tasks.

Among these studies, Table 3 shows, in detail, the objectives and the proposed methods of the 35 studies with novel model development.

Table 3. Content analysis of the top 35 records in the development category.

Number	Development objectives	Methods (proposed model)
D1	Segment brain anatomical structures in 3D MRI[a]	Voxelwise Residual Network: trained through residual learning of volumetric feature representation and integrated with contextual information of different modalities and levels
D2	Estimate poses to track body parts in various animal behaviors	DeeperCut's subset DeepLabCut: network fine-tuned on labeled body parts, with deconvolutional layers producing spatial probability densities to predict locations
D3	Predict isocitrate dehydrogenase 1 mutation in low-grade glioma with MRI radiomics analysis	Deep learning–based radiomics: segment tumor regions and directly extract radiomics image features from the last convolutional layer, which is encoded for feature selection and prediction
D4	Predict protein-ligand binding affinities represented by 3D descriptors	KDEEP: 3D network to predict binding affinity using voxel representation of protein-ligand complex with assigned property according to its atom type
D5	Predict phenotype from genotype through the biological hierarchy of cellular subsystems	DCell: visible neural network with structure following cellular subsystem hierarchy to predict cell growth phenotype and genetic interaction from genotype
D6	Classify and localize thoracic diseases in chest radiographs	DenseNet-based CheXNeXt: networks trained for each pathology to predict its presence and ensemble and localize indicative parts using class activation mappings
D7	Multi-classification of breast cancer from histopathological images	CSDCNN[b]: trained through end-to-end learning of hierarchical feature representation and optimized feature space distance between breast cancer classes
D8	Interactive segmentation of 2D and 3D medical images fine-tuned on a specific image	Bounding box and image-specific fine-tuning–based segmentation: trained for interactive image segmentation using bounding box and fine-tuned for specific image with or without scribble and weighted loss function
D9	Facial image analysis for identifying phenotypes of genetic syndromes	DeepGestalt: preprocessed for face detection and multiple regions and extracts phenotype to predict syndromes per region and aggregate probabilities for classification
D10	Predict cancer outcomes with genomic profiles through survival models optimization	SurvivalNet: deep survival model with high-dimensional genomic input and Bayesian hyperparameter optimization, interpreted using risk backpropagation
D11	Predict synergy effect of novel drug combinations for cancer treatment	DeepSynergy: predicts drug synergy value using cancer cell line gene expressions and chemical descriptors, which are normalized and combined through conic layers
D12	Classify liver fibrosis stages in chronic hepatitis B using radiomics of SWE[c]	DLRE[d]: predict the probability of liver fibrosis stages with quantitative radiomics approach through automatic feature extraction from SWE images
D13	Predict protein residue contact map at pixel level with protein features	RaptorX-Contact: combined networks to learn contact occurrence patterns from sequential and pairwise protein features to predict contacts simultaneously at pixel level
D14	Segment liver and tumor in abdominal CT[e] scans	Hybrid Densely connected U-net: 2D and 3D networks to extract intra- and interslice features with volumetric contexts, optimized through hybrid feature fusion layer
D15	Reconstruct compressed sensing MRI to dealiased image	DAGAN[f]: conditional GAN[g] stabilized by refinement learning, with the content loss combined adversarial loss incorporating frequency domain data
D16	Reconstruct sparse localization microscopy to superresolution image	Artificial Neural Network Accelerated–Photoactivated Localization Microscopy: trained with superresolution PALM[h] as the target, compares reconstructed and target with loss functions containing conditional GAN
D17	Generate novel chemical compound design with desired properties	Reinforcement Learning for Structural Evolution: generate chemically feasible molecule as strings and predict its property, which is integrated with reinforcement learning to bias the design
D18	Reduce metal artifacts in reconstructed x-ray CT images	CNN[i]-based Metal Artifact Reduction: trained on images processed by other Metal Artifact Reduction methods and generates prior images through tissue processing and replaces metal-affected projections
D19	Predict *Bacillus* species to identify anthrax spores in single cell holographic images	HoloConvNet: trained with raw holographic images to directly recognize interspecies difference through representation learning using error backpropagation
D20	Classify and detect malignant pulmonary nodules in chest radiographs	Deep learning–based automatic detection: predict the probability of nodules per radiograph for classification and detect nodule location per nodule from activation value
D21	Predict tissue-specific gene expression and genomic variant effects on the expression	ExPecto: predict regulatory features from sequences and transform to spatial features and use linear models to predict tissue-specific expression and variant effects
D22	Reconstruct MRF[j] to obtain tissue parameter maps	Deep reconstruction network: trained with a sparse dictionary that maps magnitude image to quantitative tissue parameter values for MRF reconstruction
D23	Generate high-resolution Hi-C interaction matrix of chromosomes from a low-resolution matrix	HiCPlus: predict high-resolution matrix through mapping regional interaction features of low-resolution to high-resolution submatrices using neighboring regions

Number	Development objectives	Methods (proposed model)
D24	Estimate poses to track body parts of freely moving animals	LEAP[k]: videos preprocessed for egocentric alignment and body parts labeled using GUI[l] and predicts each location by confidence maps with probability distributions
D25	Jointly segment optic disc and cup in fundus images for glaucoma screening	M-Net: multi-scale network for generating multi-label segmentation prediction maps of disc and cup regions using polar transformation
D26	Reconstruct limited-view PAT[m] to high-resolution 3D images	Deep gradient descent: learned iterative image reconstruction, incorporated with gradient information of the data fit separately computed from training
D27	Predict classifications of and localize knee injuries from MRI	MRNet: networks trained for each diagnosis according to a series to predict its presence and combine probabilities for classification using logistic regression
D28	Predict binding affinities between 3D structures of protein-ligand complexes	Pafnucy: structure-based prediction using 3D grid representation of molecular complexes with different orientations as having same atom types
D29	Classify electrocardiogram signals based on wavelet transform	Deep bidirectional LSTM[n] network–based wavelet sequences: generate decomposed frequency subbands of electrocardiogram signal as sequences by wavelet-based layer and use as input for classification
D30	Generate novel small molecule structures with possible biological activity	Reinforced Adversarial Neural Computer: combined with GAN and reinforcement learning, generates sequences matching the key feature distributions in the training molecule data
D31	Detect and localize breast cancer metastasis in digitized lymph nodes slides	LYmph Node Assistant: predict the likelihood of tumor in tissue area and generate a heat map for slides identifying likely areas
D32	Transform low-resolution thick slice knee MRI to high-resolution thin slices	DeepResolve: trained to compute residual images, which are added to low-resolution images to generate their high-resolution images
D33	Reconstruct sparse-view CT to suppress artifact and preserve feature	Learned Experts' Assessment–Based Reconstruction Network: iterative reconstruction using previous compressive sensing methods, with fields of expert-applied regularization terms learned iteration dependently
D34	Unsupervised affine and deformable aligning of medical images	Deep Learning Image Registration: multistage registration network and unsupervised training to predict transformation parameters using image similarity and create warped moving images
D35	Classify subcellular localization patterns of proteins in microscopy images	Localization Cellular Annotation Tool: predict localization per cell for image-based classification of multi-localizing proteins, combined with gamer annotations for transfer learning

[a]MRI: magnetic resonance imaging.

[b]CSDCNN: class structure-based deep convolutional neural network.

[c]SWE: shear wave elastography.

[d]DLRE: deep learning radiomics of elastography.

[e]CT: computed tomography.

[f]DAGAN: Dealiasing Generative Adversarial Networks.

[g]GAN: generative adversarial network.

[h]PALM: photoactivated localization microscopy.

[i]CNN: convolutional neural network.

[j]MRF: magnetic resonance fingerprinting.

[k]LEAP: LEAP Estimates Animal Pose.

[l]GUI: graphical user interface.

[m]PAT: photoacoustic tomography.

[n]LSTM: long short-term memory.

Black Box Problem

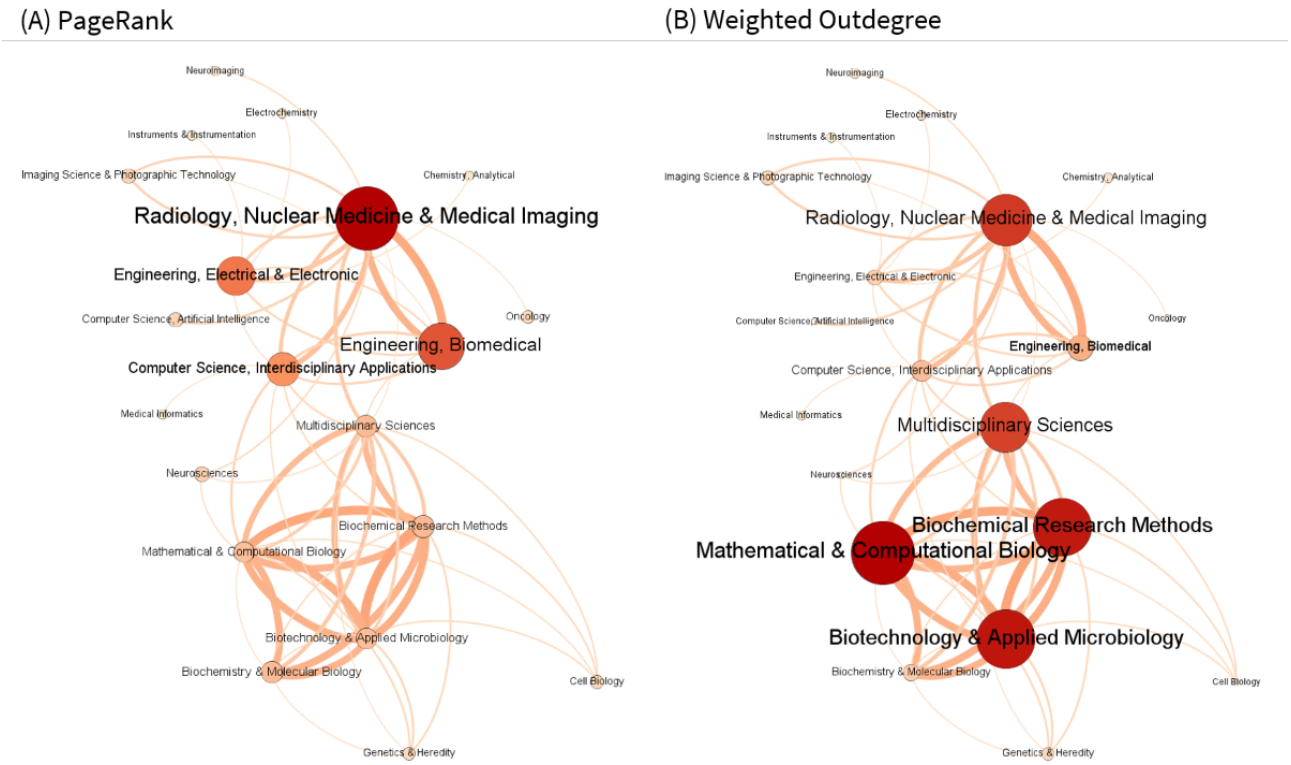
In quite a few of the reviewed studies, the *black box* problem of deep learning was partly addressed, as researchers implemented various methods to improve model interpretability. To understand the prediction results of image analysis models, most used one of the following two techniques to visualize the important regions: (1) activation-based heatmaps [45,54,65,70], especially class activation maps [57,61,77,92], and saliency maps [59] and (2) occlusion testing [39,75,82,94]. For models analyzing data other than images, there were no generally accepted techniques for model interpretation, and researchers suggested some methods, including adopting an interpretable hierarchical structure such as the cellular subsystem [122] or anatomical division [125], using backpropagation [123], observing gate activations of cells in the neural network [114], or investigating how corrupted input data affect the prediction and how identical predictions are made for different inputs [93]. As such, various methods were found to be used to tackle this well-known limitation of deep learning.

Cited Reference Analysis

On average, each examined deep learning study with at least one PubMed indexed citation (429/978, 43.9%) had 25.8 (SD 20.0) citations. These cited references comprised 9373 unique records that were cited 1.27 times on average (SD 2.16). Excluding the ones that were unindexed in the WoS Core Collection (8618/9373, 8.06% of the unique records), an average of 1.77 (SD 1.07) categories were assigned to a record. The top ten WoS categories, which were assigned to the greatest number of total cited references, pertained to the following three major groups: (1) biomedicine (*Radiology, Nuclear Medicine, and Medical Imaging*: 2025/11,033, 18.35%; *Biochemical Research Methods*: 1118/11,033, 10.13%; *Mathematical and Computational Biology*: 1066/11,033, 9.66%; *Biochemistry and Molecular Biology*: 1043/11,033, 9.45%; *Engineering, Biomedical*: 981/11,033, 8.89%; *Biotechnology and Applied Microbiology*: 916/11,033, 8.3%; *Neurosciences*: 844/11,033, 7.65%), (2) computer science and engineering (*Computer Science, Interdisciplinary Applications*: 1041/11,033, 9.44%; *Engineering, Electrical and Electronic*: 645/11,033, 5.85%), and (3) *Multidisciplinary Sciences* (with 1411/11,033, 12.79% records).

To understand the intellectual structure of how knowledge is transferred among different areas of study through citations, we visualized the citation network of WoS subject categories. In the directed citation network shown in Figure 5, the edges were directed clockwise with the source nodes as the WoS categories of the deep learning studies we examined and the target nodes as the WoS categories of the cited references from which knowledge was obtained. To enhance legibility, we filtered out categories with <100 weighted degrees, excluding self-loops, to form a network of 20 nodes (20/158, 12.7% of the total) and 59 edges (59/2380, 2.48% of the total). In the figure, the node color and size are proportional to the PageRank score (probability 0.85; ε=0.001; Figure 5A) and weighted-out degree (Figure 5B), and the edge size and color are proportional to the link strength. PageRank considers not only the quantity but also the quality of incoming edges, identifying important exporters for knowledge diffusion based on how often and by which fields a node is cited. On the other hand, the weighted outdegree measures outgoing edges and identifies major knowledge importers that frequently cite other fields.

Figure 5. Citation network of the Web of Science subject categories assigned to the reviewed publications and their cited references according to (A) PageRank and (B) weighted outdegree (number of nodes=20; number of edges=59).



As depicted in Figure 5A, categories with high PageRank scores mostly coincided with the frequently cited fields identified above and were grouped into two communities through modularity (upper half and lower half). The upper half region centered on Radiology, *Nuclear Medicine, and Medical Imaging*, which had the highest PageRank score (0.191) and proved to be a field with a significant influence on deep learning studies in biomedicine. Meanwhile, important knowledge exporters to this field included *Engineering, Biomedical* (0.134); *Engineering, Electrical and Electronic* (0.110); and *Computer Science, Interdisciplinary Applications* (0.091). The lower half region mainly comprised categories with comparable PageRank scores in which knowledge was frequently exchanged between one another, including *Biochemical Research Methods* (0.053), *Multidisciplinary Sciences* (0.053), *Biochemistry and Molecular Biology* (0.052), *Biotechnology and Applied Microbiology* (0.050), and *Mathematical and Computational Biology* (0.048). Specifically, in Figure 5B, *Mathematical and Computational*

Biology (1992), *Biotechnology and Applied Microbiology* (1836), and *Biochemical Research Methods* (1807) were identified as major knowledge importers with the highest weighted outdegrees, whereas *Biochemistry and Molecular Biology* (344) had a relatively low weighted outdegree, indicating their role as a source of knowledge for these fields.

We analyzed the 10 most frequently cited studies to gain an in-depth understanding of the most influential works and assigned these papers to one of the three categories: review, application, or development. Review articles provided comprehensive overviews of the development and applications of deep learning [1,3], with 1 focusing on applications to medical image analysis [4]. We summarize the 7 application (denoted by *A*) or development (denoted by *D*) studies in Table 4.

In these studies, excluding the study by Hochreiter and Schmidhuber [135], whose research topic pertained to computer science, deep learning was used for diagnostic image analysis of various areas [12-14,136] and for sequence analysis of proteins [21] or genomes [22]. The main architectures implemented to achieve the different research objectives mostly comprised CNNs [12-14,136] or CNN-based novel models [21,22] and RNNs [135]. The findings indicated that these deep neural networks either outperformed previous methods or achieved a performance comparable with that of human experts.

Table 4. Content analysis matrix of the highly cited references in the application or development category.

Category	Citation count, n	Research topic: task type	Objectives	Methods (deep learning architectures)
A1 [12]	53	Diagnostic image analysis: classification	Apply CNN[a] to classifying skin lesions from clinical images	Inception version 3 fine-tuned end to end with images; tested against dermatologists on 2 binary classifications
A2 [13]	51	Diagnostic image analysis: classification	Apply CNN to detecting referrable diabetic retinopathy on retinal fundus images	Inception version 3 trained and validated using 2 data sets of images graded by ophthalmologists
D1 [135]	34	Computer science	Develop a new gradient-based RNN[b] to solve error backflow problems	LSTM[c] achieved constant error flow through memory cells regulated by gate units; tested numerous times against other methods
D2 [21]	33	Sequence analysis: binding (variant effects) prediction	Propose a predictive model for sequence specificities of DNA- and RNA-binding proteins	CNN-based DeepBind trained fully automatically through parallel implementation to predict and visualize binding specificities and variation effects
A3 [14]	27	Diagnostic image analysis: classification	Evaluate factors of using CNNs for thoracoabdominal lymph node detection and interstitial lung disease classification	Compare performances of AlexNet, CifarNet, and GoogLeNet trained with transfer learning and different data set characteristics
D3 [22]	23	Sequence analysis: chromatin profiles (variant effects) prediction	Propose a model for predicting noncoding variant effects from genomic sequence	CNN-based DeepSEA trained for chromatin profile prediction to estimate variant effects with single nucleotide sensitivity and prioritize functional variants
A4 [136]	23	Diagnostic image analysis: classification	Evaluate CNNs for tuberculosis detection on chest radiographs	Compare performances of AlexNet and GoogLeNet and ensemble of 2 trained with transfer learning, augmented data set, and radiologist-augmented approach

[a]CNN: convolutional neural network.

[b]RNN: recurrent neural network.

[c]LSTM: long short-term memory.

Discussion

Principal Findings

With the increase in biomedical research using deep learning techniques, we aimed to gain a quantitative and qualitative understanding of the scientific domain, as reflected in the published literature. For this purpose, we conducted a scientometric analysis of deep learning studies in biomedicine.

Through the metadata and content analyses of bibliographic records, we identified the current leading fields and research topics, the most prominent being radiology and medical imaging. Other biomedical fields that have led this domain included biomedical engineering, mathematical and computational biology, and biochemical research methods. As part of interdisciplinary research, computer science and electrical engineering were important fields as well. The major research topics that were studied included computer-assisted image interpretation and diagnosis (which involved localizing or segmenting certain areas for classifying or predicting diseases), image processing such as medical image reconstruction or registration, and sequence analysis of proteins or RNA to understand protein structure and discover or design drugs. These topics were particularly prevalent in their application to neoplasms.

Furthermore, although deep learning techniques that had been proposed for these themes were predominantly CNN based, different architectures are preferred for different research tasks. The findings showed that CNN-based models mostly focused on analyzing medical image data, with RNN architectures for sequential data analysis and AEs for unsupervised dimensionality reduction yet to be actively explored. Other deep learning methods, such as deep belief networks [137,138], deep Q network [139], and dictionary learning [140], have also been applied to biomedical research but were excluded from the content analysis because of low citation count. As deep learning is a rapidly evolving field, future biomedical researchers should pay attention to the emerging trends and keep aware of state-of-the-art models for enhanced performance, such as transformer-based models, including bidirectional encoder representations from transformers for NLP [141]; wav2vec for speech recognition [142]; and the Swin transformer for computer vision tasks of image classification, segmentation, and object detection [143].

The findings from the analysis of the cited references revealed patterns of knowledge diffusion. In the analysis, radiology and medical imaging appeared to be the most significant knowledge source and an important field in the knowledge diffusion network. Relatedly, we identified knowledge exporters to this field, including biomedical engineering, electrical engineering, and computer science, as important, despite their relatively low citation counts. Furthermore, citation patterns revealed clique-like relationships among the four fields—biochemical research methods, biochemistry and molecular biology, biotechnology and applied microbiology, and mathematical and computational biology—with each being a source of knowledge and diffusion for the others.

Beyond knowledge diffusion, knowledge integration was also encouraged through collaboration among authors from different organizations and academic disciplines. Coauthorship analysis revealed active research collaboration between universities and hospitals and between hospitals and companies. Separately, we identified an engineering-oriented cluster and biomedicine-oriented clusters of disciplines, among which we observed a range of disciplinary collaborations, with the most prominent 2 between radiology and medical imaging and computer science and electrical engineering, which were the 3 disciplines that were most involved in publishing and collaboration. Meanwhile, pathology and public health showed a high collaborative research to publications ratio, whereas computational biology showed a low collaborative ratio.

Limitations

This study has the following limitations that may have affected data analysis and interpretation. First, focusing only on published studies may have underrepresented the field. Second, publication data were only retrieved from PubMed; although PubMed is one of the largest databases for biomedical literature, other databases such as DataBase systems and Logic Programming may also include relevant studies. Third, the use of PubMed limited our data to biomedical journals and proceedings. Given that deep learning is an active research area in computer science, computer science conference articles are valuable sources of data that were not considered in this study. Finally, our current data retrieval strategy involved searching *deep learning* as the major MeSH term, which increased precision but may have omitted relevant studies that were not explicitly tagged as *deep learning*. We plan to expand our scope in future work to consider other bibliographic databases and search terms as well.

Conclusions

In this study, we investigated the landscape of deep learning research in biomedicine and identified major research topics, influential works, knowledge diffusion, and research collaboration through scientometric analyses. The results showed a predominant focus on research applying deep learning techniques, especially CNNs, to radiology and medical imaging and confirmed the interdisciplinary nature of this domain, especially between engineering and biomedical fields. However, diverse biomedical applications of deep learning in the fields of genetics and genomics, medical informatics focusing on text or speech data, and signal processing of various activities (eg, brain, heart, and human) will further boost the contribution of deep learning in addressing biomedical research problems. As such, although deep learning research in biomedicine has been successful, we believe that there is a need for further exploration, and we expect the results of this study to help researchers and communities better align their present and future work.

Authors' Contributions

SN and YZ designed the study. SN, DK, and WJ analyzed the data. SN took the lead in the writing of the manuscript. YZ supervised and implemented the study. All authors contributed to critical edits and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]
2. Zaharchuk G, Gong E, Wintermark M, Rubin D, Langlotz CP. Deep learning in neuroradiology. AJNR Am J Neuroradiol 2018 Oct;39(10):1776-1784 [FREE Full text] [doi: 10.3174/ajnr.A5543] [Medline: 29419402]

3. Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw 2015 Jan;61:85-117. [doi: 10.1016/j.neunet.2014.09.003] [Medline: 25462637]
4. Litjens G, Kooi T, Bejnordi BE, Setio AA, 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]
5. Dilsizian ME, Siegel EL. Machine meets biology: a primer on artificial intelligence in cardiology and cardiac imaging. Curr Cardiol Rep 2018 Oct 18;20(12):139. [doi: 10.1007/s11886-018-1074-8] [Medline: 30334108]
6. Hu Z, Tang J, Wang Z, Zhang K, Zhang L, Sun Q. Deep learning for image-based cancer detection and diagnosis − a survey. Pattern Recognit 2018 Nov;83:134-149. [doi: 10.1016/j.patcog.2018.05.014]
7. Xue Y, Chen S, Qin J, Liu Y, Huang B, Chen H. Application of deep learning in automated analysis of molecular images in cancer: a survey. Contrast Media Mol Imaging 2017;2017:9512370 [FREE Full text] [doi: 10.1155/2017/9512370] [Medline: 29114182]
8. Meyer P, Noblet V, Mazzara C, Lallement A. Survey on deep learning for radiotherapy. Comput Biol Med 2018 Jul 01;98:126-146. [doi: 10.1016/j.compbiomed.2018.05.018] [Medline: 29787940]
9. Mamoshina P, Vieira A, Putin E, Zhavoronkov A. Applications of deep learning in biomedicine. Mol Pharm 2016 May 02;13(5):1445-1454. [doi: 10.1021/acs.molpharmaceut.5b00982] [Medline: 27007977]
10. Cao C, Liu F, Tan H, Song D, Shu W, Li W, et al. Deep learning and its applications in biomedicine. Genomics Proteomics Bioinformatics 2018 Feb;16(1):17-32 [FREE Full text] [doi: 10.1016/j.gpb.2017.07.003] [Medline: 29522900]
11. Wainberg M, Merico D, Delong A, Frey BJ. Deep learning in biomedicine. Nat Biotechnol 2018 Oct;36(9):829-838. [doi: 10.1038/nbt.4233] [Medline: 30188539]
12. 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]
13. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 2016 Dec 13;316(22):2402-2410. [doi: 10.1001/jama.2016.17216] [Medline: 27898976]
14. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging 2016 May;35(5):1285-1298 [FREE Full text] [doi: 10.1109/TMI.2016.2528162] [Medline: 26886976]
15. de Vos BD, Wolterink JM, de Jong PA, Viergever MA, Išgum I. 2D image classification for 3D anatomy localization: employing deep convolutional neural networks. In: Proceedings Volume 9784, Medical Imaging 2016: Image Processing. 2016 Presented at: SPIE '16; February 27-March 3, 2016; San Diego, CA, USA. [doi: 10.1117/12.2216971]
16. Wang G, Li W, Zuluaga MA, Pratt R, Patel PA, Aertsen M, et al. Interactive medical image segmentation using deep learning with image-specific fine tuning. IEEE Trans Med Imaging 2018 Jul;37(7):1562-1573 [FREE Full text] [doi: 10.1109/TMI.2018.2791721] [Medline: 29969407]
17. Miao S, Wang ZJ, Liao R. A CNN regression approach for real-time 2D/3D registration. IEEE Trans Med Imaging 2016 May;35(5):1352-1363. [doi: 10.1109/TMI.2016.2521800] [Medline: 26829785]
18. de Vos BD, Berendsen FF, Viergever MA, Sokooti H, Staring M, Išgum I. A deep learning framework for unsupervised affine and deformable image registration. Med Image Anal 2019 Feb;52:128-143. [doi: 10.1016/j.media.2018.11.010] [Medline: 30579222]
19. Lin Z, Lanchantin J, Qi Y. MUST-CNN: a multilayer shift-and-stitch deep convolutional architecture for sequence-based protein structure prediction. In: Proceedings of the 13th AAAI Conference on Artificial Intelligence. 2016 Presented at: IAAI '16; February 12-17, 2016; Phoenix, AZ, USA p. 27-34.
20. Wang S, Li W, Liu S, Xu J. RaptorX-Property: a web server for protein structure property prediction. Nucleic Acids Res 2016 Jul 08;44(W1):W430-W435 [FREE Full text] [doi: 10.1093/nar/gkw306] [Medline: 27112573]
21. Alipanahi B, Delong A, Weirauch MT, Frey BJ. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nat Biotechnol 2015 Aug;33(8):831-838. [doi: 10.1038/nbt.3300] [Medline: 26213851]
22. Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning-based sequence model. Nat Methods 2015 Oct;12(10):931-934 [FREE Full text] [doi: 10.1038/nmeth.3547] [Medline: 26301843]
23. Quang D, Xie X. DanQ: a hybrid convolutional and recurrent deep neural network for quantifying the function of DNA sequences. Nucleic Acids Res 2016 Jun 20;44(11):e107 [FREE Full text] [doi: 10.1093/nar/gkw226] [Medline: 27084946]
24. dos Santos BS, Steiner MT, Fenerich AT, Lima RH. Data mining and machine learning techniques applied to public health problems: a bibliometric analysis from 2009 to 2018. Comput Ind Eng 2019 Dec;138:106120. [doi: 10.1016/j.cie.2019.106120]
25. Shukla N, Merigó JM, Lammers T, Miranda L. Half a century of computer methods and programs in biomedicine: a bibliometric analysis from 1970 to 2017. Comput Methods Programs Biomed 2020 Jan;183:105075. [doi: 10.1016/j.cmpb.2019.105075] [Medline: 31526946]
26. Entrez help. Bethesda, MD, USA: National Center for Biotechnology Information (US); 2006.
27. Introduction to MeSH. National Library of Medicine. 2019. URL: https://www.nlm.nih.gov/mesh/introduction.html [accessed 2020-05-25]

28. Chapman D. Advanced search features of PubMed. J Can Acad Child Adolesc Psychiatry 2009 Feb;18(1):58-59 [FREE Full text] [Medline: 19270851]
29. Chang J, Chapman B, Friedberg I, Hamelryck T, de Hoon M, Cock P, et al. Biopython tutorial and cookbook. Biopython. 2020. URL: http://biopython.org/DIST/docs/tutorial/Tutorial.pdf [accessed 2020-05-25]
30. Sugimoto CR, Larivière V. Measuring research: what everyone needs to know. Oxford, UK: Oxford University Press; 2018.
31. Van Eck NJ, Waltman L. Manual for VOSviewer version 1.6.15. VOSviewer. 2020. URL: https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.15.pdf [accessed 2020-05-25]
32. Van Eck NJ, Waltman L. Visualizing bibliometric networks. In: Ding Y, Rousseau R, Wolfram D, editors. Measuring scholarly impact. Berlin, Germany: Springer; 2014:285-320.
33. Waltman L, van Eck NJ, Noyons EC. A unified approach to mapping and clustering of bibliometric networks. J Informetr 2010 Oct;4(4):629-635. [doi: 10.1016/j.joi.2010.07.002]
34. Verborgh R, De Wilde M. Using OpenRefine. Birmingham, UK: Packt Publishing; 2013.
35. Martín-Martín A, Orduna-Malea E, Thelwall M, Delgado López-Cózar E. Google Scholar, Web of Science, and Scopus: a systematic comparison of citations in 252 subject categories. J Informetr 2018 Nov;12(4):1160-1177. [doi: 10.1016/j.joi.2018.09.002]
36. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. In: Proceedings of the 3rd International AAAI Conference on Web and Social Media. 2009 Presented at: ICWSM '09; May 17-20, 2009; San Jose, CA, USA p. 17-20.
37. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech 2008 Oct 09;2008(10):P10008. [doi: 10.1088/1742-5468/2008/10/P10008]
38. Brin S, Page L. The anatomy of a large-scale hypertextual web search engine. Comput Netw 1998 Apr;30(1-7):107-117. [doi: 10.1016/S0169-7552(98)00110-X]
39. Kermany DS, Goldbaum M, Cai W, Valentim CC, Liang H, Baxter SL, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 2018 Feb 22;172(5):1122-31.e9 [FREE Full text] [doi: 10.1016/j.cell.2018.02.010] [Medline: 29474911]
40. 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]
41. Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, Reader study level-Ilevel-II Groups, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol 2018 Aug 01;29(8):1836-1842 [FREE Full text] [doi: 10.1093/annonc/mdy166] [Medline: 29846502]
42. Lao J, Chen Y, Li ZC, Li Q, Zhang J, Liu J, et al. A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. Sci Rep 2017 Sep 04;7(1):10353 [FREE Full text] [doi: 10.1038/s41598-017-10649-8] [Medline: 28871110]
43. Saltz J, Gupta R, Hou L, Kurc T, Singh P, Nguyen V, Cancer Genome Atlas Research Network, et al. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. Cell Rep 2018 Apr 03;23(1):181-93.e7 [FREE Full text] [doi: 10.1016/j.celrep.2018.03.086] [Medline: 29617659]
44. Chen H, Dou Q, Yu L, Qin J, Heng PA. VoxResNet: deep voxelwise residual networks for brain segmentation from 3D MR images. Neuroimage 2018 Apr 15;170:446-455. [doi: 10.1016/j.neuroimage.2017.04.041] [Medline: 28445774]
45. Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nat Biomed Eng 2018 Mar;2(3):158-164. [doi: 10.1038/s41551-018-0195-0] [Medline: 31015713]
46. Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nat Neurosci 2018 Sep;21(9):1281-1289. [doi: 10.1038/s41593-018-0209-y] [Medline: 30127430]
47. Byrne MF, Chapados N, Soudan F, Oertel C, Linares Pérez M, Kelly R, et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. Gut 2019 Jan;68(1):94-100 [FREE Full text] [doi: 10.1136/gutjnl-2017-314547] [Medline: 29066576]
48. Ciompi F, Chung K, van Riel SJ, Setio AA, Gerke PK, Jacobs C, et al. Towards automatic pulmonary nodule management in lung cancer screening with deep learning. Sci Rep 2017 Apr 19;7:46479 [FREE Full text] [doi: 10.1038/srep46479] [Medline: 28422152]
49. Schlegl T, Waldstein SM, Bogunovic H, Endstraßer F, Sadeghipour A, Philip AM, et al. Fully automated detection and quantification of macular fluid in OCT using deep learning. Ophthalmology 2018 Apr;125(4):549-558 [FREE Full text] [doi: 10.1016/j.ophtha.2017.10.031] [Medline: 29224926]
50. Li Z, Wang Y, Yu J, Guo Y, Cao W. Deep learning based radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma. Sci Rep 2017 Jul 14;7(1):5467 [FREE Full text] [doi: 10.1038/s41598-017-05848-2] [Medline: 28710497]

51. Jiménez J, Škalič M, Martínez-Rosell G, De Fabritiis G. K deep: protein-ligand absolute binding affinity prediction via 3d-convolutional neural networks. J Chem Inf Model 2018 Feb 26;58(2):287-296. [doi: 10.1021/acs.jcim.7b00650] [Medline: 29309725]
52. Nie D, Zhang H, Adeli E, Liu L, Shen D. 3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients. Med Image Comput Comput Assist Interv 2016 Oct;9901:212-220 [FREE Full text] [doi: 10.1007/978-3-319-46723-8_25] [Medline: 28149967]
53. Li Z, He Y, Keel S, Meng W, Chang RT, He M. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. Ophthalmology 2018 Aug;125(8):1199-1206. [doi: 10.1016/j.ophtha.2018.01.023] [Medline: 29506863]
54. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. PLoS Med 2018 Nov;15(11):e1002683 [FREE Full text] [doi: 10.1371/journal.pmed.1002683] [Medline: 30399157]
55. Chilamkurthy S, Ghosh R, Tanamala S, Biviji M, Campeau NG, Venugopal VK, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. Lancet 2018 Dec 01;392(10162):2388-2396. [doi: 10.1016/S0140-6736(18)31645-3] [Medline: 30318264]
56. Chang P, Grinband J, Weinberg BD, Bardis M, Khy M, Cadena G, et al. Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas. AJNR Am J Neuroradiol 2018 Jul;39(7):1201-1207 [FREE Full text] [doi: 10.3174/ajnr.A5667] [Medline: 29748206]
57. Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, et al. Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Med 2018 Nov;15(11):e1002686 [FREE Full text] [doi: 10.1371/journal.pmed.1002686] [Medline: 30457988]
58. Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, et al. Fully automated echocardiogram interpretation in clinical practice. Circulation 2018 Oct 16;138(16):1623-1635 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.118.034338] [Medline: 30354459]
59. Ding Y, Sohn JH, Kawczynski MG, Trivedi H, Harnish R, Jenkins NW, et al. A deep learning model to predict a diagnosis of Alzheimer disease by using 18 F-FDG PET of the brain. Radiology 2019 Feb;290(2):456-464 [FREE Full text] [doi: 10.1148/radiol.2018180958] [Medline: 30398430]
60. Han Z, Wei B, Zheng Y, Yin Y, Li K, Li S. Breast cancer multi-classification from histopathological images with structured deep learning model. Sci Rep 2017 Jun 23;7(1):4172 [FREE Full text] [doi: 10.1038/s41598-017-04075-z] [Medline: 28646155]
61. Han SS, Kim MS, Lim W, Park GH, Park I, Chang SE. Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. J Invest Dermatol 2018 Jul;138(7):1529-1538 [FREE Full text] [doi: 10.1016/j.jid.2018.01.028] [Medline: 29428356]
62. González G, Ash SY, Vegas-Sánchez-Ferrero G, Onieva Onieva J, Rahaghi FN, Ross JC, COPDGene and ECLIPSE Investigators. Disease staging and prognosis in smokers using deep learning in chest computed tomography. Am J Respir Crit Care Med 2018 Jan 15;197(2):193-203 [FREE Full text] [doi: 10.1164/rccm.201705-0860OC] [Medline: 28892454]
63. Dolz J, Desrosiers C, Ben Ayed I. 3D fully convolutional networks for subcortical segmentation in MRI: a large-scale study. Neuroimage 2018 Apr 15;170:456-470. [doi: 10.1016/j.neuroimage.2017.04.039] [Medline: 28450139]
64. Wen H, Shi J, Zhang Y, Lu KH, Cao J, Liu Z. Neural encoding and decoding with deep learning for dynamic natural vision. Cereb Cortex 2018 Dec 01;28(12):4136-4160 [FREE Full text] [doi: 10.1093/cercor/bhx268] [Medline: 29059288]
65. Gurovich Y, Hanani Y, Bar O, Nadav G, Fleischer N, Gelbman D, et al. Identifying facial phenotypes of genetic disorders using deep learning. Nat Med 2019 Jan;25(1):60-64. [doi: 10.1038/s41591-018-0279-0] [Medline: 30617323]
66. Falk T, Mai D, Bensch R, Çiçek Ö, Abdulkadir A, Marrakchi Y, et al. U-Net: deep learning for cell counting, detection, and morphometry. Nat Methods 2019 Jan;16(1):67-70. [doi: 10.1038/s41592-018-0261-2] [Medline: 30559429]
67. Ribli D, Horváth A, Unger Z, Pollner P, Csabai I. Detecting and classifying lesions in mammograms with Deep Learning. Sci Rep 2018 Mar 15;8(1):4165 [FREE Full text] [doi: 10.1038/s41598-018-22437-z] [Medline: 29545529]
68. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med 2019 Jun;25(6):954-961. [doi: 10.1038/s41591-019-0447-x] [Medline: 31110349]
69. Song Q, Zhao L, Luo X, Dou X. Using deep learning for classification of lung nodules on computed tomography images. J Healthc Eng 2017;2017:8314740 [FREE Full text] [doi: 10.1155/2017/8314740] [Medline: 29065651]
70. Hosny A, Parmar C, Coroller TP, Grossmann P, Zeleznik R, Kumar A, et al. Deep learning for lung cancer prognostication: a retrospective multi-cohort radiomics study. PLoS Med 2018 Nov;15(11):e1002711 [FREE Full text] [doi: 10.1371/journal.pmed.1002711] [Medline: 30500819]
71. Betancur J, Commandeur F, Motlagh M, Sharir T, Einstein AJ, Bokhari S, et al. Deep learning for prediction of obstructive disease from fast myocardial perfusion SPECT: a multicenter study. JACC Cardiovasc Imaging 2018 Nov;11(11):1654-1663 [FREE Full text] [doi: 10.1016/j.jcmg.2018.01.020] [Medline: 29550305]

72. Wang X, Yang W, Weinreb J, Han J, Li Q, Kong X, et al. Searching for prostate cancer by fully automated magnetic resonance imaging classification: deep learning versus non-deep learning. Sci Rep 2017 Nov 13;7(1):15415 [FREE Full text] [doi: 10.1038/s41598-017-15720-y] [Medline: 29133818]
73. Wang K, Lu X, Zhou H, Gao Y, Zheng J, Tong M, et al. Deep learning Radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study. Gut 2019 Apr;68(4):729-741 [FREE Full text] [doi: 10.1136/gutjnl-2018-316204] [Medline: 29730602]
74. Li X, Chen H, Qi X, Dou Q, Fu CW, Heng PA. H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. IEEE Trans Med Imaging 2018 Dec;37(12):2663-2674. [doi: 10.1109/TMI.2018.2845918] [Medline: 29994201]
75. Grassmann F, Mengelkamp J, Brandl C, Harsch S, Zimmermann ME, Linkohr B, et al. A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. Ophthalmology 2018 Sep;125(9):1410-1420 [FREE Full text] [doi: 10.1016/j.ophtha.2018.02.037] [Medline: 29653860]
76. Cha KH, Hadjiiski L, Chan HP, Weizer AZ, Alva A, Cohan RH, et al. Bladder cancer treatment response assessment in CT using radiomics with deep-learning. Sci Rep 2017 Aug 18;7(1):8738 [FREE Full text] [doi: 10.1038/s41598-017-09315-w] [Medline: 28821822]
77. Arvaniti E, Fricker KS, Moret M, Rupp N, Hermanns T, Fankhauser C, et al. Automated Gleason grading of prostate cancer tissue microarrays via deep learning. Sci Rep 2018 Aug 13;8(1):12054 [FREE Full text] [doi: 10.1038/s41598-018-30535-1] [Medline: 30104757]
78. Jo Y, Park S, Jung J, Yoon J, Joo H, Kim MH, et al. Holographic deep learning for rapid optical screening of anthrax spores. Sci Adv 2017 Aug;3(8):e1700606 [FREE Full text] [doi: 10.1126/sciadv.1700606] [Medline: 28798957]
79. Oakden-Rayner L, Carneiro G, Bessen T, Nascimento JC, Bradley AP, Palmer LJ. Precision radiology: predicting longevity using feature engineering and deep learning methods in a radiomics framework. Sci Rep 2017 May 10;7(1):1648 [FREE Full text] [doi: 10.1038/s41598-017-01931-w] [Medline: 28490744]
80. Nam JG, Park S, Hwang EJ, Lee JH, Jin KN, Lim KY, et al. Development and validation of deep learning-based automatic detection algorithm for malignant pulmonary nodules on chest radiographs. Radiology 2019 Jan;290(1):218-228. [doi: 10.1148/radiol.2018180237] [Medline: 30251934]
81. Chung SW, Han SS, Lee JW, Oh KS, Kim NR, Yoon JP, et al. Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. Acta Orthop 2018 Aug;89(4):468-473 [FREE Full text] [doi: 10.1080/17453674.2018.1453714] [Medline: 29577791]
82. Wang Y, Kosinski M. Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. J Pers Soc Psychol 2018 Feb;114(2):246-257. [doi: 10.1037/pspa0000098] [Medline: 29389215]
83. Zreik M, Lessmann N, van Hamersvelt RW, Wolterink JM, Voskuil M, Viergever MA, et al. Deep learning analysis of the myocardium in coronary CT angiography for identification of patients with functionally significant coronary artery stenosis. Med Image Anal 2018 Feb;44:72-85. [doi: 10.1016/j.media.2017.11.008] [Medline: 29197253]
84. Lindsey R, Daluiski A, Chopra S, Lachapelle A, Mozer M, Sicular S, et al. Deep neural network improves fracture detection by clinicians. Proc Natl Acad Sci U S A 2018 Nov 06;115(45):11591-11596 [FREE Full text] [doi: 10.1073/pnas.1806905115] [Medline: 30348771]
85. Han Y, Kim D. Deep convolutional neural networks for pan-specific peptide-MHC class I binding prediction. BMC Bioinformatics 2017 Dec 28;18(1):585 [FREE Full text] [doi: 10.1186/s12859-017-1997-x] [Medline: 29281985]
86. Jiang H, Ma H, Qian W, Gao M, Li Y, Jiang H, et al. An automatic detection system of lung nodule based on multigroup patch-based deep learning network. IEEE J Biomed Health Inform 2018 Jul;22(4):1227-1237. [doi: 10.1109/JBHI.2017.2725903] [Medline: 28715341]
87. Aubreville M, Knipfer C, Oetter N, Jaremenko C, Rodner E, Denzler J, et al. Automatic classification of cancerous tissue in laserendomicroscopy images of the oral cavity using deep learning. Sci Rep 2017 Sep 20;7(1):11979 [FREE Full text] [doi: 10.1038/s41598-017-12320-8] [Medline: 28931888]
88. Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat Med 2019 Aug;25(8):1301-1309 [FREE Full text] [doi: 10.1038/s41591-019-0508-1] [Medline: 31308507]
89. Pereira TD, Aldarondo DE, Willmore L, Kislin M, Wang SS, Murthy M, et al. Fast animal pose estimation using deep neural networks. Nat Methods 2019 Jan;16(1):117-125 [FREE Full text] [doi: 10.1038/s41592-018-0234-5] [Medline: 30573820]
90. Fu H, Cheng J, Xu Y, Wong DW, Liu J, Cao X. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. IEEE Trans Med Imaging 2018 Jul;37(7):1597-1605. [doi: 10.1109/TMI.2018.2791488] [Medline: 29969410]
91. Bernard O, Lalande A, Zotti C, Cervenansky F, Yang X, Heng PA, et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE Trans Med Imaging 2018 Nov;37(11):2514-2525. [doi: 10.1109/TMI.2018.2837502] [Medline: 29994302]

92. Bien N, Rajpurkar P, Ball RL, Irvin J, Park A, Jones E, et al. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of MRNet. PLoS Med 2018 Nov;15(11):e1002699 [FREE Full text] [doi: 10.1371/journal.pmed.1002699] [Medline: 30481176]
93. Stepniewska-Dziubinska MM, Zielenkiewicz P, Siedlecki P. Development and evaluation of a deep learning model for protein-ligand binding affinity prediction. Bioinformatics 2018 Nov 01;34(21):3666-3674 [FREE Full text] [doi: 10.1093/bioinformatics/bty374] [Medline: 29757353]
94. Sharma K, Rupprecht C, Caroli A, Aparicio MC, Remuzzi A, Baust M, et al. Automatic segmentation of kidneys using deep learning for total kidney volume quantification in autosomal dominant polycystic kidney disease. Sci Rep 2017 May 17;7(1):2049 [FREE Full text] [doi: 10.1038/s41598-017-01779-0] [Medline: 28515418]
95. Liu F, Zhou Z, Samsonov A, Blankenbaker D, Larison W, Kanarek A, et al. Deep learning approach for evaluating knee MR images: achieving high diagnostic performance for cartilage lesion detection. Radiology 2018 Oct;289(1):160-169 [FREE Full text] [doi: 10.1148/radiol.2018172986] [Medline: 30063195]
96. Lehman CD, Yala A, Schuster T, Dontchos B, Bahl M, Swanson K, et al. Mammographic breast density assessment using deep learning: clinical implementation. Radiology 2019 Jan;290(1):52-58. [doi: 10.1148/radiol.2018180694] [Medline: 30325282]
97. Coenen A, Kim YH, Kruk M, Tesche C, De Geer J, Kurata A, et al. Diagnostic accuracy of a machine-learning approach to coronary computed tomographic angiography-based fractional flow reserve: result from the MACHINE consortium. Circ Cardiovasc Imaging 2018 Jun;11(6):e007217. [doi: 10.1161/CIRCIMAGING.117.007217] [Medline: 29914866]
98. Steiner DF, MacDonald R, Liu Y, Truszkowski P, Hipp JD, Gammage C, et al. Impact of deep learning assistance on the histopathologic review of lymph nodes for metastatic breast cancer. Am J Surg Pathol 2018 Dec;42(12):1636-1646 [FREE Full text] [doi: 10.1097/PAS.0000000000001151] [Medline: 30312179]
99. Sullivan DP, Winsnes CF, Åkesson L, Hjelmare M, Wiking M, Schutten R, et al. Deep learning is combined with massive-scale citizen science to improve large-scale image classification. Nat Biotechnol 2018 Oct;36(9):820-828. [doi: 10.1038/nbt.4225] [Medline: 30125267]
100. Chang K, Balachandar N, Lam C, Yi D, Brown J, Beers A, et al. Distributed deep learning networks among institutions for medical imaging. J Am Med Inform Assoc 2018 Aug 01;25(8):945-954 [FREE Full text] [doi: 10.1093/jamia/ocy017] [Medline: 29617797]
101. Han Y, Yoo J, Kim HH, Shin HJ, Sung K, Ye JC. Deep learning with domain adaptation for accelerated projection-reconstruction MR. Magn Reson Med 2018 Sep;80(3):1189-1205. [doi: 10.1002/mrm.27106] [Medline: 29399869]
102. Wang H, Rivenson Y, Jin Y, Wei Z, Gao R, Günaydın H, et al. Deep learning enables cross-modality super-resolution in fluorescence microscopy. Nat Methods 2019 Jan;16(1):103-110 [FREE Full text] [doi: 10.1038/s41592-018-0239-0] [Medline: 30559434]
103. Yang G, Yu S, Dong H, Slabaugh G, Dragotti PL, Ye X, et al. DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. IEEE Trans Med Imaging 2018 Jun;37(6):1310-1321. [doi: 10.1109/TMI.2017.2785879] [Medline: 29870361]
104. Ouyang W, Aristov A, Lelek M, Hao X, Zimmer C. Deep learning massively accelerates super-resolution localization microscopy. Nat Biotechnol 2018 Jun;36(5):460-468. [doi: 10.1038/nbt.4106] [Medline: 29658943]
105. Nie D, Trullo R, Lian J, Wang L, Petitjean C, Ruan S, et al. Medical image synthesis with deep convolutional adversarial networks. IEEE Trans Biomed Eng 2018 Dec;65(12):2720-2730 [FREE Full text] [doi: 10.1109/TBME.2018.2814538] [Medline: 29993445]
106. Zhang Y, Yu H. Convolutional neural network based metal artifact reduction in X-ray computed tomography. IEEE Trans Med Imaging 2018 Jun;37(6):1370-1381 [FREE Full text] [doi: 10.1109/TMI.2018.2823083] [Medline: 29870366]
107. Gong E, Pauly JM, Wintermark M, Zaharchuk G. Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI. J Magn Reson Imaging 2018 Aug;48(2):330-340. [doi: 10.1002/jmri.25970] [Medline: 29437269]
108. Cohen O, Zhu B, Rosen MS. MR fingerprinting deep reconstruction network (DRONE). Magn Reson Med 2018 Sep;80(3):885-894 [FREE Full text] [doi: 10.1002/mrm.27198] [Medline: 29624736]
109. Zhang Y, An L, Xu J, Zhang B, Zheng WJ, Hu M, et al. Enhancing Hi-C data resolution with deep convolutional neural network HiCPlus. Nat Commun 2018 Feb 21;9(1):750 [FREE Full text] [doi: 10.1038/s41467-018-03113-2] [Medline: 29467363]
110. Hyun CM, Kim HP, Lee SM, Lee S, Seo JK. Deep learning for undersampled MRI reconstruction. Phys Med Biol 2018 Jun 25;63(13):135007. [doi: 10.1088/1361-6560/aac71a] [Medline: 29787383]
111. Hauptmann A, Lucka F, Betcke M, Huynh N, Adler J, Cox B, et al. Model-based learning for accelerated, limited-view 3-D photoacoustic tomography. IEEE Trans Med Imaging 2018 Jun;37(6):1382-1393. [doi: 10.1109/TMI.2018.2820382] [Medline: 29870367]
112. Chaudhari AS, Fang Z, Kogan F, Wood J, Stevens KJ, Gibbons EK, et al. Super-resolution musculoskeletal MRI using deep learning. Magn Reson Med 2018 Nov;80(5):2139-2154 [FREE Full text] [doi: 10.1002/mrm.27178] [Medline: 29582464]

113. Chen H, Zhang Y, Chen Y, Zhang J, Zhang W, Sun H, et al. LEARN: learned experts' assessment-based reconstruction network for sparse-data CT. IEEE Trans Med Imaging 2018 Jun;37(6):1333-1347 [FREE Full text] [doi: 10.1109/TMI.2018.2805692] [Medline: 29870363]
114. Popova M, Isayev O, Tropsha A. Deep reinforcement learning for de novo drug design. Sci Adv 2018 Jul;4(7):eaap7885 [FREE Full text] [doi: 10.1126/sciadv.aap7885] [Medline: 30050984]
115. Blaschke T, Olivecrona M, Engkvist O, Bajorath J, Chen H. Application of generative autoencoder in de novo molecular design. Mol Inform 2018 Jan;37(1-2):1700123 [FREE Full text] [doi: 10.1002/minf.201700123] [Medline: 29235269]
116. Zhou J, Theesfeld CL, Yao K, Chen KM, Wong AK, Troyanskaya OG. Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. Nat Genet 2018 Aug;50(8):1171-1179 [FREE Full text] [doi: 10.1038/s41588-018-0160-6] [Medline: 30013180]
117. Putin E, Asadulaev A, Ivanenkov Y, Aladinskiy V, Sanchez-Lengeling B, Aspuru-Guzik A, et al. Reinforced adversarial neural computer for de novo molecular design. J Chem Inf Model 2018 Jun 25;58(6):1194-1204 [FREE Full text] [doi: 10.1021/acs.jcim.7b00690] [Medline: 29762023]
118. Merk D, Friedrich L, Grisoni F, Schneider G. De novo design of bioactive small molecules by artificial intelligence. Mol Inform 2018 Jan;37(1-2):1700153 [FREE Full text] [doi: 10.1002/minf.201700153] [Medline: 29319225]
119. Veltri D, Kamath U, Shehu A. Deep learning improves antimicrobial peptide recognition. Bioinformatics 2018 Aug 15;34(16):2740-2747 [FREE Full text] [doi: 10.1093/bioinformatics/bty179] [Medline: 29590297]
120. Wang S, Sun S, Xu J. Analysis of deep learning methods for blind protein contact prediction in CASP12. Proteins 2018 Mar;86 Suppl 1:67-77 [FREE Full text] [doi: 10.1002/prot.25377] [Medline: 28845538]
121. Chaudhary K, Poirion OB, Lu L, Garmire LX. Deep learning-based multi-omics integration robustly predicts survival in liver cancer. Clin Cancer Res 2018 Mar 15;24(6):1248-1259 [FREE Full text] [doi: 10.1158/1078-0432.CCR-17-0853] [Medline: 28982688]
122. Ma J, Yu MK, Fong S, Ono K, Sage E, Demchak B, et al. Using deep learning to model the hierarchical structure and function of a cell. Nat Methods 2018 Apr;15(4):290-298 [FREE Full text] [doi: 10.1038/nmeth.4627] [Medline: 29505029]
123. Yousefi S, Amrollahi F, Amgad M, Dong C, Lewis JE, Song C, et al. Predicting clinical outcomes from large scale cancer genomic profiles with deep survival models. Sci Rep 2017 Sep 15;7(1):11707 [FREE Full text] [doi: 10.1038/s41598-017-11817-6] [Medline: 28916782]
124. Preuer K, Lewis RP, Hochreiter S, Bender A, Bulusu KC, Klambauer G. DeepSynergy: predicting anti-cancer drug synergy with deep learning. Bioinformatics 2018 May 01;34(9):1538-1546 [FREE Full text] [doi: 10.1093/bioinformatics/btx806] [Medline: 29253077]
125. Liang H, Tsui BY, Ni H, Valentim CC, Baxter SL, Liu G, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nat Med 2019 Mar;25(3):433-438. [doi: 10.1038/s41591-018-0335-9] [Medline: 30742121]
126. Wang L, You ZH, Chen X, Xia SX, Liu F, Yan X, et al. A computational-based method for predicting drug-target interactions by using stacked autoencoder deep neural network. J Comput Biol 2018 Mar;25(3):361-373. [doi: 10.1089/cmb.2017.0135] [Medline: 28891684]
127. Yildirim Ö. A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. Comput Biol Med 2018 May 01;96:189-202. [doi: 10.1016/j.compbiomed.2018.03.016] [Medline: 29614430]
128. Chambon S, Galtier MN, Arnal PJ, Wainrib G, Gramfort A. A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series. IEEE Trans Neural Syst Rehabil Eng 2018 Apr;26(4):758-769. [doi: 10.1109/TNSRE.2018.2813138] [Medline: 29641380]
129. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016 Presented at: CVPR '16; June 27-30, 2016; Las Vegas, NV, USA p. 770-778. [doi: 10.1109/cvpr.2016.90]
130. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016 Presented at: CVPR '16; June 27-30, 2016; Las Vegas, NV, USA p. 2818-2826. [doi: 10.1109/cvpr.2016.308]
131. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015 Presented at: CVPR '15; June 7-12, 2015; Boston, MA, USA p. 1-9. [doi: 10.1109/cvpr.2015.7298594]
132. Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. IEEE Trans Pattern Anal Mach Intell 2017 Apr;39(4):640-651. [doi: 10.1109/TPAMI.2016.2572683] [Medline: 27244717]
133. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: International Conference on 18th Medical Image Computing and Computer-Assisted Intervention. 2015 Presented at: MICCAI '15; October 5-9, 2015; Munich, Germany p. 234-241. [doi: 10.1007/978-3-319-24574-4_28]
134. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: Proceedings of the 27th International Conference on Neural Information Processing Systems. 2014 Presented at: NIPS '14; December 8-13, 2014; Montreal, Canada p. 2672-2680. [doi: 10.1145/3422622]
135. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997 Nov 15;9(8):1735-1780. [doi: 10.1162/neco.1997.9.8.1735] [Medline: 9377276]

136. Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology 2017 Aug;284(2):574-582. [doi: 10.1148/radiol.2017162326] [Medline: 28436741]
137. Chen G, Tsoi A, Xu H, Zheng WJ. Predict effective drug combination by deep belief network and ontology fingerprints. J Biomed Inform 2018 Sep;85:149-154 [FREE Full text] [doi: 10.1016/j.jbi.2018.07.024] [Medline: 30081101]
138. Kim JK, Choi MJ, Lee JS, Hong JH, Kim CS, Seo SI, et al. A deep belief network and dempster-shafer-based multiclassifier for the pathology stage of prostate cancer. J Healthc Eng 2018;2018:4651582 [FREE Full text] [doi: 10.1155/2018/4651582] [Medline: 29755715]
139. Alansary A, Oktay O, Li Y, Folgoc LL, Hou B, Vaillant G, et al. Evaluating reinforcement learning agents for anatomical landmark detection. Med Image Anal 2019 Apr;53:156-164 [FREE Full text] [doi: 10.1016/j.media.2019.02.007] [Medline: 30784956]
140. Shu X, Tang J, Li Z, Lai H, Zhang L, Yan S. Personalized age progression with bi-level aging dictionary learning. IEEE Trans Pattern Anal Mach Intell 2018 Apr;40(4):905-917. [doi: 10.1109/TPAMI.2017.2705122] [Medline: 28534768]
141. Devlin J, Chang MW, Lee K, Toutanova K. Bert: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019 Presented at: NAACL-HLT '19; June 2-7, 2019; Minneapolis, MN, USA p. 4171-4186. [doi: 10.18653/v1/2021.naacl-main.152]
142. Baevski A, Zhou H, Mohamed A, Auli M. wav2vec 2.0: a framework for self-supervised learning of speech representations. In: Proceedings of the 34th Conference on Neural Information Processing Systems. 2020 Dec Presented at: NeurIPS '20; December 6-12, 2020; Vancouver, Canada.
143. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin transformer: hierarchical vision transformer using shifted windows. In: IEEE/CVF International Conference on Computer Vision. 2021 Oct Presented at: ICCV '21; October 10-17, 2021; Montreal, Canada p. 9992-10002. [doi: 10.1109/iccv48922.2021.00986]

Abbreviations

AE: autoencoder
CNN: convolutional neural network
FCNN: fully convolutional neural network
MeSH: Medical Subject Heading
NLP: natural language processing
ResNet: residual neural network
RNN: recurrent neural network
WoS: Web of Science

Edited by A Mavragani; submitted 22.02.21; peer-reviewed by Y Zhao, C Su, Y Zhang; comments to author 17.03.21; revised version received 30.05.21; accepted 20.02.22; published 22.04.22.

Please cite as:
Nam S, Kim D, Jung W, Zhu Y
Understanding the Research Landscape of Deep Learning in Biomedical Science: Scientometric Analysis
J Med Internet Res 2022;24(4):e28114
URL: https://www.jmir.org/2022/4/e28114
doi:10.2196/28114
PMID:35451980

Original Paper

Mechanism of Impact of Big Data Resources on Medical Collaborative Networks From the Perspective of Transaction Efficiency of Medical Services: Survey Study

Junyi Yuan[1], PhD; Sufen Wang[2], PhD; Changqing Pan[3], MD

[1]Information Center, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China

[2]Department of Management Science and Engineering, Glorious Sun School of Business & Management, Donghua University, Shanghai, China

[3]Hospital's Office, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China

Corresponding Author:
Changqing Pan, MD
Hospital's Office
Shanghai Chest Hospital
Shanghai Jiao Tong University
No 241 West Huaihai Road, Xuhui District
Shanghai, 200030
China
Phone: 86 21 62805080
Email: panchangqing@shchest.org

Abstract

Background: The application of big data resources and the development of medical collaborative networks (MCNs) boost each other. However, MCNs are often assumed to be exogenous. How big data resources affect the emergence, development, and evolution of endogenous MCNs has not been well explained.

Objective: This study aimed to explore and understand the influence of the mechanism of a wide range of shared and private big data resources on the transaction efficiency of medical services to reveal the impact of big data resources on the emergence and development of endogenous MCNs.

Methods: This study was conducted by administering a survey questionnaire to information technology staff and medical staff from 132 medical institutions in China. Data from information technology staff and medical staff were integrated. Structural equation modeling was used to test the direct impact of big data resources on transaction efficiency of medical services. For those big data resources that had no direct impact, we analyzed their indirect impact.

Results: Sharing of diagnosis and treatment data (β=.222; *P*=.03) and sharing of medical research data (β=.289; *P*=.04) at the network level (as big data itself) positively directly affected the transaction efficiency of medical services. Network protection of the external link systems (β=.271; *P*=.008) at the level of medical institutions (as big data technology) positively directly affected the transaction efficiency of medical services. Encryption security of web-based data (as big data technology) at the level of medical institutions, medical service capacity available for external use, real-time data of diagnosis and treatment services (as big data itself) at the level of medical institutions, and policies and regulations at the network level indirectly affected the transaction efficiency through network protection of the external link systems at the level of medical institutions.

Conclusions: This study found that big data technology, big data itself, and policy at the network and organizational levels interact with, and influence, each other to form the transaction efficiency of medical services. On the basis of the theory of neoclassical economics, the study highlighted the implications of big data resources for the emergence and development of endogenous MCNs.

(J Med Internet Res 2022;24(4):e32776) doi:10.2196/32776

KEYWORDS

medical collaborative networks; big data resources; transaction efficiency

Introduction

Background

There has been a long-term coexistence of imbalanced allocation and low use efficiency of medical resources in China. Most health care reforms have tried to encourage a variety of medical collaborative practices as a means to improve the quality and efficiency of health care delivery. For example, the New Rural Cooperative Medical Scheme was launched to protect rural households from catastrophic medical expenditure [1] and various medical consortia were mainly used to improve the system of tiered medical services to balance inadequate medical resources [2]. These studies often assumed that medical collaborative networks (MCNs) are exogenous and had already been formed. However, many medical collaborative practices have not achieved the desired results. Su et al [3] showed that there was no statistically significant difference between the distribution of inpatients in county and township hospitals before and after the implementation of the New Rural Cooperative Medical Scheme in China. The practice of collaborative health care will produce various forms of MCNs. The MCNs' structures are always complex [4]. It was corroborated that the MCNs' structures and collaborative practices influence each other [5], the mutual recursive influence becoming meaningful through a complex net of organizational and institutional features, as well as patients' nosological profiles [6]. MCNs are often assumed to be exogenous; however, they are endogenous. It is very important to pay attention to how endogenous MCNs emerge and develop.

At the same time, the development of the internet and big data technology has promoted the transformation of medical service patterns and management modes [7,8], leading to the emergence of various MCNs, such as collaboration between hospitals of different levels [2,9]. Furthermore, many internet companies (such as Hao Daifu, Chunyu Doctor, and Weiyi) have been pouring into the medical service industry to lead more diverse forms of medical collaborative practices [10]. Big data resources in health care have advanced the development of MCNs, which in turn further promotes the application of big data in the health care field [11]. It is generally believed that big data resources affect the emergence and development of MCNs; yet, there is a lack of understanding of the mechanism of the impact of big data resources on the emergence and development of MCNs.

As the organizational network has increasingly become an important form of business operation, the commercial value of information technology (IT) to the organizational network has gradually become an issue of concern. Han et al [12] analyzed the value of the relationship, based on the enterprise resource planning system, between suppliers of the enterprise resource planning system and their partners through case studies. Ceccagnoli et al [13] explored the cocreation of value in a platform ecosystem based on the resource-based view of the firm. These studies have emphasized the organizational privatization of traditional IT resources [14,15] without considering the particularity of big data resources or the coexistence of shared and private resources in the organizational network [16]. The value realization of big data should be analyzed from the work practice, organizational, and supraorganizational levels [17] and be integrated information, technology, policy, and so on [18,19].

This study aims to explore and understand the influence of the mechanism of shared and private big data resources on the emergence and development of MCNs. The coexistence of labor division and cooperation is not only the most basic phenomenon of MCNs, but also the most basic driving force of survival and development. On the basis of neoclassical economics, this paper took the transaction efficiency of medical services as a key variable to represent the emergence and development of endogenous MCNs. Next, we classified big data resources related to value cocreation of MCNs according to two dimensions: (1) public big data resources at the network level versus private big data resources at the medical institution level and (2) the three elements of big data value (data itself, technology, and various organizational elements). At the level of medical institutions in the MCN, there are external web-based big data (health care big data itself) and outward interaction security (big data technology); at the public level of the MCN, there are sharing of big data (health care big data itself) and policies and regulations related to big data (data policy). Finally, we empirically analyzed the direct and intermediary effects of all kinds of big data resources on the transaction efficiency of medical services.

Hypotheses and Modeling

Transaction Efficiency of Medical Services

Medical collaboration refers to a process that occurs when a group of autonomous stakeholders with various medical resources communicate and coordinate with each other to share decision-making, goal setting, and implementation of a plan of care [2,5,6,20].

Extant empirical studies often assumed that MCNs are exogenous and found that medical collaborative practices may be affected by factors at individual, organizational, and system levels, such as mutual trust [20,21], IT infrastructure [22-24], medical policies, investment of public funds [9], and remuneration methods [4]. However, the conclusions drawn regarding the influence of these factors are inconsistent and contradictory [5]. Because of the interaction of many factors, it is necessary to analyze the nature of the impact of these factors on medical collaborative practices from the perspective of system and process [5,6].

From the perspective of system and process, various forms of medical collaborative practices have been explored. Touati et al [6] elicited three specific modalities of collaboration: quasi-inexistent, restrained, and extended. Braun and Cusick [25] explored four innovative care models that aimed to expand access to dental care: expanded coordinated care, colocated care, integrated care, and virtual dental home. Huang and Li [26] divided the medical alliance into three types (compact, semicompact, and loose) according to the closeness of the contact. The recursive interaction between structures and collaborative practices has been corroborated [5] and becomes meaningful through a complex network of organizational and institutional characteristics and the nosological profiles of

patients [6]. However, to explore the influence of the mechanism of big data resources on the emergence and development of MCNs, we need to integrate factor research and structure research to determine a theoretical construct that can reflect the changes in network structure and embody various factors influencing collaborative practices.

The neoclassical economics framework proposed by Yang and Ng [27] studied organizational topological properties by introducing transaction costs. The increase in division of labor will increase the number of transactions, and each transaction will produce transaction costs. If the transaction efficiency is low, the transaction cost is greater than the specialized economy generated by the division of labor and individuals will choose to be self-sufficient. If the transaction efficiency is fully improved, the transaction cost is offset by the specialized economy and individuals will choose division of labor. Therefore, organizational topological properties are closely related to transaction efficiency: the smaller the size of the organization, the more the cooperation with the outside world [28].

MCN members are afforded both cooperation and division of labor. Touati et al [6] emphasized that transaction cost cannot be ignored in all kinds of collaboration involving various factors at individual, organizational, and clinical levels. Collaborative practice requires collaborators to share rules, beliefs, and codes of conduct [5], regarding which there are often differences in the collaborators' cognitions. These differences will incur transaction costs, affecting the results of collaborative practice. McComb et al [20] showed that physicians and nurses in general medical units have different perceptions of role, responsibility, and mutual trust, which act as obstacles to cooperation in these units. Communication problems among collaborators often persist and seriously affect the implementation of collaborative practices. Without videoconferencing, some diagnostic pathways (visual and clinical examination) would be lost in the interaction between cardiologists and family physicians [23]. The traditional written referral usually led to incomplete information, thus affecting the quality and comprehensiveness of communication [24]. There are also some factors at the system level, such as poor public infrastructure [9], that lead to low transaction efficiency and high transaction cost.

Because of the characteristics of autonomy and limited resources, there is division of labor everywhere in MCNs. At the same time, the collaborative community is different from the simple addition of the original individuals and relies on value rationality among members to create a unique social structure oriented to the ultimate goal of common commitment, which can support members to work collaboratively [29]. The decision of whether to choose medical collaborative practice is based on the trade-off between the health care specialized economy and transaction cost. Collaborators make decisions in their own self-interest under a specific MCN, but their decisions are affected by other decision-makers in the MCN. Finally, through the interaction of all parties and the balance of interests, a specific structure will emerge. The MCN's structure and individual decision-making are entangled to produce and reproduce. To sum up, MCNs are endogenous and the transaction efficiency of health care is the key variable for the emergence and development of MCNs.

In this paper, the transaction efficiency of medical services refers to the quality of the medical transaction service. The higher the quality of the transaction service, the smaller the transaction cost and the higher the transaction efficiency. At this time, it is more likely that MCNs will be chosen to provide medical services in a cooperative way.

Big Data Resources

There were 2 main concepts of big data. The first is based on the characteristics of the generated data, such as the 3V model [30], 4V model [31], and 5V model [32]. The second is focused on various technologies and methods such as big data storage and management [33], cloud computing and cloud service [34], big data security and privacy [31], real-time data-processing technology [7], and various big data analysis technologies [35]. De Mauro et al [18] proposed that the four elements (technology, method, information, and impact) that affect the value of big data should be integrated. Wamba et al [19] believed that the business value of big data is enabled through data policy, technology, organizational change, data access, industry structure, and so on. However, the classification of these value factors lack a theoretical basis.

The IT resources of a single organization were often conceptualized and classified based on the resource-based view of the firm [14,15], which emphasized the organizational privatization of resources with a clear definition of property rights. Dover [16] studied the business value of IT based on the relationship theory, expanded the limitations of the resource-based view of the firm on the assumption of ownership and control of resources, and distinguished shared resources from nonshared resources. In network organizations, IT resources (especially big data resources) are both publicly owned by the network and privately owned by a specific organization.

We applied and further extended the classification of IT resources for a single organization [14,15] to that of big data for MCNs and extended the process of realizing IT business value to the process of realizing big data business value. Big data resources for MCNs involve health care big data itself, big data technology, and data policy at both the public level of MCNs and the institution level in MCNs. At the level of medical institutions in the MCN, external web-based big data (health care big data itself) and outward interaction security (big data technology) form the conditions and basis for medical institutions to export or import medical services as decision-makers. At the public level of MCNs, the sharing of big data (health care big data itself) and policies and regulations related to big data (data policy) affect all kinds of support conditions and constraints for the operation of medical institutions in MCNs by forming or changing the public environment at the network level.

At the level of medical institutions in the MCN, external web-based big data resources (big data itself) play a balancing and optimizing role in ensuring the supply of medical service resources to other hospitals or institutions and include real-time data of diagnosis and treatment services and medical service

capacity available for external use. Real-time data of diagnosis and treatment services refers to the degree to which a medical institution provides information on physician suspending the diagnosis and treatment and opening consultations for external systems (such as remote consultation platforms, government public platforms, and medical networking). Real-time data of diagnosis and treatment services are the data source of the catalog of external services provided by medical institutions [36,37]. Medical service capacity available for external use is a medical institution's ability to determine medical service resources such as consultation services and appointment services that can be provided to other hospitals or institutions and can be obtained by comparing the real-time use status of the medical service resources with the ideal status [38,39]. Medical service capacity available for external use is a relevance index of health care big data that reflects the connectivity of health care data [40,41].

As big data technology, outward interaction security at the level of medical institutions provides security for stable and continuous connection of data distributed at different medical institutions. It includes encryption security of web-based data and network protection of external link systems. Encryption security of web-based data is the perceived ability of a medical institution to ensure data security during interaction with other hospitals or institutions [42,43]. Network protection of the external link systems is the perceived ability of a medical institution to deploy the physical security foundation for the connection between medical institutions and the outside world [24,44].

At the public level of MCNs, sharing of big data may improve medical service and research capabilities by sharing health care big data with each other [45]. This includes the sharing of diagnosis and treatment data as well as medical research data. Sharing of diagnosis and treatment data refers to the degree to which a medical institution within MCNs can obtain diagnosis and treatment data from other medical institutions through government public platforms or third-party platforms [41,46]. Sharing of research data refers to the degree to which a medical institution within MCNs can obtain research data from other medical institutions through Chinese National Knowledge Infrastructure, PubMed, and so on. Policies and regulations related to big data at the public level of MCNs refers to the degree to which policies, laws, and regulations (such as 3-level referral from the Health and Family Planning Commission, medical consortium, and regional medical treatment center) can support the construction of the regional medical service platform [19].

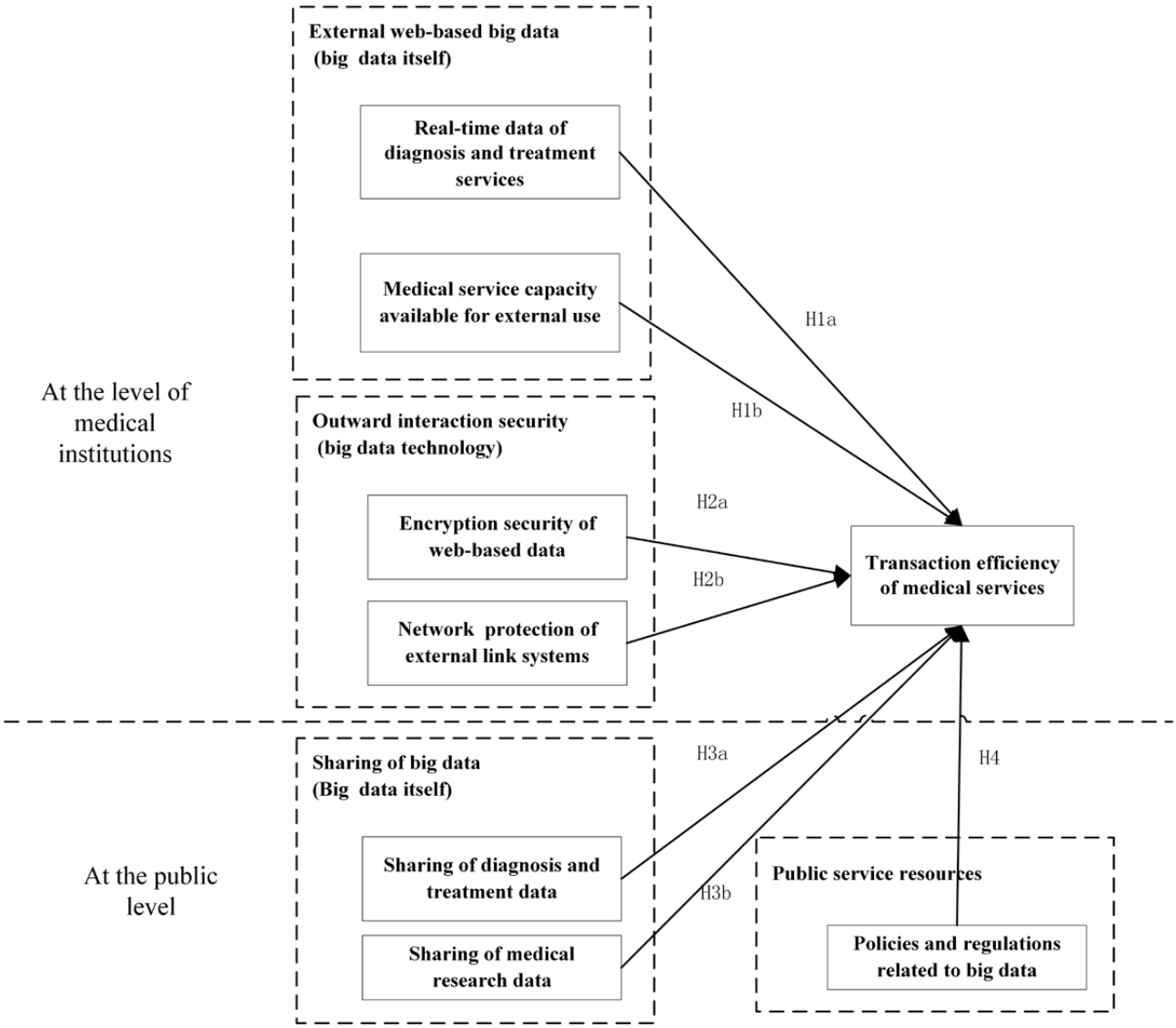
Model

On the basis of the assumption that MCNs are endogenous and that the transaction efficiency of health care is the key variable for the emergence and development of MCNs, this study aims to explore and understand the mechanism of the influence of shared and private big data resources in MCNs on transaction efficiency to reveal the impact of big data resources on the emergence and development of MCNs. The research questions are as follows:

1. What big data resources at the two levels (shared and private) directly affect transaction efficiency?
2. When there is no direct impact, what are the paths of indirect influence of these big data resources on transaction efficiency?

Figure 1 presents the model examined in this research. It shows relationships that are hypothesized to exist among big data resources at the level of medical institutions in the MCN, big data resources at the public level of the MCN, and the transaction efficiency of medical services.

Figure 1. Conceptual model: the impact of big data resources of medical collaborative networks on transaction efficiency of medical services. H: hypothesis.



Methods

Measurement Instruments and Questionnaire Development

Overview

For most constructs, measures validated in previous studies were adapted. For constructs unique to the model, multiple operational measures based on field interviews were developed. All constructs were measured using a 7-point Likert-scale ranging from 1=*strongly disagree* to 7=*strongly agree*. Details of the measures are presented in Multimedia Appendix 1.

Transaction Efficiency of Medical Services

In this study, collaborative medical care was mainly carried out through third-party platforms such as Baiyulan and cloud hospitals. The transaction efficiency of medical services depends on the quantity and quality of medical service resources provided by the platform. The first concerns the scope and level of experts available on the platform. The accuracy, real-time nature, and comprehensiveness of information on the experts enable the requester to know the experts in time, make correct judgments, and reduce unnecessary transaction costs caused by the provision of asymmetric information. The second concerns the performance of the communication mechanism provided by the platform for all partners. To better cooperate with collaborative diagnosis and treatment, the platform needs to support multiple medical institutions to read medical records and images on the web and in real time at the same time to ensure that the image data can be transmitted to the consultation experts without distortion. Inefficient web-based reading will lead to long waiting periods, resulting in uncontrollable average visit time. The smooth reading of required data is not only a powerful guarantee for the rapid completion of services, but also the basis for the continuous demand for collaborative medical services.

On the basis of the studies by DeLone and McLean [47] and Taylor and Todd [48], the transaction efficiency of medical services was measured with 6 items that reflect the extent to which the platform provides reliability, timeliness, and comprehensiveness of expert information, as well as timeliness and stability of communication.

Big Data Resources

To ensure content validity, the measures for most constructs were used, expanded, and modified from the studies by DeLone and McLean [47], Taylor and Todd [48], Bailey and Pearson [49], and Goodhue [50]. For constructs unique to the big data resources for MCNs, items were self-developed.

The real-time data of diagnosis and treatment services were measured by 3 items that reflect the extent of timeliness, accuracy, and accessibility of physician suspending the diagnosis and treatment as well as the opening information provided by medical institutions to the external systems. The medical service capacity available for external use was measured with 4 items that reflect the extent of the ability and accuracy of external consultation and appointment services provided by medical institutions according to the physician's workload. The encryption security of web-based data was measured by 4 items that reflect the extent of effect, convenience, transmission efficiency, and coverage of the encryption and decryption technology used by medical institutions when interacting with external systems. The network protection of external link systems was measured with 4 items that reflect the extent of effect, convenience, satisfaction, and coverage of network protection and application protection deployed by medical institutions.

The sharing of diagnosis and treatment data with other medical institutions was measured by 7 items that reflect the extent of accessibility, accuracy, and integrity of diagnosis and treatment data of other medical institutions, as well as the effect of the data sharing on effectively shortening diagnosis time, avoiding repeated examination, avoiding repeated medication, and avoiding adverse drug-drug reactions. Sharing of medical research data with other medical institutions was measured by 3 items that reflect the extent of convenience, functional completeness, and accuracy of research data provided by other medical institutions. The policies and regulations were measured by 3 items that reflect the extent of rationality, existence, and functional completeness of relevant policies, laws, and regulations supporting the construction of a regional medical service platform.

Data Collection and Demographic Profiles

Data were collected using a survey questionnaire. In China, public hospitals are the main institutions providing health care services. Accordingly, we mainly chose public hospitals, along with some private hospitals. It is very important for medical staff to cooperate closely with IT staff to ensure the implementation of collaborative medical services. Accordingly, each medical institution selected 1 medical staff member and 1 IT staff member as respondents.

The specific data collection plan was designed as follows:

1. Contact the relevant personnel at the target medical institution through WeChat and ask whether they were willing to participate in the survey.
2. Through the relevant personnel, ask the medical institution to determine the respondents, administer the questionnaire on-site, and collect it after completion.
3. If the medical institution is located far away and if the relevant person agrees, provide the questionnaire through WeChat to the person responsible for administering it.

The survey packages were mailed to the appropriate IT executive at each target hospital, with a request that the recipient complete the survey.

The survey packages were also mailed to the appropriate business executive at each target hospital. Part A of the questionnaire was distributed among the appropriate medical staff to complete the measurement items regarding sharing of diagnosis and treatment data, sharing of research data, and transaction efficiency of medical services. Part B was distributed among the appropriate IT executive staff to complete the items related to real-time data of diagnosis and treatment services, medical service capacity available for external use, network protection of external link systems, encryption security of web-based data, and policies and regulations. The questionnaire was administered between August 1, 2017, and October 31, 2017.

Of the 150 medical institutions (involving 18 provinces, autonomous regions, and municipalities) that participated in the survey, 132 (88%) provided valid questionnaires. A total of 264 respondents took part: 132 (50%) IT staff and 132 (50%) medical staff. The sample profile is shown in Table 1.

Table 1. Statistical description of the sample (N=132).

Variables and categories	Values, n (%)
Hospitals	
Hospital level	
Tertiary general hospitals	39 (29.5)
Tertiary specialty hospitals	15 (11.4)
Second-class general hospitals	75 (56.8)
Second-class specialty hospitals	4 (3)
Community hospitals	1 (0.8)
Type of hospital	
Public hospitals	126 (95.5)
General practice	6 (4.5)
IT[a] staff	
Sex	
Male	83 (62.9)
Female	49 (37.1)
Age (years)	
20-30	35 (26.5)
31-40	79 (59.8)
41-50	18 (13.6)
Education	
High school graduate	8 (6.1)
Bachelor’s degree	114 (86.4)
Master’s degree	10 (7.6)
Medical staff	
Sex	
Male	74 (56.1)
Female	58 (43.9)
Age (years)	
20-30	23 (17.4)
31-40	72 (54.5)
41-50	30 (22.7)
51-60	7 (5.3)
Education	
High school graduate	5 (3.8)
Bachelor’s degree	75 (56.8)
Master’s degree	50 (37.9)
Doctorate	2 (1.5)

[a]IT: information technology.

Ethics Approval

The study protocol was reviewed and approved by the ethics review committee at the Shanghai Chest Hospital (IS[P]22003). Before the research was conducted, all participants gave their consent in writing after being informed of the purpose and procedure of the study. We ensured the confidentiality and anonymity of the information collected from the participants.

Data Analysis Process

SmartPLS is a component-based path-modeling software tool based on the partial least squares regression method. We used

SmartPLS (version 2.0) to evaluate the measurement properties and test our hypotheses. Our strategy for data analysis was as follows. First, we evaluated the measurement model by analyzing reliability and validity (including convergent and discriminant validity). Next, applying SmartPLS by using the standard bootstrap resampling procedure (5000 samples) to estimate the significance of the paths, the direct impact of big data resources on transaction efficiency of medical services was examined. For those big data resources that had no direct impact on transaction efficiency, we analyzed their indirect impact.

Results

Reliability and Validity

The measurement model was evaluated using the following criteria:

1. Reliability: The outer loading for the indicator should be ≥0.70 (indicator reliability). The cutoff value for Cronbach α was .70 and that for composite reliability was 0.70 (internal consistency reliability) [51].
2. Validity: The average variance extracted (AVE) should be ≥0.50 (convergent validity), based on the Fornell-Larcker criterion [52] (discriminant validity).

As shown in Table 2, the factor loading values of all items were higher than 0.89 and significant at *P*=.001, with composite reliability value=0.9, above the normal value of 0.7. All values met the minimum requirement for indicator reliability and internal consistency reliability. In addition, the AVE used to assess the convergent validity was >0.70 for all constructs, proving that the model had good convergence validity.

Table 2. Reliability and convergence validity test results.

Constructs and items	Values, mean (SD)	Load value	Composite reliability	Average variance extracted
Encryption security of web-based data			0.970	0.891
ES[a]_1	4.99 (1.532)	0.959		
ES_2	4.95 (1.536)	0.928		
ES_3	5.09 (1.395)	0.951		
ES_4	4.85 (1.515)	0.937		
Network protection of external link systems			0.961	0.862
NP[b]_1	5.72 (1.236)	0.912		
NP_2	5.69 (1.253)	0.941		
NP_3	5.51 (1.224)	0.944		
NP_4	5.47 (1.383)	0.916		
Real-time data of diagnosis and treatment services			0.995	0.983
RT[c]_1	5.15 (1.619)	0.990		
RT_2	5.11 (1.644)	0.995		
RT_3	5.11 (1.611)	0.991		
Medical service capacity available for external use			0.995	0.982
SC[d]_1	4.33 (1.812)	0.991		
SC_2	4.30 (1.788)	0.992		
SC_3	4.44 (1.798)	0.988		
SC_4	4.31 (1.781)	0.993		
Policies and regulations related to big data			0.956	0.879
PR[e]_1	5.5 (1.297)	0.968		
PR_2	5.64 (1.151)	0.919		
PR_3	5.33 (1.292)	0.925		
Sharing of diagnosis and treatment data			0.990	0.931
TS[f]_1	4.4 (1.654)	0.964		
TS_2	4.57 (1.687)	0.950		
TS_3	4.39 (1.681)	0.973		
TS_4	4.54 (1.656)	0.958		
TS_5	4.47 (1.820)	0.972		
TS_6	4.56 (1.715)	0.968		
TS_7	4.45 (1.836)	0.970		
Sharing of medical research data			0.984	0.952
RS[g]_1	4.66 (1.690)	0.966		
RS_2	4.82 (1.587)	0.978		
RS_3	4.79 (1.717)	0.984		
Transaction efficiency of medical services			0.973	0.859
TE[h]_1	4.84 (1.621)	0.937		
TE_2	4.92 (1.574)	0.947		
TE_3	4.89 (1.580)	0.953		
TE_4	4.91 (1.551)	0.925		

Constructs and items	Values, mean (SD)	Load value	Composite reliability	Average variance extracted
TE_5	4.86 (1.528)	0.906		
TE_6	4.95 (1.541)	0.890		

[a]ES: encryption security of web-based data.
[b]NP: network protection of external link systems.
[c]RT: real-time data of diagnosis and treatment services.
[d]SC: medical service capacity available for external use.
[e]PR: policies and regulations.
[f]TS: sharing of diagnosis and treatment data.
[g]RS: sharing of medical research data.
[h]TE: transaction efficiency of medical services.

Table 3 presents the test results of discriminant validity. The square root of the AVE values of each construct were greater than the correlation coefficient between the constructs, which conforms to the Fornell-Larcker criterion [52], proving that the measurement model had good discriminant validity.

Table 3. Discriminant validity test results.

	ES[a]	NP[b]	RT[c]	SC[d]	PR[e]	TS[f]	RS[g]	TE[h]
ES	0.944	—[i]	—	—	—	—	—	—
NP	0.540	0.928	—	—	—	—	—	—
RT	0.475	0.613	0.992	—	—	—	—	—
SC	0.690	0.432	0.615	0.991	—	—	—	—
PR	0.637	0.601	0.527	0.658	0.938	—	—	—
TS	0.359	0.286	0.423	0.417	0.346	0.965	—	—
RS	0.430	0.318	0.508	0.433	0.521	0.698	0.976	—
TE	0.466	0.527	0.554	0.500	0.519	0.581	0.621	0.927

[a]ES: encryption security of web-based data.
[b]NP: network protection of external link systems.
[c]RT: real-time data of diagnosis and treatment services.
[d]SC: medical service capacity available for external use.
[e]PR: policies and regulations.
[f]TS: sharing of diagnosis and treatment data.
[g]RS: sharing of medical research data.
[h]TE: transaction efficiency of medical services.
[i]Not applicable.

Influence Path

Overview

The results of the influence path analysis, including the standardized regression weights and levels of significance, are presented in Table 4 and Figure 2. The coefficient of determination R^2 was used to measure the explained variance of the latent dependent variables compared with the total variance. The cutoff levels were as follows: 0.190, weak; 0.333, moderate; and 0.670, substantial; 55.3% of the variance in transaction efficiency of medical services, 53.3% of the variance in the network protection of external link systems, and 48.7% of the variance in sharing of diagnosis and treatment data were moderately explained, whereas 27.2% of the variance in sharing of medical research data was weakly explained, but met the cutoff level.

Table 4. Direct effect test results.

Hypothesis	Direct path	β coefficient (SE)	*P* value	Support
H1a	RT[a] to TE[b]	.070 (0.121)	.56	Not supported
H1b	SC[c] to TE	.116 (0.123)	.35	Not supported
H2a	ES[d] to TE	−.011 (0.115)	.93	Not supported
H2b	NP[e] to TE	.271 (0.101)	.008	Supported
H3a	TS[f] to TE	.220 (0.105)	.03	Supported
H3b	RS[g] to TE	.289 (0.135)	.04	Supported
H4	PR[h] to TE	.023 (0.118)	.85	Not supported

[a]RT: real-time data of diagnosis and treatment services.

[b]TE: transaction efficiency of medical services.

[c]SC: medical service capacity available for external use.

[d]ES: encryption security of web-based data.

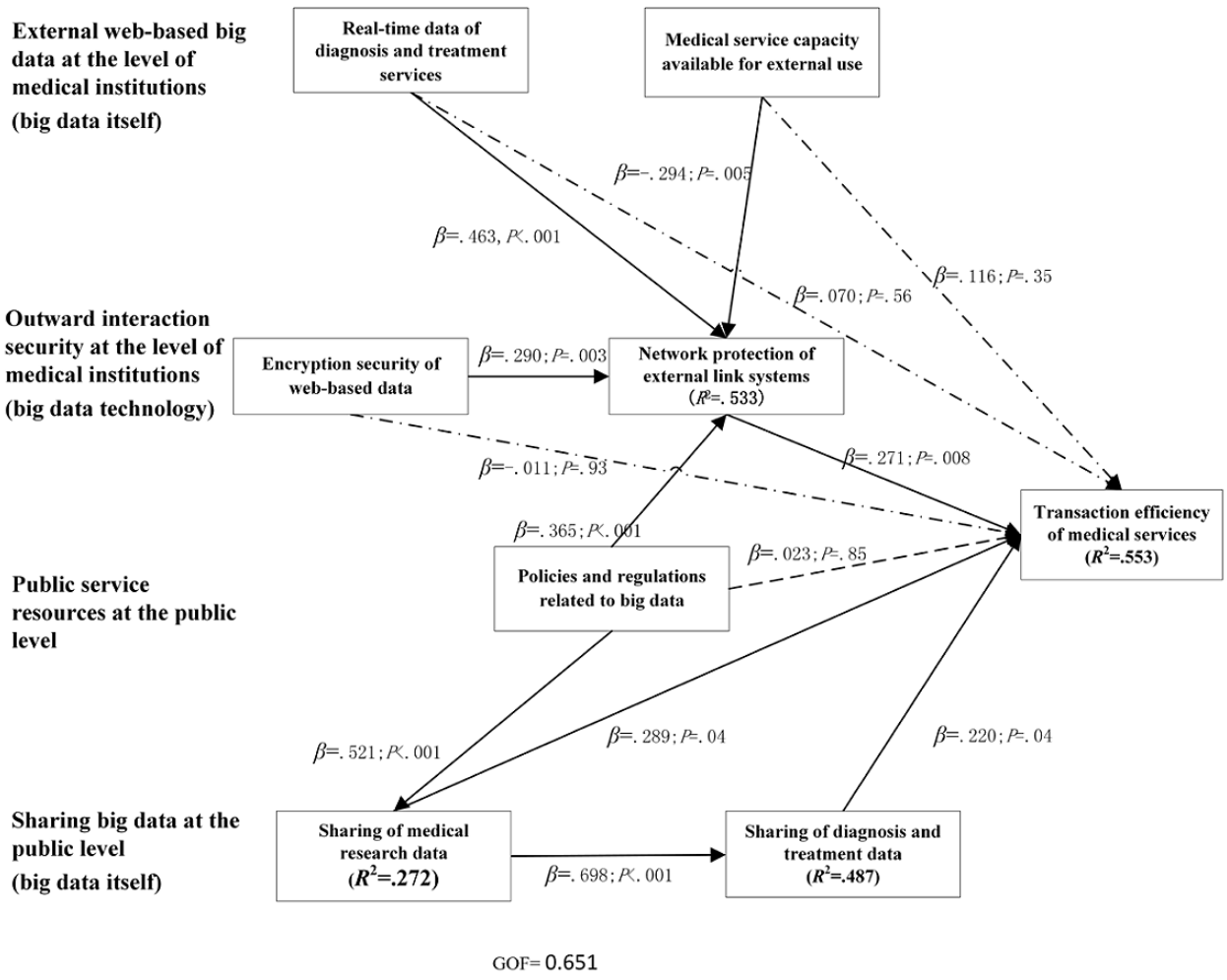
[e]NP: network protection of external link systems.

[f]TS: sharing of diagnosis and treatment data.

[g]RS: sharing of medical research data.

[h]PR: policies and regulations.

Figure 2. Model results, including direct and indirect effects. GOF: goodness of fit.



The model's goodness of fit was our last criterion to assess the overall fit of the model. The model's goodness of fit for this study as calculated was 0.651, which was deemed large [53].

Direct Influence Path

From Figure 2, it can be observed that the direct effects of the network protection of external link systems (β=.271; *P*=.008), sharing of diagnosis and treatment data (β=.220; *P*=.04), and sharing of medical research data (β=.289; *P*=.04) on transaction efficiency of medical services were significant. Hypotheses H2b, H3a, and H3b gained empirical support.

The direct effects of real-time data of diagnosis and treatment services, medical service capacity available for external use, encryption security of web-based data, and policies and regulations on transaction efficiency of medical services were not significant. Hypotheses H1a, H1b, H2a, and H4 did not gain empirical support.

Indirect Influence Analysis

As the encryption security of web-based data, real-time data of diagnosis and treatment services, medical service capacity available for external use, and policies and regulations had no direct impact on transaction efficiency of medical services, the indirect effects of these 4 variables on transaction efficiency of medical services were further analyzed. The results of the mediation test are presented in Table 5 and Figure 2. To assess the magnitude of the indirect effect [54], the variance accounted for (VAF) value was calculated, which represents the relationship between the indirect effect and the total effect.

From Table 5, we can observe the following:

1. The indirect impact of policies and regulations. Although the policies and regulations had no direct impact on transaction efficiency of medical services, there was a completely mediated path (policies and regulations → network protection of external link systems → transaction efficiency of medical services) in which the network protection of external link systems played a mediating role in the effect of policies and regulations on transaction efficiency of medical services (VAF=0.945; *P*=.03). It indicated that the government's establishment of regulations in network security should be conducive to ensuring transaction efficiency and data security.
2. The indirect impact of the encryption security of web-based data. Although the encryption security of web-based data in the external web-based security environment had no direct impact on the transaction efficiency of medical services, there was a completely mediated path (encryption security of web-based data → network protection of external link systems → transaction efficiency of medical services) in which the network protection of external lnk systems played a mediating role in the effect of encryption security of web-based data on transaction efficiency of medical services (VAF=0.879; *P*=.03). It indicated that the encryption security of web-based data improved people's perception of the degree of network protection of external systems and indirectly affected the transaction efficiency of medical services.
3. The indirect impact of real-time data of diagnosis and treatment services. Although the real-time data of diagnosis and treatment services in the external big data analysis environment had no direct impact on the transaction efficiency of medical services, there was a completely mediated path (real-time data of diagnosis and treatment services → network protection of external link systems → transaction efficiency of medical services) in which the network protection of external link systems played a mediating role in the effect of real-time data of diagnosis and treatment services on transaction efficiency of medical services (VAF=0.678; *P*=.02). It indicated that the stronger the ability of internal data extraction, the safer the external data pipeline and the higher the transaction efficiency.
4. The indirect impact of medical service capacity available for external use. Although the medical service capacity available for external use in the external big data analysis environment had no direct impact on the transaction efficiency of medical services, there was a completely mediated path (medical service capacity available for external use → network protection of external link systems → transaction efficiency of medical services) in which there was the indirect effect of medical service capacity available for external use through the network protection of external link systems on transaction efficiency of medical services (VAF=0.391; *P*=.05). From Figure 2, it can be observed that medical service capacity available for external use has a significant direct negative effect on the network protection of external link systems (β=−0.294; *P*=.005), which indicated that frequent service adjustment will increase the complexity of security control and indirectly reduce the transaction efficiency of medical services.
5. The indirect impact of sharing of medical research data. In addition to the direct and significant impact of medical services on the transaction efficiency, there was a partial mediated path (sharing of medical research data → sharing of diagnosis and treatment data → transaction efficiency of medical services) in which sharing of diagnosis and treatment data played a mediating role in the effect of sharing of medical research data on transaction efficiency of medical services (VAF=0.345; *P*=.04). The sharing of research data was conducive to the ability of physicians to interpret the patient's past medical history to issue an accurate diagnosis faster, promote the sharing of diagnosis and treatment data, and indirectly promote transaction efficiency of medical services.

Table 5. Mediation test results.

Indirect effect/direct path	*P* value	Mediated paths	Sobel test		VAF[a]	Type of relationship
			Sobel statistic (SE)	*P* value		
PR[b] to TE[c]			2.170 (0.046)	.03	0.945	Full mediation
PR to TE	.85					
PR to NP[d]	<.001					
NP to TE	.008					
ES[e] to TE			2.122 (0.042)	.03	0.879	Full mediation
ES to TE	.93					
ES to NP	.003					
NP to TE	.008					
RT[f] to TE			2.313 (0.054)	.02	0.678	Full mediation
RT to TE	.56					
RT to NP	<.001					
NP to TE	.008					
SC[g] to TE			−1.958 (0.041)	.05	0.391	Full mediation
SC to TE	.35					
SC to NP	.005					
NP to TE	.008					
RS[h] to TE			2.086 (0.075)	.04	0.345	Partial mediation
RS to TE	.04					
RS to TS[i]	<.001					
TS to TE	.04					

[a]VAF: variance accounted for.

[b]PR: policies and regulations.

[c]TE: transaction efficiency of medical services.

[d]NP: network protection of external link systems.

[e]ES: encryption security of web-based data.

[f]RT: real-time data of diagnosis and treatment services.

[g]SC: medical service capacity available for external use.

[h]RS: sharing of medical research data.

[i]TS: sharing of diagnosis and treatment data.

Discussion

Principal Findings

On the basis of the assumption that MCNs are endogenous and that service transaction efficiency is the key variable for the emergence and development of MCNs, this study empirically analyzed the impact of big data resources of MCNs on the transaction efficiency of health care and provided evidence regarding the following:

1. Sharing of diagnosis and treatment data (big data itself) at the network level directly affected the transaction efficiency of medical services.
   An important challenge of implementing precision medicine based on big data is to share data in MCNs [45]. Sharing diagnosis and treatment data with other hospitals or institutions is an important part of the big data–sharing environment [41]. Only by formulating the classification, grading, and domain-sharing system of medical big data can we steadily promote the opening of medical big data. The sharing of diagnosis and treatment data can result in many obvious benefits, including timely and effective improvement in diagnosis accuracy, strengthening of physician-patient communication and coordination, reduction in repeated treatments, and decrease in the risk of medical errors. By accessing the entire treatment record of the patient through government or third-party platforms, physicians can quickly review the patient's condition, reduce medical expenses, and avoid adverse medical events such as drug-drug interactions and drug contraindications,

thus improving the overall transaction efficiency of medical services.

2. Sharing of research data (big data itself) at the network level directly affected the transaction efficiency of medical services.
   The sharing of research data is another important factor in the overall improvement of medical service quality. Be it clinical effectiveness research, new drug development, or basic medical research, each is often based on the research results of others [40,55]. There are already many shared and free medical research databases such as the electrocardiogram database of the National Institutes of Health, Brain-CODE [43], and Alzheimer disease big data [56] that have advanced related medical research. Integrating the research data of multiple medical institutions is conducive to overcoming the limitations of scientific research and improving the scientific research ability of physicians. With the advent of the era of precision medicine, more and more knowledge-sharing methods have come into being, which has promoted the improvement of multidisciplinary diagnosis and treatment ability and improved the transaction efficiency of medical services.
3. Network protection of external link systems (big data technologies) at the level of medical institutions directly affected the transaction efficiency of medical services.
   Outward interaction security (big data technologies) at the level of medical institutions provides a safe and efficient web-based environment in which a medical institution can be connected with other hospitals or institutions and exchange data. To connect data distributed in different medical institutions steadily and continuously, the first thing to address is the security problem [7,31,57].
   In the past, medical institutions only needed to pay attention to the security of the internal network, which was basically isolated from the outside world. The local area network had high security but poor interoperability. With the development of the internet and big data, the applications of telemedicine are changing rapidly [23,24] and medical institutions are facing increasing need for connections to other hospitals or institutions. The network protection of an outreach system is an important security guarantee for contact between medical institutions and the outside world. Network protection must take into account both security and efficiency, and it should not reduce the efficiency and availability of facilities while ensuring the security of data exchanged by external systems. Abbasi et al [58] point out that through a secure and stable link, the activities of the cooperating parties in the network can be more closely linked and the transaction is more efficient.
4. Real-time data of diagnosis and treatment services (big data itself), medical service capacity available for external use (big data itself), encryption security of web-based data (big data technologies) at the level of medical institutions, and policies and regulations at the network level indirectly affected the transaction efficiency of medical services through network protection of the outreach system (big data technology) at the level of medical institutions. These 4 big data resources will affect the perception of physicians regarding the deployment of a physical security foundation for the connection between medical institutions and the outside world [24,44]. These results highlight that big data technology, big data, and policy at the network and organizational levels interact with, and influence, each other to form the service transaction efficiency of various MCNs.

Theoretical Implications

This study contributes to research in 3 ways. First, we highlighted the important role of service transaction efficiency in MCN research. Prior research has largely emphasized that service transaction efficiency is one of the factors that affect the operation effect of specific MCNs [6]. In these studies, it was often assumed that MCNs are exogenous and that there is an absolute standard for the quality of MCNs. But this paper emphasized that an MCN is not exogenous; rather, many factors are responsible for its emergence and development. On the basis of the theory of neoclassical economics [27], this study took service transaction efficiency as the key variable for the emergence and development of MCNs and connected the 2 perspectives of factor-oriented research and process-oriented research in current collaborative medical research. From the perspective of MCN being endogenous, the foothold of the study was not the absolute quality of the MCN but the fitness of the MCN to the specific environment. On the basis of transaction efficiency, the study provided the basis for future research on the emergence and development of MCNs. This logic may help explain why there are various contradictions in prior studies on the factors responsible.

Second, we conceptualized big data resources oriented to MCNs from the network and medical institution levels, including big data itself, big data technology, and policy. The combination of big data resources at the level of medical institutions in the MCN and the network public level of the MCN thus affected the transaction efficiency of medical services as a key variable for the emergence and development of MCNs. It emphasized the coexistence and intertwined influence of public big data resources of MCNs and private big data resources in MCNs. This study expanded the limitation of the existing IT-enabling value based on the resource-based view of the firm, which emphasized the private and exclusive nature of IT resources. It also corresponded to the call for research on analyzing the value realization of big data from the work practice, organizational, and supraorganizational levels [17].

Third, this study provided empirical support for De Mauro et al [18] and Wamba et al [19], who proposed integrating big data technology, big data itself, and policy to realize the value of big data. The results further refined and enriched this insight to reveal the detailed impact path of big data technologies, big data itself, and policies on transaction efficiency of medical services. Big data itself was divided into the network level and the organizational level. Big data assets at the network level have a direct impact on transaction efficiency of medical services. However, big data assets at the organizational level affected the transaction efficiency by affecting people's perception of outward interaction security technology at the organizational level. The negative impact of medical service capacity available for external use on network protection of external link systems indicated that an increase in external

services would make people develop a great sense of insecurity. Policies and regulations related to big data at the public level cannot directly affect the services' transaction efficiency, but they affected the overall formation and operation of MCNs by affecting the public big data resources and the perception of outward interaction security technology at the organizational level.

Practical Implications

The results have several implications for practice. This study provided the corresponding theoretical guidance for the government to formulate policies. The government should specify corresponding strategies to develop policies regarding sharing of big data resources at the public level and promote various institutions to strengthen the security of external collaborative networks. These policies will affect the ecological service environment of an MCN's operation to improve transaction efficiency and ultimately enhance the development of MCNs. In addition, all kinds of medical institutions that are willing to interact with the outside world to form an MCN must first strengthen network security, which can especially balance the negative effects caused by the increase in external collaborative services.

Study Limitations

This study includes several limitations. The data collection was based on the convenient sampling method. Although the medical institutions covered were basically in line with the relative proportion of public and private hospitals in China's medical institutions, the selection of regions was based on the principle of convenient sampling. Furthermore, this study only considered the transaction efficiency of medical services to reveal the impact of big data resources on the emergence and development of MCNs. In fact, other variables, such as the learning cost of medical services, can affect the emergence and development of MCNs. Future research can analyze the impact of big data resources on the emergence and development of MCNs from the perspective of the learning cost of medical services.

Conclusions

Our study contributes to both theory and practice. First, it focused on the effects of big data resources on the transaction efficiency of medical services and highlighted how MCNs emerge and develop. Second, it theorized that there are two levels of big data resources—network level and medical institution level—and highlighted the intertwined effect of public and private big data resources on transaction efficiency (including direct impact and intermediary impact). Third, it focused on the effects of health care big data itself, big data technology, and policy on transaction efficiency and revealed the interaction and influence mechanism of these 3 elements of big data value as well as their impact on the formation and development of MCNs.

Acknowledgments

This work was funded by the Information Development Project of the Shanghai Economic and Information Commission (202002009).

Authors' Contributions

JY and SW were responsible for the study conception and design. JY was responsible for the acquisition of data. SW analyzed and interpreted the data. SW and JY drafted the manuscript, and CP was responsible for its critical revision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire items.
[DOCX File , 16 KB - jmir_v24i4e32776_app1.docx]

References

1. Li L. New cooperative medical scheme and medical expenditure in rural China. Pac Econ Rev 2017 Jul 09;24(1):46-68. [doi: 10.1111/1468-0106.12228]
2. Yang F, Yang Y, Liao Z. Evaluation and analysis for Chinese medical alliance's governance structure modes based on Preker-harding model. Int J Integr Care 2020 Nov 24;20(4):14 [FREE Full text] [doi: 10.5334/ijic.5417] [Medline: 33281526]
3. Su D, Chen Y, Gao H, Li H, Chang J, Lei S, et al. Is there a difference in the utilisation of inpatient services between two typical payment methods of health insurance? Evidence from the new rural cooperative medical scheme in China. Int J Environ Res Public Health 2019 Apr 19;16(8):1410 [FREE Full text] [doi: 10.3390/ijerph16081410] [Medline: 31010133]
4. Dutta M, Sun K. Complex structures: meaning formation amid China's new rural cooperative medical scheme. Int J Commun 2017:1618-1634 [FREE Full text]
5. Touati N, Rodríguez C, Paquette M, Denis J. Understanding the emergence and development of medical collaboration across organizational boundaries: a longitudinal case study. Health (London) 2018 Nov;22(6):580-602. [doi: 10.1177/1363459317724852] [Medline: 28795577]

6. Touati N, Rodríguez C, Paquette M, Maillet L, Denis J. Professional role identity: at the heart of medical collaboration across organisational boundaries. Int J Integr Care 2019 Apr 02;19(2):1 [FREE Full text] [doi: 10.5334/ijic.4184] [Medline: 30971867]
7. Yaqoob I, Hashem IA, Gani A, Mokhtar S, Ahmed E, Anuar NB, et al. Big data: from beginning to future. Int J Inf Manag 2016 Dec;36(6):1231-1247. [doi: 10.1016/j.ijinfomgt.2016.07.009]
8. Borges do Nascimento IJ, Marcolino MS, Abdulazeem HM, Weerasekara I, Azzopardi-Muscat N, Gonçalves MA, et al. Impact of big data analytics on people's health: overview of systematic reviews and recommendations for future studies. J Med Internet Res 2021 Apr 13;23(4):e27275 [FREE Full text] [doi: 10.2196/27275] [Medline: 33847586]
9. Zeng Y, Li J, Yuan Z, Fang Y. The effect of China's new cooperative medical scheme on health expenditures among the rural elderly. Int J Equity Health 2019 Feb 06;18(1):27 [FREE Full text] [doi: 10.1186/s12939-019-0933-2] [Medline: 30728018]
10. China iunternet + health care white paper report. Kantar Consulting. URL: https://mp.weixin.qq.com/s/Ix-h_vQ_EpNfqGP5W3mu1w [accessed 2022-04-09]
11. van Elten HJ, Sülz S, van Raaij EM, Wehrens R. Big data health care innovations: performance dashboarding as a process of collective sensemaking. J Med Internet Res 2022 Feb 22;24(2):e30201 [FREE Full text] [doi: 10.2196/30201] [Medline: 35191847]
12. Han K, Oh W, Im KS, Chang RM, Oh H, Pinsonneault A. Value cocreation and wealth spillover in open innovation alliances. MIS Q 2012;36(1):291-351. [doi: 10.2307/41410418]
13. Ceccagnoli M, Forman C, Huang P, Wu DJ. Cocreation of value in a platform ecosystem! The case of enterprise software. MIS Q 2012;36(1):263-290. [doi: 10.2307/41410417]
14. Banker RD, Bardhan IR, Chang H, Lin L. Plant information systems, manufacturing capabilities, and plant performance. MIS Q 2006;30(2):315-337. [doi: 10.2307/25148733]
15. Wade M, Hulland J. Review: the resource-based view and information systems research: review, extension, and suggestions for future research. MIS Q 2004;28(1):107-142. [doi: 10.2307/25148626]
16. Lavie D. The competitive advantage of interconnected firms: an extension of the resource-based view. Acad Manag Rev 2006 Jul;31(3):638-658. [doi: 10.5465/amr.2006.21318922]
17. Günther WA, Rezazade Mehrizi MH, Huysman M, Feldberg F. Debating big data: a literature review on realizing value from big data. J Strat Inf Syst 2017 Sep;26(3):191-209. [doi: 10.1016/j.jsis.2017.07.003]
18. De Mauro A, Greco M, Grimaldi M. What is big data? A consensual definition and a review of key research topics. AIP Conference Proc 2015 Sep 05;1644(1):97. [doi: 10.1063/1.4907823]
19. Fosso Wamba S, Akter S, Edwards A, Chopin G, Gnanzou D. How ‘big data’ can make big impact: findings from a systematic review and a longitudinal case study. Int J Production Econ 2015 Jul;165:234-246. [doi: 10.1016/j.ijpe.2014.12.031]
20. McComb SA, Lemaster M, Henneman EA, Hinchey KT. An evaluation of shared mental models and mutual trust on general medical units: implications for collaboration, teamwork, and patient safety. J Patient Saf 2017 Dec;13(4):237-242. [doi: 10.1097/PTS.0000000000000151] [Medline: 25706910]
21. Bussu S, Marshall M. (Dis)Integrated care? Lessons from East London. Int J Integr Care 2020 Oct 19;20(4):2 [FREE Full text] [doi: 10.5334/ijic.5432] [Medline: 33177965]
22. Ding WY, Beresford MW, Saleem MA, Ramanan AV. Big data and stratified medicine: what does it mean for children? Arch Dis Child 2019 Apr;104(4):389-394. [doi: 10.1136/archdischild-2018-315125] [Medline: 30266876]
23. Nicolini D. Stretching out and expanding work practices in time and space: the case of telemedicine. Human Relations 2016 Apr 22;60(6):889-920. [doi: 10.1177/0018726707080080]
24. Lundvoll Nilsen L. Collaborative work by using videoconferencing: opportunities for learning in daily medical practice. Qual Health Res 2011 Aug;21(8):1147-1158. [doi: 10.1177/1049732311405683] [Medline: 21483025]
25. Braun PA, Cusick A. Collaboration between medical providers and dental hygienists in pediatric health care. J Evid Based Dent Pract 2016 Jun;16 Suppl:59-67. [doi: 10.1016/j.jebdp.2016.01.017] [Medline: 27236997]
26. Huang E, Li W. Evaluation on public health reform models in China. J Health Econ 2010;9:5-12. [doi: 10.14055/j.cnki.33-1056/f.2010.09.002.html]
27. Yang X, Ng Y. Specialization and Economic Organization A New Classical Microeconomic Framework. Amsterdam: North-Holland; 1993:1-40.
28. Liu P, Yang X. The theory of irrelevance of the size of the firm. J Econ Behav Org 2000 Jun;42(2):145-165. [doi: 10.1016/s0167-2681(00)00083-4]
29. Adler PS, Kwon S, Heckscher C. Perspective—professional work: the emergence of collaborative community. Org Sci 2008 Apr;19(2):359-376. [doi: 10.1287/orsc.1070.0293]
30. Laney D. 3-D data management: controlling data volume, velocity and variety. META Group Research Note. URL: http://www.researchgate.net/publication/304929258_3-D_Data_Management_Controlling_Data_Volume_Velocity_and_Variety [accessed 2022-04-09]
31. Terzi D, Terzi R, Sagiroglu S. A survey on security and privacy issues in big data. In: Proceedings of the 2015 10th International Conference for Internet Technology and Secured Transactions (ICITST). 2015 Presented at: 2015 10th

International Conference for Internet Technology and Secured Transactions (ICITST); Dec 14-16, 2015; London, UK. [doi: 10.1109/icitst.2015.7412089]
32. Kuo M, Sahama T, Kushniruk A, Borycki E, Grunwell D. Health big data analytics: current perspectives, challenges and potential solutions. Int J Big Data Intell 2014;1(1/2):114-126 [FREE Full text] [doi: 10.1504/ijbdi.2014.063835]
33. Zhang H, Chen G, Ooi BC, Tan K, Zhang M. In-memory big data management and processing: a survey. IEEE Trans Knowl Data Eng 2015 Jul 1;27(7):1920-1948. [doi: 10.1109/tkde.2015.2427795]
34. Hashem IA, Yaqoob I, Anuar NB, Mokhtar S, Gani A, Ullah Khan S. The rise of “big data” on cloud computing: review and open research issues. Inf Syst 2015 Jan;47:98-115. [doi: 10.1016/j.is.2014.07.006]
35. Saggi MK, Jain S. A survey towards an integration of big data analytics to big insights for value-creation. Inf Process Manag 2018 Sep;54(5):758-790. [doi: 10.1016/j.ipm.2018.01.010]
36. Aktepe A, Turker AK, Ersoz S. Internet based intelligent hospital appointment system. Intell Automation Soft Comput 2014 Sep 16;21(2):135-146. [doi: 10.1080/10798587.2014.949094]
37. Ariyaluran Habeeb RA, Nasaruddin F, Gani A, Targio Hashem IA, Ahmed E, Imran M. Real-time big data processing for anomaly detection: a survey. Int J Inf Manag 2019 Apr;45:289-307. [doi: 10.1016/j.ijinfomgt.2018.08.006]
38. Wang Y, Kung L, Byrd TA. Big data analytics: understanding its capabilities and potential benefits for healthcare organizations. Technol Forecast Soc Change 2018 Jan;126:3-13. [doi: 10.1016/j.techfore.2015.12.019]
39. Wang Y, Kung L, Wang WY, Cegielski CG. An integrated big data analytics-enabled transformation model: application to health care. Inf Manag 2018 Jan;55(1):64-79. [doi: 10.1016/j.im.2017.04.001]
40. Al-Harazi O, El Allali A, Colak D. Biomolecular databases and subnetwork identification approaches of interest to big data community: an expert review. OMICS 2019 Mar;23(3):138-151. [doi: 10.1089/omi.2018.0205] [Medline: 30883301]
41. Bardram JE, Houben S. Collaborative affordances of medical records. Comput Support Coop Work 2017 Nov 20;27(1):1-36. [doi: 10.1007/s10606-017-9298-5]
42. Chen W, Chen Z, Cui F. Collaborative and secure transmission of medical data applied to mobile healthcare. Biomed Eng Online 2019 May 20;18(1):60 [FREE Full text] [doi: 10.1186/s12938-019-0674-x] [Medline: 31109320]
43. Manogaran G, Thota C, Lopez D, Sundarasekar R. Big data security intelligence for healthcare industry 4.0. In: Cybersecurity for Industry 4.0. Cham: Springer; 2017:103-126.
44. Saheb T, Izadi L. Paradigm of IoT big data analytics in the healthcare industry: a review of scientific literature and mapping of research trends. Telematics Informatics 2019 Aug;41:70-85. [doi: 10.1016/j.tele.2019.03.005]
45. Hulsen T, Jamuar SS, Moody AR, Karnes JH, Varga O, Hedensted S, et al. From big data to precision medicine. Front Med (Lausanne) 2019 Mar 1;6:34 [FREE Full text] [doi: 10.3389/fmed.2019.00034] [Medline: 30881956]
46. Lefaivre S, Behan B, Vaccarino A, Evans K, Dharsee M, Gee T, et al. Big data needs big governance: best practices from brain-CODE, the Ontario-brain institute's neuroinformatics platform. Front Genet 2019 Mar 29;10:191 [FREE Full text] [doi: 10.3389/fgene.2019.00191] [Medline: 30984233]
47. Delone WH, McLean ER. The DeLone and McLean model of information systems success: a ten-year update. J Manag Inf Syst 2014 Dec 23;19(4):9-30. [doi: 10.1080/07421222.2003.11045748]
48. Taylor S, Todd PA. Understanding information technology usage: a test of competing models. Inf Syst Res 1995 Jun;6(2):144-176. [doi: 10.1287/isre.6.2.144]
49. Bailey JE, Pearson SW. Development of a tool for measuring and analyzing computer user satisfaction. Manag Sci 1983 May;29(5):530-545. [doi: 10.1287/mnsc.29.5.530]
50. Goodhue DL. Development and measurement validity of a task-technology fit instrument for user evaluations of information system. Decision Sci 1998 Jan;29(1):105-138. [doi: 10.1111/j.1540-5915.1998.tb01346.x]
51. Al-Fraihat D, Joy M, Masa'deh R, Sinclair J. Evaluating E-learning systems success: an empirical study. Comput Human Behav 2020 Jan;102:67-86. [doi: 10.1016/j.chb.2019.08.004]
52. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Market Res 2018 Nov 28;18(1):39-50. [doi: 10.1177/002224378101800104]
53. Wetzels M, Odekerken-Schröder G, van Oppen C. Using PLS path modeling for assessing hierarchical construct models: guidelines and empirical illustration. MIS Q 2009;33(1):177. [doi: 10.2307/20650284]
54. Helm S, Eggert A, Garnefeld I. Modeling the impact of corporate reputation on customer satisfaction and loyalty using partial least squares. In: Handbook of Partial Least Squares. Berlin, Heidelberg: Springer; 2010:515-534.
55. Tobert CM, Mott SL, Nepple KG. Malnutrition diagnosis during adult inpatient hospitalizations: analysis of a multi-institutional collaborative database of academic medical centers. J Acad Nutr Diet 2018 Jan;118(1):125-131. [doi: 10.1016/j.jand.2016.12.019] [Medline: 28416434]
56. Zhang R, Simon G, Yu F. Advancing Alzheimer's research: a review of big data promises. Int J Med Inform 2017 Oct;106:48-56 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.07.002] [Medline: 28870383]
57. Chen H, Chiang RH, Storey VC. Business intelligence and analytics: from big data to big impact. MIS Q 2012;36(4):1165-1188. [doi: 10.2307/41703503]
58. Abbasi A, Sarker S, Chiang R. Big data research in information systems: toward an inclusive research agenda. J Assoc Inf Syst 2016 Feb;17(2):I-XXXII. [doi: 10.17705/1jais.00423]

Abbreviations

AVE: average variance extracted
IT: information technology
MCN: medical collaborative network
VAF: variance accounted for

Edited by R Kukafka; submitted 10.08.21; peer-reviewed by H Shi, P Seth; comments to author 23.10.21; revised version received 16.03.22; accepted 21.03.22; published 21.04.22.

Please cite as:
Yuan J, Wang S, Pan C
Mechanism of Impact of Big Data Resources on Medical Collaborative Networks From the Perspective of Transaction Efficiency of Medical Services: Survey Study
J Med Internet Res 2022;24(4):e32776
URL: https://www.jmir.org/2022/4/e32776
doi:10.2196/32776
PMID:35318187

Original Paper

Effectiveness of a Pharmacist-Led Web-Based Medication Adherence Tool With Patient-Centered Communication: Results of a Clustered Randomized Controlled Trial

Jan van Lieshout[1*], MD, PhD; Joyca Lacroix[2*], PhD; Aart van Halteren[3*], PhD; Martina Teichert[1,4*], PhD

[1]Scientific Center for Quality of Healthcare (IQ healthcare), Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, Netherlands

[2]Department of Brain, Behavior & Cognition, Philips Research, Eindhoven, Netherlands

[3]Department of Chronic Disease Management, Philips Research, Eindhoven, Netherlands

[4]Department of Clinical Pharmacy and Toxicology, Leiden University Medical Centre, Leiden, Netherlands

[*]all authors contributed equally

Corresponding Author:
Jan van Lieshout, MD, PhD
Scientific Center for Quality of Healthcare (IQ healthcare)
Radboud Institute for Health Sciences
Radboud University Medical Center
P.O. Box 9101
Nijmegen, 6500HB
Netherlands
Phone: 31 024 361 53 05
Email: jan.vanlieshout@radboudumc.nl

Abstract

Background: Growing numbers of people use medication for chronic conditions; nonadherence is common, leading to poor disease control. A web-based tool to identify an increased risk for nonadherence with related potential individual barriers might facilitate tailored interventions and improve adherence.

Objective: This study aims to assess the effectiveness of a newly developed tool aimed at improving medication adherence.

Methods: We performed a cluster randomized controlled trial in patients initiating cardiovascular or oral blood glucose–lowering medication. Participants were recruited from community pharmacies. They completed an online questionnaire comprising assessments of their risk for medication nonadherence and subsequently of barriers to adherence. In pharmacies belonging to the intervention group, individual barriers displayed in a graphical profile on a tablet were discussed by pharmacists and patients with high nonadherence risk in face-to-face meetings and shared with their general practitioners and practice nurses. Tailored interventions were initiated by pharmacists. Barriers of control patients were not presented nor discussed and these patients received usual care. The primary outcome was the effectiveness of the intervention on medication adherence at 8 months' follow-up between patients with an increased nonadherence risk from the intervention and control groups, calculated from dispensing data.

Results: Data from 492 participants in 15 community pharmacies were available for analyses (intervention 253, 7 pharmacies; control 239, 8 pharmacies). The intervention had no effect on medication adherence (B=–0.01; 95% CI –0.59 to 0.57; *P*=.96), nor in the post hoc per-protocol analysis (B=0.19; 95% CI –0.50 to 0.89; *P*=.58).

Conclusions: This study showed no effectiveness of a risk stratification and tailored intervention addressing personal barriers for medication adherence. Various potential explanations for lack of effectiveness were identified. These explanations relate, for instance, to high medication adherence in the control group, study power, and fidelity. Process evaluation should elicit possible improvements and inform the redesign of intervention and implementation.

Trial Registration: The Netherlands National Trial Register NTR5186; https://tinyurl.com/5d8w99hk

(J Med Internet Res 2022;24(4):e16141) doi:10.2196/16141

KEYWORDS

medication adherence; improvement; intervention; web-based; tailored intervention; patient centered; barriers; primary care; cardiovascular diseases; diabetes

Introduction

Adherence to chronic medication is problematic, leading to poor disease control with a burden on patients' quality of life and health care systems [1]. Studies show that 17%-80% of patients with a chronic condition were not adherent, especially in asymptomatic conditions [2-6]. Various causes can hamper adherence and additionally, adherence varies between types of diseases and within patients over time [4,7,8]. Vrijens et al [9] made a taxonomy of nonadherence based on phases in the process of medication use: initiation, implementation, and persistence.

The multifaceted nature of the adherence problem illustrates that improving adherence needs interventions that are tailored to the individual patient [7,10]. Accordingly, recent high-quality randomized controlled trials in a systematic review on interventions for enhancing medication adherence tailored their interventions. These methods of improving medication adherence for chronic health problems were mostly complex and lacked effectivity [11]. In an overview of systematic reviews, Ryan et al [12] found self-monitoring of medicines and self-management programs to be generally effective. Simplified dosage regimens and pharmacists involvement in medication reviews and pharmaceutical care services on adherence involving patient education on good medication use were considered promising.

Systematic reviews showed that the pharmacist can play an important role in the prevention of cardiovascular diseases, mainly through patient education and counseling, drug safety management, medication reviews, monitoring and reconciliation, detection and control of risk factors, and clinical outcomes [13,14]. Community pharmacies are a natural location for patient recruitment at first dispensing to target patients with an increased risk for nonadherence and to perform a tailored intervention at the second dispensing moment.

In recent reviews nonadherence was reported to be high for cardiovascular and oral blood glucose–lowering medication [15-18]. In several studies, the risk for nonadherence was shown to be the highest in the first year after the start of chronic medication [19,20]. Consequently, interventions to warrant adherence are potentially most effective at the initiation of a chronic medication treatment.

At present, there is no tool that combines selection of those patients who are at risk for nonadherence and assessment of their individual barriers for good medication use in combination with offering tailored interventions by care professionals to overcome individual barriers. We have now developed a user-friendly medication adherence tool comprising a nonadherence risk and barrier assessment in an online patient questionnaire, pharmacists' equipment to perform a tailored intervention based on a graphic barrier profile, and the intervention itself.

The primary research question reported in this paper was: What is the effectiveness of using the medication adherence tool on medication adherence of patients starting with cardiovascular or oral blood glucose–lowering medication identified as being at high risk for nonadherence at 8 months' follow-up compared with usual care? Medication adherence was measured by pharmacy dispensing data. Furthermore, we assessed predictive values of the medication nonadherence risk assessment and the barrier questionnaire. Parallel to the effectiveness evaluation, we performed a process evaluation. In this publication we only report on the effectiveness evaluation.

Methods

Study Design

This was a cluster randomized trial with an intervention group of pharmacies (using the medication adherence tool) and a control group of pharmacies (providing usual care). The study design is explained in detail in the study protocol [21]. The patient inclusion period was from 2015 to 2017.

Study Setting

In the Netherlands, the vast majority of the patient groups included (diabetes, cardiovascular risk management) is treated in primary care. General practices provide cardiovascular risk management and diabetes care, generally supported by care groups.

In the Netherlands, at the start of chronic medication patients with a first dispensing usually receive medication for 2 weeks. A second dispensing after 2 weeks is intended to assess first patient experiences with the drug. Only when the patient is willing to use the drug chronically, a follow-up dispensing for mostly 90 days is supplied. Consequently, patients starting with the study medication (intervention group) were expected to have a second dispensing after 2 weeks. The follow-up dispensing of chronic medication is expected to take place every 3 months.

The Medication Adherence Tool

Overview

The medication adherence tool developed was based on the emotional, cognitive, and practical components in nonadherence. It comprised 3 elements: a patient questionnaire, pharmacy equipment, and the tailored intervention. All patients filled the questionnaires; however, the intervention and control groups differed in patient information from the questionnaire available to the pharmacist at the second dispensing, additional pharmacy training, and recommendations to address potential risk and barriers by a tailored intervention.

Patient Questionnaire

The online questionnaire consisted of 2 parts. First, the Probabilistic Medication Adherence Scale (ProMAS) measuring the nonadherence risk. The ProMAS is an 18-item validated questionnaire to assess nonadherence behavior in general [22].

One question from the original questionnaire was excluded because the validation study results showed that it had a substantially lower model fit than the other questions [22]. Two other questions needed to be excluded for those patients who did not already use medication chronically at that point of time. To know whether that was the case for the specific patient, we preceded the ProMAS with a question on whether the patient was on medication already. If answered with "no," the 15-item questionnaire was presented excluding the questions about longer medication use; otherwise the 17-item questionnaire was presented. Patients were invited to participate in the study at the first dispense. Participating patients received the questionnaire shortly before the second dispense. If it appeared that they had not answered the questionnaire at the second dispensing moment, they were offered to answer the questionnaire in the pharmacy. Consequently, patient's answers evaluated their experiences in medication adherence for at least 10 days.

The items consist of statements with yes/no answer categories. Examples of statements are "It has happened at least once that I forgot to take (one of) my medicines"; and "When I am away from home, I occasionally do not take (one of) my medicines." The items skipped related to making changes in the medication use and being late for refills, as these questions consider a longer experience in chronic medication use. One original question considering longer medication use ("In the past month, I have forgotten to take my medication at least once") was adapted to "Since the first dispensing, I have forgotten to take my medication at least once."

Second, the barrier questionnaire measured to what extent the following emotional, cognitive, and practical barriers for nonadherence were present: feelings with regard to medication (emotional), fear of side effects (emotional), concerns about medication usage (emotional and cognitive), necessity beliefs (cognitive), attitude with regard to medication (cognitive), self-efficacy (cognitive and practical), inconvenience (practical), and applying the medication scheme (practical). The barrier questionnaire was developed and validated in an earlier study (data not shown). First, a list of 25 adherence determinants was composed, based on existing evidence in the literature about their relationship with medication nonadherence [8,23-27]. Second, an extensive questionnaire was composed by assessing this list of 25 determinants making use of existing (shorter versions of) instruments for the constructs for which they existed and self-composed items when instruments did not exist. Third, the questionnaire was administered to 1247 patients taking medication for their chronic condition in the Netherlands and United Kingdom. Furthermore, their medication adherence was assessed through pharmacy refills. Finally, based on predictive modeling, the determinants that were significantly predictive of medication nonadherence were selected and the less predictive ones were excluded. This resulted in a short and manageable questionnaire that included 24 items that screen for a set of 8 determinants (named barriers) that have been shown to have a significant impact on medication nonadherence. The questionnaire entails a screening for the potential presence of the barriers in the patient rather than a validated assessment. The result is verified in the conversation between the patient and the care provider. Examples of items and answering categories are as follows:

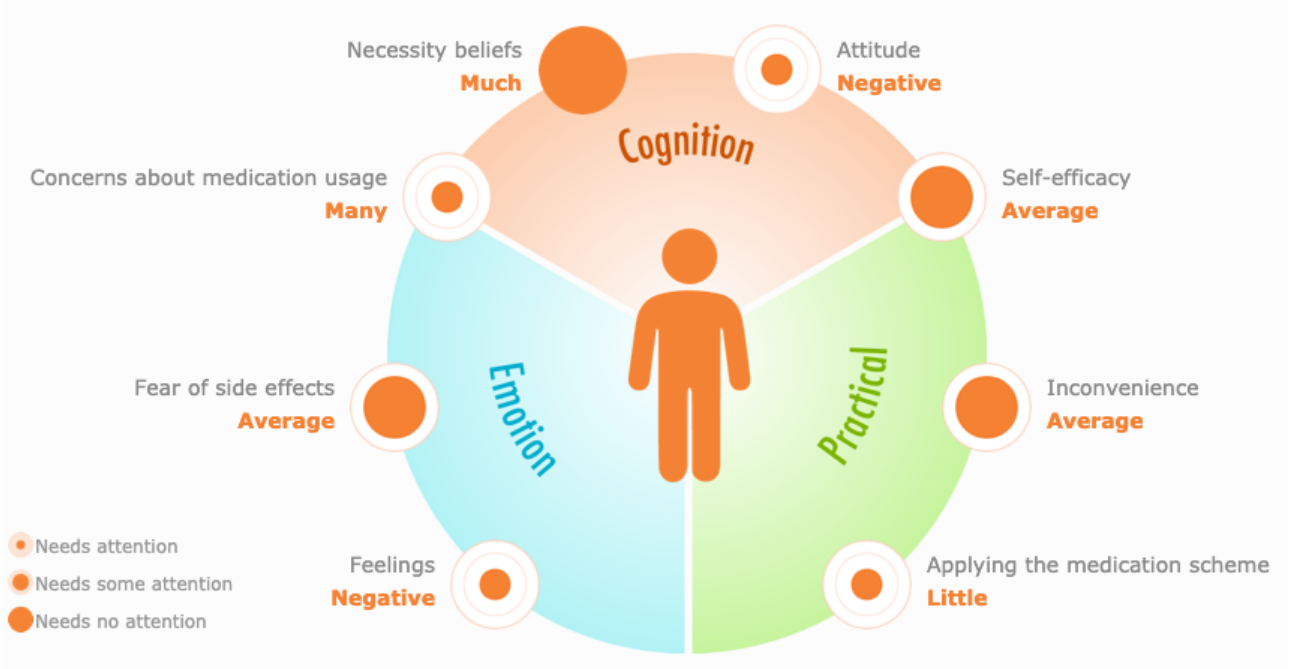
- In the domain self-efficacy: "If I do my best, I will succeed in taking my medicines according to my doctor's prescriptions." Answering categories: Strongly disagree/ Disagree/ Uncertain/ Agree/ Strongly agree.
- In the domain attitude: "How positive or negative are you about your prescribed medication?" Possible answers: Negative/ Somewhat negative/ Neutral/ Somewhat positive/ Positive.
- In the domain feelings: "I feel that I would rather stay away from my medication." Answering categories: Strongly disagree/ Disagree/ Uncertain/ Agree/ Strongly agree (scored reversed).

Barrier Profile

The answers to the barrier questionnaire were translated into a visual barrier profile that presents each domain as a circle: a small circle corresponding to a barrier "asking for much attention"; a larger circle corresponding to "asking for some attention"; and the largest circle corresponding to "no barrier present" (Figure 1). The visual representation of the profile deliberately showed the largest circle when no barrier was found, to emphasize a patient's strengths (full circle) and represent barriers as opportunities for growth (from small to larger circles).

The profile shows 8 potential adherence barriers: 2 emotional barriers (*feelings with regard to medication* and *fear of side effects*), 1 emotional/cognitive barrier (*concerns about medication usage*), 2 cognitive barriers (*necessity beliefs* and *attitude with regard to medication*), 1 cognitive/practical barrier (*self-efficacy*), and 2 practical barriers (*inconvenience* and *applying the medication scheme*).

Figure 1. An example of a barrier profile.



Pharmacy Equipment

Intervention pharmacists and their pharmacy assistants received a 3-hour training comprising an introduction to patient-centered motivational communication, a demonstration of a second dispense discussion with a barriers profile, and skills practice using role play. The training was provided by a psychologist with expertise in patient-centered communication technique. The psychologist instructed the intervention pharmacists and their pharmacy assistants on discussing the profile with the patients to tailor the intervention to overcome personally relevant cognitive, emotional, or practical barriers.

Intervention pharmacies received a manual containing instructions for discussing and overcoming the various potential barriers as reference material at their disposal when they would need it. This manual was developed based on principles of patient-centered communication and with experts' input. The manual held recommendations on how to address each of the possible barriers. In the developmental phase a stakeholder group was involved. This group comprised pharmacists, general practitioners (GPs), a communication expert, technicians, and researchers.

Each pharmacy was provided with a tablet that had an app (BOMM) installed, which was specially developed for registration of participating patients.

It also offered the opportunity to the patient to fill out the questionnaire at the second dispensing as a second chance to those who wanted to participate but failed to do so before. In the intervention pharmacies, the pharmacist used the tablet to review the graphical barrier profile during the second dispensing moment, and to make any notes regarding the conversation and applied intervention. Control pharmacies did not receive the information from the questionnaires, and their consultation during the second dispense was performed as usual.

Tailored Intervention

At the second dispensing moment, a tailored intervention was initiated for patients in the intervention group with an increased risk for nonadherence. The intervention started with a presentation and discussion of the barriers profile. In a personal face-to-face consultation, the pharmacist discussed the relevant barriers in the profile trying to take away or diminish these barriers using a patient-centered communication technique. The pharmacist could use the manual as a reference source to rely on for addressing the relevant barriers (the manual contained specific instructions for addressing each barrier). To address, for instance, the cognitive barrier of necessity beliefs, the intervention, for example, focused on emphasizing the necessity of the medication or stimulating the patient to start self-monitoring of blood pressure or blood glucose levels to make the effect of the medication more visible to him/her and thereby improve the belief that it is necessary to take the medication. Self-monitoring devices were made available by the research group for patients to use during the study. To overcome the practical barrier of a lack of applying the medication scheme, the pharmacist could give additional explanation about the medication scheme or a pill organizer or multidose drug dispensing systems; in some cases, a simplification of the dosage scheme could also be offered. The pharmacist could plan a follow-up consultation.

The pharmacist registered the intervention type in the automated information system, which is shared with the pharmacy team and the GP. This enabled the health care professionals in the general practice to take notice of the intervention and to pay attention to the adherence in line with the pharmacists' intervention.

The feasibility of implementing the use of the adherence tool in the daily pharmacy workflow was piloted in 2 intervention pharmacies. Based on the experiences during the pilot period, workflow adaptations to increase convenience and efficiency

were made. Study information for patients and pharmacists, comprising clarification for the patients and flowcharts for the pharmacies, was adapted after the pilot period based on feedback from the pilot pharmacies.

Implementation of the Medication Adherence Tool

Members of the research team visited all participating pharmacists to explain the study in detail and the use of the app on the tablet and to provide them with an easy-to-understand explanation of the workflow concept. They subsequently also discussed how the workflow concept could be optimally tailored to existing routine work procedures in their pharmacy. One of the project group members from the care group was available during working hours to answer questions arising from the study. Finally, pharmacists had the opportunity to contact one of the pilot pharmacists for questions and advice about study procedures.

Furthermore, intervention pharmacies received a follow-up group session from the trainer on the patient-centered communication technique to provide support and exchange experiences on discussing the barriers profiles.

Recruitment of Pharmacies, General Practices, and Patients

Drug Classes and ATC Codes

The trial was carried out with community pharmacists, GPs, and their patients with a first prescription of cardiovascular or oral blood glucose–lowering medication with Anatomic Therapeutical Chemical (ATC) codes A10B, B01AC, C01A, C01D, C03, C07, C08, C09, or C10 within the study period (see Multimedia Appendix 1 for drug classes and ATC codes and [28]). A first prescription was defined as no drug dispensing from that drug class to the patient in the preceding year. In the Netherlands, patients are listed with a GP. Furthermore, patients generally receive their medication from 1 community pharmacy.

The primary care collaborative DOH (De Ondernemende Huisarts or The Innovative General Practitioner) recruited the pharmacies and general practices for this study. DOH is a general practice collaborative in the South of the Netherlands that developed and implemented structured care for several prevalent chronic conditions including cardiovascular diseases and diabetes. The DOH works in close collaboration with the regional pharmacist organization Stichting Categorale Zorg voor apothekers in Zuidoost Brabant (CaZo). During the trial period patient inclusion failed compared with the expected numbers. For that reason we allowed pharmacies to also include patients listed in general practices not part of the DOH collaborative.

Pharmacies

DOH closely cooperates with 25 pharmacies and invited all for the study. In a pilot phase, 2 pharmacies tested the use of the medication adherence tool in their daily practice; for that reason they participated in the intervention group. The other participating pharmacies were randomly assigned to the intervention or control group by drawing lots (performed by an independent research assistant), and were informed about their assignment. To ensure that small and large pharmacies were evenly spread over the intervention and control groups, pharmacies were stratified by “size,” dichotomized as “large” or “small” (ie, the number of registered patients from DOH GPs).

General Practices

DOH general practices received written information about the study before the start. They were encouraged to contact a member of the research team for posing their questions in case things were unclear.

Furthermore, they were offered waiting room materials including information on the study for narrowcasting, to raise awareness about the study, and thus enhance patient inclusion. During the study period, GPs were informed about the study progress in terms of number of participants recruited. The barrier profile and the pharmacists’ notes on the intervention were available to the GPs and practice nurses in the electronic chain system, so they could build on this information related to these patients in the intervention group to increase medication adherence in their follow-up contacts with the patient.

Patients

Patients (>18 years) from a DOH GP with a first dispensing of cardiovascular or oral blood glucose–lowering medication (ATC codes mentioned above), without cognitive impairments, and able to read and speak Dutch were eligible for inclusion. The pharmacies’ automatized computer system alerts the pharmacist at first dispensing of any drug at the ATC level 5 (eg, simvastatin). However, the inclusion criterion for this study was more strict in defining a first prescription within a drug class (ATC level 3, eg, statins). Consequently, upon receiving an alert for a first dispensing, the pharmacist had to check whether the prescription was related to the medication studied and whether this dispensing concerned the first drug that the patient received within the drug class during the last year. Pharmacists informed eligible patients about the study, provided them with an information package that also contained the informed consent form, and invited them to participate. Those interested in participating were registered. We included only those patients who returned a signed informed consent form. One week after the first dispensing, these patients received a link to the online questionnaire by email or a paper-based questionnaire if preferred.

For all patients starting with cardiovascular or oral blood glucose–lowering medication with a first prescription from a DOH GP, the GP starts a treatment plan in the electronic information system (referred to as chain information system) that GPs and pharmacies use in their current daily practice to report about the condition and laboratory results of their patients and this is accessible to the different caregivers in the chain who are involved in the care of the patient (eg, GP, pharmacist, nurse practitioner). For patients in intervention pharmacies with a high nonadherence risk, the pharmacists could access the graphical barrier profile (see Figure 1 for an example) in the chain system and in the app on the tablet by entering the credentials of the patient.

Ethical Approval and Informed Consent

The study was reviewed and approved by the local Medical Research Ethics Committee, the Commissie Mensgebonden Onderzoek (CMO) region Arnhem-Nijmegen (registration number 2015-1604).

During their pharmacy visit for a first dispense, eligible patients were invited to participate and received a package with study information and an informed consent form. Additionally, we asked informed consent from participating pharmacies and GPs to share data regarding the condition and the medication use of participating patients.

Measures

Nonadherence Risk

The revised ProMAS questionnaire consisted of 15 or 17 questions depending on the use of chronic medication prior to the study. Each question on the ProMAS has 2 answer categories, leading to a maximum sum score of 15 or 17. In the 18-item ProMAS the cut-off value for being at risk for nonadherence was 14. As we skipped 1 question, we pragmatically lowered the cut-off value to 13: patients with a score of ≤13 were classified as having an increased risk for nonadherence, whereas those with a score above 13 were not. We used the same cut-off for patients who had to answer only 15 questions to minimize the probability of excluding patients that might have an increased risk for nonadherence. The revised ProMAS was applied 2 weeks after treatment initiation (baseline) and at 8 months' follow-up, both in the intervention and in the control group. In the follow-up measurement all patients received the 17-item revised ProMAS questionnaire. The revised ProMAS score at baseline signaled nonadherence risk and showed whether the patient was eligible for an intervention while the control group data were used to assess the predictive values of the revised ProMAS. The effectiveness of the intervention on the revised ProMAS score at follow-up was a secondary outcome measure.

Barriers to Adherence

Similar to the nonadherence risk, barriers were assessed at inclusion and follow-up, and in both the intervention group and the control group. The result at baseline informed the infographic and consequently the intervention in those patients in the intervention group that had a revised ProMAS score indicative of a high nonadherence risk. Data at inclusion in the control group were used to assess the predictive values of the barrier questionnaire. The results at follow-up were used to assess the effectiveness of the intervention on the barriers.

Adherence

Medication adherence was calculated as the percentage of days covered (PDC) by medication based on pharmacy dispensing data of 1 drug group from therapy initiation to follow-up after 8 months. The denominator of the PDC was the number of days in 8 months from the first dispensing. For the numerator of the PDC we counted the days covered by medication. Gaps in availability due to late follow-up dispensing led to a lower PDC.

We used the PDC in a dichotomized way and as a continuous measure. Applying Haynes's empirical definition of adequate adherence (ie, at least 80% of drugs taken) to antihypertensive medication, patients with a PDC of at least 80% were labeled as "adherent." Although it may depend on the specific medication in use, this cut-off point is commonly used in the literature as a critical value for nonadherence [29,30]. As second main outcome, we assessed the effectiveness of the intervention on the PDC as continuous measure.

PDCs were calculated per medication group (Multimedia Appendix 1), except for blood glucose–lowering drugs, which were calculated at the subclass level, for example, biguanides (ATC code A10BA) and sulfonylurea derivatives (ATC code A10BB).

The follow-up period was set at 8 months to include at least two follow-up prescriptions covering 3 months each after the first dispensing for 2 weeks. Within participating pharmacies information on the PDC of included patients was received in an anonymized way for intervention and control pharmacies.

Outcomes

Primary Outcome

The primary outcome was the effectiveness of the intervention on medication adherence (PDC dichotomized as ≥80% of days covered and PDC as a continuous measure) comparing intervention and control group patients after an 8-month follow-up.

Secondary Outcomes

- The difference in percentage of patients with an increased risk for nonadherence based on the revised ProMAS between the intervention and control groups after an 8-month follow-up.
- The effectiveness of the intervention on the composite barrier score (See the "Data Analysis" section).
- The positive and negative predictive values of (1) the revised ProMAS score and (2) the barriers profile measured at baseline in the control group in relation to medication adherence at 8 months' follow-up.

In the study protocol we formulated a secondary outcome relating to medication adherence in the subgroup of patients with a follow-up period of at least one year. As data might be available easily, this would be a way to study whether an effect would sustain. We refrained from this outcome because we found no effectiveness in the primary outcome after 8 months.

Sample Size

For the sample size calculation we assumed that 60% of the patients at high risk in our sample, based on their revised ProMAS score, would be nonadherent (defined as PDC <80%), with a 20% increase in patients with a PDC ≥80% in the intervention group compared with the control group [31]. Concerning the effect of clustering of patients within pharmacies, we assumed an intracluster correlation of 0.05.

For this trial in the care group setting, we expected at least 14 community pharmacies to participate. The sample size calculation indicated that 39 patients at high risk for nonadherence are needed per pharmacy (power 80%, type 1 error 5%; PASS software version 11).

Data Analysis

To assess differences between the groups, we performed linear and logistic mixed model multilevel analyses, with adjustment for potential confounders (patient age, gender, diagnosis [diabetes or not]). We planned to control for the number of comedication in chronic use, but lacked the data for this.

We compared differences in medication adherence between the intervention and control groups both as a dichotomous (PDC <80% versus PDC ≥80%) and as a continuous outcome measure. We performed intention-to-treat analyses.

During the study period data from the process evaluation (eg, patient interviews) showed that the intervention was not always applied. Post hoc we performed the same analyses per protocol. From the intervention group we included only those patients from whom we had proof that they actually received an intervention (based on a note about the intervention in the electronic chain system or from questionnaire or interview data from the process evaluation) and compared these with the patients in the control group.

During analysis we discovered much higher adherence rates in the control condition than anticipated. We therefore tested whether the dichotomized adherence differed between all patients starting research medication in control pharmacies and our research sample in the control group that was part of this larger group of patients. Adherence was based on the PDC, assessed in the same way as for the study population. We used a chi-square test.

We compared differences in the revised ProMAS score between the intervention and control groups by a fixed cut-off point (score ≤13 versus >13) and the mean score between the groups at follow-up.

Further secondary analyses assessed the difference in the composite barrier profile score between the intervention and control groups. For this outcome we computed a composite barrier score based on the profile presentation. Each barrier was scored as 1 (indicative for serious barrier), 2 (possible barrier asking for some attention), or 3 (no barrier). We added the 8 barrier scores to form the composite barrier score with a range from 8 to 24, with a higher score indicating fewer barriers.

The predictive values of the revised ProMAS score were computed based on cross tabulation of revised ProMAS and PDC scores dichotomized. All these analyses were performed using SPSS software (version 25; IBM Corp).

Finally, the predictive value of the barrier profile was assessed using machine learning, performed in R [32,33]. The barrier profile consists of 8 individual barrier scores that were used to train a machine learning algorithm to predict adherence at 8 months. The benefit of using a machine learning technique is that combinations of the individual barrier scores that are indicative of nonadherence will be discovered during the training phase of the machine learning model. To create a machine-learned predictive model, the data set was split into training, test, and validation sets. The training set was used for learning the parameters of the predictive model. The test set was used to tune the parameters of a predictive model and the validation set to evaluate the performance of the predictive model. Two-thirds of the sample was used as the test and training sets and one-third for validation.

For evaluating the performance of the predictive model, a 10-fold cross validation was applied [34]. To find the most appropriate machine learning technique for our data, 3 techniques were tested: random forest, kernel support vector machines, and generalized linear models [35,36].

Results

Overview

A total of 15 community pharmacies participated (7 pharmacies in the intervention group and 8 in the control group).

In total, pharmacies registered 1405 patients for the study. Of them, 806 completed the first questionnaire and returned a signed informed consent form. Of these 806 patients, pharmacy data were available for 684 patients. We had to exclude 192 patients as they turned out to be no initiators of their drug group according to study criteria (but switchers or restarters) or because they did not start chronic medication from our predefined groups and thus were not eligible for inclusion. So, for analyses we finally had available data from 492 patients. In the intervention group 129/253 patients (51.0%) had a revised ProMAS score, indicating high nonadherence risk; in the control group this concerned 115/239 patients (48.1%). Table 1 describes the sample in terms of the basic patient characteristics; the intervention group and the control group showed no important differences. The questionnaire at 8-month follow-up was filled in by 370 of the 492 patients. Patient inclusion numbers per pharmacy varied from 3 to 107 patients.

Table 1. Patient characteristics.

Characteristics	Intervention group (n=253)	Control group (n=239)
Female, n (%)	127 (50.2)	107 (44.8)
Age, mean (SD)	63.7 (10.8)	62.5 (10.6)
On chronic medication before starting the study medication, n (%)	194 (76.7)	185 (77.4)
In a disease management program for diabetes, n (%)	39 (15.4)	43 (18.0)
Revised ProMAS[a] score, mean (SD)	12.8 (3.3)	12.9 (3.0)
Revised ProMAS score ≤13, n (%)	129 (51.0)	115 (48.1)
Composite barrier score, mean (SD)	19.4 (2.34)	19.3 (2.26)
Second survey completed, n (%)	188 (74.3)	182 (76.2)

[a]ProMAS: Probabilistic Medication Adherence Scale.

Outcomes

Primary Outcome

Our primary outcome, medication adherence after 8 months, was 65.1% (84/129) in the intervention group and 66.1% (76/115) in the control group. There was no significant difference between the intervention group and the control group (B=–0.01; 95% CI –0.59 to 0.57; *P*=.96). Patients with programmed diabetes care showed a significantly better medication adherence (B=1.02; 95% CI 0.21-1.84; *P*=.01).

Analyzing the data considering the PDC as a continuous outcome gave comparable results (effect intervention: B= –0.74; 95% CI –11.0 to 9.5; *P*=.87; effect programmed diabetes care: B=16.3; 95% CI 6.2-26.5; *P*=.002).

In our post hoc per-protocol analysis we compared the medication adherence of 74/129 patients in the intervention group who received an intervention (71 based on records in the information system, and an additional 3 based on data from patient interviews) with the 115 patients in the control group. No significant difference in medication adherence was found for adherence dichotomized (B=0.19; 95% CI –0.50 to 0.89; *P*=.58) and as a continuous outcome (B=4.5; 95% CI –7.8 to 16.9; *P*=.40).

Medication adherence in all patients starting research medication in the control pharmacies during the research period was 47.79% (2471/5170). This differed significantly from the 72.8% (174/239) in our control sample (*P*<0.001).

Secondary Outcomes

ProMAS Score and Effectiveness of the Intervention

A total of 370 patients filled in the first as well as the second survey. In the intervention group 55.3% (104/188) had a low revised ProMAS score at 8 months' follow-up, indicative of a high nonadherence risk; in the control group 51.6% (94/182) had a low revised ProMAS score. The average revised ProMAS score was 12.64 and 12.74, respectively (*P*=.77).

Controlling for clustering in pharmacies, the revised ProMAS score at inclusion, age, gender, and disease management program (diabetes or not), the effect of the intervention was nonsignificant (B=0.05; 95 CI –0.46 to 0.57; *P*=.85). Revised ProMAS dichotomized gave comparable results (B=0.16; 95% CI –0.31 to 0.62; *P*=.50).

We assessed the effectiveness of the intervention on the barrier profile. The average score on the barrier profile at follow-up was 19.7 in the control group and 19.9 in the intervention group. Controlling for clustering in pharmacies, the barrier profile at inclusion, age, gender, and disease management program (diabetes or not), the effect of the intervention was nonsignificant (B=0.11; 95% CI –0.49 to 0.71; *P*=.69).

Predictive Values of Revised ProMAS

We tested the predictive value of the revised ProMAS and the barrier questionnaire on the data from all 239 patients in the control group (all data are presented in Table 2). In the control group medication adherence was 72.8% (174/239). Of the patients with a low revised ProMAS score predicting a high risk for nonadherence, 33.9% (39/115) had a PDC less than 80% (nonadherent). This was the positive predictive value of revised ProMAS. Conversely, the negative predictive value was 79% (98/124), meaning that 79% of the patients with a high revised ProMAS score not indicative for a high nonadherence risk had a PDC of 80% or more (adherent). The sensitivity of revised ProMAS was 60% (39/65).

Table 2. Revised ProMAS score versus medication adherence.

Revised ProMAS[a] score	Medication adherence		
	Nonadherent	Adherent	Total
≤13	39	76	115
≥14	26	98	124
Total	65	174	239

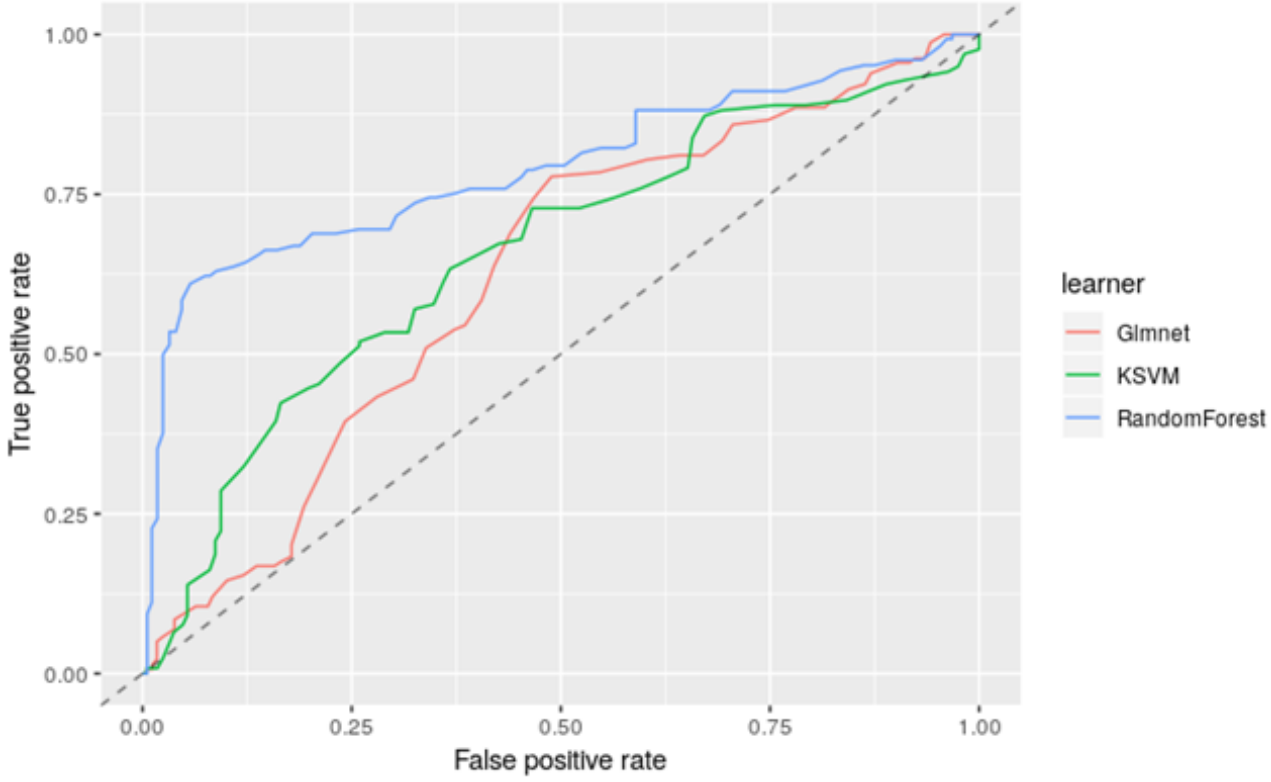
[a]ProMAS: Probabilistic Medication Adherence Scale.

Predictive Values of the Barrier Questionnaire

The barrier profile consists of 8 individual barrier scores that were used to train a machine learning algorithm to predict adherence at 8 months. Figure 2 shows the average performance of the 3 applied learners. The best performing learner was the random forest with an area under the curve of 0.795.

Applying the random forest model to the validation set (102/239 patients) yielded the following results: positive predictive value 74%; negative predictive value 80%; and sensitivity 76%.

Figure 2. The performance of 3 machine learning techniques (random forest, kernel support vector machines [KSVMs], and generalized linear models). The results are the average outcomes of a 10-fold cross validation and relate to the predictive value of the answers on the barrier assessment survey.



Discussion

Principal Findings

Among initiators of diabetic or cardiovascular drugs, we found no effect of our intervention on medication adherence as measured by pharmacy dispensing data at 8 months' follow-up. Further, the intervention did not significantly change the secondary outcomes for the medication adherence risk and barriers at 8 months' follow-up. Instead, receiving programmed diabetic care was associated with higher medication adherence.

The positive and negative predictive values of the risk assessment based on the revised ProMAS questionnaire were 34% and 79%, respectively. Sensitivity was 60%. The predictive values of the barrier questionnaire were 74 and 80%, respectively.

Below we discuss several factors that may have influenced the results: (1) the adherence in the control group was high and left little room for improvement; (2) our study lacked power, as the final patient number eligible for analysis was lower than expected; (3) the fidelity of the pharmacists in the intervention group to perform the intervention was lower than expected (many patients actually did not receive the intervention); (4) the accuracy of the revised ProMAS in predicting nonadherence was limited; (5) the quality of the intervention delivery may have been suboptimal; (6) the opportunities to enhance the impact of the intervention in the care chain were not leveraged.

An important hindrance in our study was—in itself very positive—the finding of the high medication adherence level in the control group: 66.1% (76/115) of the patients at high nonadherence risk had a PDC over 80%. Many patients had an already optimal PDC of 100% from usual care alone, which leaves no room for improvement. Although we have no data proving it, we expect that repeat prescription services and multidose dispensing systems may have led to more patients with a PDC of 100%. When offered more often in study patients, this might be seen as a form of cross over.

Our assumptions for the sample size calculation were based on average adherence data in the literature of 50%. This percentage was confirmed by the adherence of all patients starting research medication in the control pharmacies of 47.79% (2471/5170). The high medication adherence in our sample gives rise to several possible explanations. First, this may be due to selection in the pharmacies during the phase of inviting patients at the first dispensing. Selection bias might also have occurred at the patient level: patients who consider medication adherence important might be more willing to participate in the study. Another explanation of the high adherence rate in our study sample might be the so-called Hawthorn effect: knowing you participate in a study will influence your behavior [37,38]. Finally, the questionnaire used in both the intervention and control groups might have triggered patients to reflect on their behavior, their health, and the importance of medication, resulting in higher adherence rates. Therefore, filling out the questionnaire might be considered an intervention in itself.

We operationalized medication adherence by pharmacy dispense data, which might be criticized. The PDC as a surrogate measure of adherence is a conservative measure for nonpersistence without taking noninitiation into account [9].

Second, the intervention was not always delivered to those patients selected for it. All patients with a revised ProMAS score indicative of a high nonadherence risk had to be offered an intervention with notes in the multidisciplinary electronic file. From the lack of notes and from patient information gathered in the process evaluation, we learned that many eligible patients did not receive an intervention. Clear descriptions, possibilities for flexible time management, simple patient inclusion, and task delegation could increase participation in the intervention [39]. Although the pharmacists agreed to participate, the lack of flexibility relating to the timing of the intervention might have been a barrier to perform it according to the intervention protocol.

In our per-protocol analyses we only included those patients to whom the intervention was actually offered. However, in this smaller sample we did not find an effect.

Third, the number of included patients per month proved to be far lower than anticipated. The research team, therefore, put much effort in supporting the pharmacies to invite patients and we increased the potential of eligible patients by allowing them to be included from GPs other than the participating GP care group. Still, we did not manage to reach the calculated sample size. Moreover, we ended up with a large variation between numbers of included patient per pharmacy. Variation in cluster size requires even higher numbers to achieve the same statistical power level [40]. The final sample size was even further compromised because we had to exclude patients from the analysis who did not fulfil the inclusion criteria. This was mainly due to the stricter criteria for therapy initiation from our study compared with the automated alerts from the pharmacy system.

Fourth, the accuracy of the prediction of nonadherence risk was limited. We preferred to use the revised ProMAS over other adherence questionnaires as it measures behavior and not the beliefs, attitudes, and intentions [22]. The adherence results in our control group showed that 2 out of 3 patients who were offered an intervention would have been adherent at 8 months without an intervention. This is inefficient regarding time and means and dilutes any possible effect from a research perspective. By contrast, with a revised ProMAS sensitivity of 60% in 40% of nonadherent patients we missed the chance to offer these patients an intervention. Although personalized care involves risk stratification, the diagnostic ability of revised ProMAS may not have been sufficient yet. The predictive characteristics of the barrier questionnaire proved to be better.

Fifth, the quality of intervention delivery may have been suboptimal. Before the start of the project, pharmacists followed a 3-hour training in communication skills and intervention delivery. During the intervention period, pharmacists were offered extra training to improve their communication skills for the intervention. Process evaluation showed that this training was mainly used to discuss difficulties in patient inclusion and thus communication skills may not have been developed as expected.

Sixth, collaboration between pharmacy and general practice in reinforcing the intervention did not happen. Moreover, pharmacists did not offer devices to measure blood pressure or blood glucose to the patients, although we offered this as a way to improve patients' motivation [12]. Interviews during the study period with the pharmacists showed that they experienced a lack of skills to recruit and perform adherence conversation and often lacked time to execute interventions on busy days.

Improvement of our intervention might apply various elements. The basic principle of profiling a patient based on nonadherence risk and barriers for adherence seems to align well with the trend to provide personalized care in general and more specific with the trend to tailor medication adherence interventions to the needs of the patient.

Improving the instrument for patient selection would help to put the energy where it is most beneficial. The positive and negative predictive values of the revised ProMAS were 34% and 79%, respectively. In their review, Lam and Fresco [41] mentioned that the Morisky Medication Adherence Scale has advantages over other self-reporting adherence scales. Tan et al [42] found in their review 2 studies reporting predictive values of the Morisky Medication Adherence Scale. The positive predictive values were 0.41 and 0.71, the negative predictive values were 0.65 and 0.43, respectively [43,44]. Further research could include the evaluation of the individual responses of all patients to the revised ProMAS questionnaire. This would provide more insight into its test characteristics.

Additionally, from dispensing data available in the pharmacy, it should be possible to only provide the intervention to patients on chronic medication not collecting their medication in time.

The assessment of the predictive values of the barrier profile shows better results to target patients than those of the revised ProMAS, indicating that the barriers measured are relevant for medication adherence.

More collaboration in the chain of health care professionals is another possible way to strengthen the intervention. When in the general practice the barrier profile and the data from the pharmacist's intervention are directly visible in the chain system, the GP and practice nurse can build upon the intervention or at least support and underline it. In our study the intervention was not supported in the general practice.

Strengths and Limitations

As discussed above, the sample size was an important limitation of our study. We failed to include the patient numbers needed and for about 100 participants medication data could not be linked. Almost 200 patients included were not eligible because they were not a "starter" with one of the trial medications. To increase patient recruitment, we allowed pharmacists who hardly managed to include patients to include patients listed with GPs from another care group. Although we do not expect effect on the outcome, this change in eligibility is a limitation.

Another limitation was that we had to adapt the ProMAS questionnaire for those patients not on medication yet. Consequently, the version we used, excluding questions, was not formally validated as the original version. The measurement of nonadherence risk only shortly after medication initiation may result in a slightly more positive nonadherence risk due to the situation-specific nature of certain items and the decreased chance of the occurrence of these situations in a short period compared with a longer period (eg, When I am away from home, I occasionally do not take my medicine). However, we used a liberal cut-off, thereby lowering the chances of missing any patients at increased risk for nonadherence.

In our research we were not able to control for polypharmacy. Patients on medication and starting with another medication (add-on) are in another situation than patients starting with 1 first medication. While more medication can add more difficulties, it could also enhance adherence to existing regimen(s).

An important strength of our study was the trial design, with pharmacies being randomized and analyses taking patient clustering into account. This was achieved in collaboration with an industrial company providing user-friendly technology and the care professionals with their care group policy makers. This study was carried out in the daily practice after a pilot phase to customize the processes and materials.

Taking less medication by skipping dosages and stopping after a certain period may result in the same PDC indicating nonadherence. So, our PDC data do not allow for conclusion on the type of nonadherence.

Conclusion

Our tailored intervention for initiators of cardiovascular or diabetes medication did not improve medication adherence compared with usual care. However, interventions tailored to individual barriers of those patients with an increased risk for nonadherence appear to be a good strategy in line with the current policy to personalize care. A study with a better selection of those patients who could benefit and a better implementation of the intervention might well show positive results.

Acknowledgments

We acknowledge Stichting Categorale Zorg voor apothekers in Zuidoost Brabant (CaZo), the regional pharmacist organization, and De Ondernemende Huisarts (DOH), the general practitioners collaborative, for their collaboration. Funding for this study is provided by Philips.

Conflicts of Interest

Two authors (JL and AvH) are salaried employees of Philips. The other authors (JvL and MT) are salaried professionals of the Scientific Center for Quality of Healthcare (IQ healthcare), Radboudumc. IQ healthcare received a grant from Philips.

Multimedia Appendix 1
Patients starting medication from the listed ATC codes for cardiovascular or oral blood glucose lowering medication were eligible for inclusion.
[DOCX File , 17 KB - jmir_v24i4e16141_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2495 KB - jmir_v24i4e16141_app2.pdf]

References

1. Schroeder K, Fahey T, Ebrahim S. Interventions for improving adherence to treatment in patients with high blood pressure in ambulatory settings. Cochrane Database Syst Rev 2004(2):CD004804. [doi: 10.1002/14651858.CD004804] [Medline: 15106262]

2. Dolce JJ, Crisp C, Manzella B, Richards JM, Hardin JM, Bailey WC. Medication adherence patterns in chronic obstructive pulmonary disease. Chest 1991 Apr;99(4):837-841. [doi: 10.1378/chest.99.4.837] [Medline: 2009784]
3. Haynes RB, Ackloo E, Sahota N, McDonald HP, Yao X. Interventions for enhancing medication adherence. Cochrane Database Syst Rev 2008 Apr 16(2):CD000011. [doi: 10.1002/14651858.CD000011.pub3] [Medline: 18425859]
4. Krueger KP, Berger BA, Felkey B. Medication adherence and persistence: a comprehensive review. Adv Ther 2005;22(4):313-356. [doi: 10.1007/BF02850081] [Medline: 16418141]
5. Sarafino EP, Smith TW. Biopsychosocial Interactions (7th edition). Hoboken, NJ: Wiley; 2010.
6. Blaschke TF, Osterberg L, Vrijens B, Urquhart J. Adherence to medications: insights arising from studies on the unreliable link between prescribed and actual drug dosing histories. Annu Rev Pharmacol Toxicol 2012;52:275-301. [doi: 10.1146/annurev-pharmtox-011711-113247] [Medline: 21942628]
7. Horne R. Patients' beliefs about treatment: the hidden determinant of treatment outcome? J Psychosom Res 1999 Dec;47(6):491-495. [doi: 10.1016/s0022-3999(99)00058-6] [Medline: 10661596]
8. Horne R, Chapman SCE, Parham R, Freemantle N, Forbes A, Cooper V. Understanding patients' adherence-related beliefs about medicines prescribed for long-term conditions: a meta-analytic review of the Necessity-Concerns Framework. PLoS One 2013;8(12):e80633 [FREE Full text] [doi: 10.1371/journal.pone.0080633] [Medline: 24312488]
9. Vrijens B, De Geest S, Hughes DA, Przemyslaw K, Demonceau J, Ruppar T, et al. A new taxonomy for describing and defining adherence to medications. Br J Clin Pharmacol 2012 May;73(5):691-705 [FREE Full text] [doi: 10.1111/j.1365-2125.2012.04167.x] [Medline: 22486599]
10. World Health Organization (WHO). Adherence to long-term therapies: Evidence for action. WHO. 2003. URL: https://tinyurl.com/yaydj9ws [accessed 2022-03-14]
11. Nieuwlaat R, Wilczynski N, Navarro T, Hobson N, Jeffery R, Keepanasseril A, et al. Interventions for enhancing medication adherence. Cochrane Database Syst Rev 2014 Nov 20(11):CD000011 [FREE Full text] [doi: 10.1002/14651858.CD000011.pub4] [Medline: 25412402]
12. Ryan R, Santesso N, Lowe D, Hill S, Grimshaw J, Prictor M, et al. Interventions to improve safe and effective medicines use by consumers: an overview of systematic reviews. Cochrane Database Syst Rev 2014 Apr 29(4):CD007768 [FREE Full text] [doi: 10.1002/14651858.CD007768.pub3] [Medline: 24777444]
13. Omboni S, Caserini M. Effectiveness of pharmacist's intervention in the management of cardiovascular diseases. Open Heart 2018 Jan 03;5(1):e000687 [FREE Full text] [doi: 10.1136/openhrt-2017-000687] [Medline: 29344376]
14. Elnaem MH, Rosley NFF, Alhifany AA, Elrggal ME, Cheema E. Impact of Pharmacist-Led Interventions on Medication Adherence and Clinical Outcomes in Patients with Hypertension and Hyperlipidemia: A Scoping Review of Published Literature. J Multidiscip Healthc 2020;13:635-645 [FREE Full text] [doi: 10.2147/JMDH.S257273] [Medline: 32764955]
15. Fuller RH, Perel P, Navarro-Ruan T, Nieuwlaat R, Haynes RB, Huffman MD. Improving medication adherence in patients with cardiovascular disease: a systematic review. Heart 2018 Aug;104(15):1238-1243. [doi: 10.1136/heartjnl-2017-312571] [Medline: 29572248]
16. Hamdidouche I, Jullien V, Boutouyrie P, Billaud E, Azizi M, Laurent S. Drug adherence in hypertension: from methodological issues to cardiovascular outcomes. J Hypertens 2017 Jun;35(6):1133-1144. [doi: 10.1097/HJH.0000000000001299] [Medline: 28306634]
17. van Driel ML, Morledge MD, Ulep R, Shaffer JP, Davies P, Deichmann R. Interventions to improve adherence to lipid-lowering medication. Cochrane Database Syst Rev 2016 Dec 21;12:CD004371 [FREE Full text] [doi: 10.1002/14651858.CD004371.pub4] [Medline: 28000212]
18. Vignon Zomahoun HT, de Bruin M, Guillaumie L, Moisan J, Grégoire JP, Pérez N, et al. Effectiveness and Content Analysis of Interventions to Enhance Oral Antidiabetic Drug Adherence in Adults with Type 2 Diabetes: Systematic Review and Meta-Analysis. Value Health 2015 Jun;18(4):530-540 [FREE Full text] [doi: 10.1016/j.jval.2015.02.017] [Medline: 26091607]
19. Erkens JA, Panneman MMJ, Klungel OH, van den Boom G, Prescott MF, Herings RMC. Differences in antihypertensive drug persistence associated with drug class and gender: a PHARMO study. Pharmacoepidemiol Drug Saf 2005 Nov;14(11):795-803. [doi: 10.1002/pds.1156] [Medline: 16178043]
20. Penning-van Beest F, van Herk-Sukel M, Gale R, Lammers J, Herings R. Three-year dispensing patterns with long-acting inhaled drugs in COPD: a database analysis. Respir Med 2011 Feb;105(2):259-265 [FREE Full text] [doi: 10.1016/j.rmed.2010.07.007] [Medline: 20705441]
21. Hilbink M, Lacroix J, Bremer-van der Heiden L, van Halteren A, Teichert M, van Lieshout J. Effectiveness of a medication-adherence tool: study protocol for a randomized controlled trial. Trials 2016 Jun 03;17(1):274 [FREE Full text] [doi: 10.1186/s13063-016-1393-2] [Medline: 27255080]
22. Kleppe M, Lacroix J, Ham J, Midden C. The development of the ProMAS: a Probabilistic Medication Adherence Scale. Patient Prefer Adherence 2015;9:355-367 [FREE Full text] [doi: 10.2147/PPA.S76749] [Medline: 25784791]
23. Kleppe M, Lacroix J, Ham J, Midden C. 'A necessary evil': associations with taking medication and their relationship with medication adherence. Psychol Health Med 2017 Dec;22(10):1217-1223. [doi: 10.1080/13548506.2017.1316412] [Medline: 28393554]

24. Rich A, Brandes K, Mullan B, Hagger MS. Theory of planned behavior and adherence in chronic illness: a meta-analysis. J Behav Med 2015 Aug;38(4):673-688. [doi: 10.1007/s10865-015-9644-3] [Medline: 25994095]
25. Strack F, Deutsch R. Reflective and impulsive determinants of social behavior. Pers Soc Psychol Rev 2004;8(3):220-247. [doi: 10.1207/s15327957pspr0803_1] [Medline: 15454347]
26. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. Fam Med 2004 Sep;36(8):588-594 [FREE Full text] [Medline: 15343421]
27. Jamous RM, Sweileh WM, El-Deen Abu Taha AS, Zyoud SH. Beliefs About Medicines and Self-reported Adherence Among Patients with Chronic Illness: A Study in Palestine. J Family Med Prim Care 2014 Jul;3(3):224-229 [FREE Full text] [doi: 10.4103/2249-4863.141615] [Medline: 25374859]
28. ATC codes. WHO Collaborating Centre for Drug Statistics. URL: https://www.whocc.no/atc_ddd_index/ [accessed 2022-10-03]
29. Doshi JA, Zhu J, Lee BY, Kimmel SE, Volpp KG. Impact of a prescription copayment increase on lipid-lowering medication adherence in veterans. Circulation 2009 Jan 27;119(3):390-397 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.108.783944] [Medline: 19139387]
30. Haynes RB, Taylor DW, Sackett DL, Gibson ES, Bernholz CD, Mukherjee J. Can simple clinical measurements detect patient noncompliance? Hypertension 1980;2(6):757-764. [doi: 10.1161/01.hyp.2.6.757] [Medline: 7007235]
31. 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]
32. Bischl B, Lang M, Kotthoff L, Schiffner J, Richter J, Studerus E, et al. Mlr: Machine Learning in R. JMLR 2016;17(170):1-5 [FREE Full text]
33. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2018. URL: https://www.R-project.org/ [accessed 2022-03-14]
34. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI'95) - Volume 2. San Francisco, CA: Morgan Kaufmann Publishers; 1995 Presented at: Proceedings of the 14th International Joint Conference on Artificial Intelligence; August 20-25, 1995; Montreal, QC p. 1137-1143.
35. Li W, Qian Z, Li H. Interpretation and Classification of P-Series Recommendations in ITU-R. IJCNS 2016;09(05):117-125. [doi: 10.4236/ijcns.2016.95010]
36. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 2010;33(1):1-22 [FREE Full text] [Medline: 20808728]
37. Hawthorne Effect. Wikipedia. URL: https://en.wikipedia.org/wiki/Hawthorne_effect [accessed 2022-03-14]
38. Adair JG. The Hawthorne effect: A reconsideration of the methodological artifact. Journal of Applied Psychology 1984;69(2):334-345. [doi: 10.1037/0021-9010.69.2.334]
39. Kuipers E, Wensing M, De Smet PAGM, Teichert M. Barriers and facilitators for community pharmacists' participation in pharmacy practice research: a survey. Int J Pharm Pract 2019 Aug;27(4):399-402 [FREE Full text] [doi: 10.1111/ijpp.12522] [Medline: 30784142]
40. Kerry SM, Bland JM. Unequal cluster sizes for trials in English and Welsh general practice: implications for sample size calculations. Stat Med 2001 Feb 15;20(3):377-390. [doi: 10.1002/1097-0258(20010215)20:3<377::aid-sim799>3.0.co;2-n] [Medline: 11180308]
41. Lam WY, Fresco P. Medication Adherence Measures: An Overview. Biomed Res Int 2015;2015:217047 [FREE Full text] [doi: 10.1155/2015/217047] [Medline: 26539470]
42. Tan X, Patel I, Chang J. Review of the four item Morisky Medication Adherence Scale (MMAS-4) and eight item Morisky Medication Adherence Scale (MMAS-8). Innov Pharm 2014 Jan 01;5(3):165 [FREE Full text] [doi: 10.24926/iip.v5i3.347]
43. van de Steeg N, Sielk M, Pentzek M, Bakx C, Altiner A. Drug-adherence questionnaires not valid for patients taking blood-pressure-lowering drugs in a primary health care setting. J Eval Clin Pract 2009 Jun;15(3):468-472. [doi: 10.1111/j.1365-2753.2008.01038.x] [Medline: 19366389]
44. Sakthong P, Chabunthom R, Charoenvisuthiwongs R. Psychometric properties of the Thai version of the 8-item Morisky Medication Adherence Scale in patients with type 2 diabetes. Ann Pharmacother 2009 May;43(5):950-957. [doi: 10.1345/aph.1L453] [Medline: 19366872]

Abbreviations

ATC: Anatomic Therapeutical Chemical
BOMM: Begeleiding op Maat bij Medicatie
CaZo: Stichting Categorale Zorg voor apothekers in Zuidoost Brabant
CMO: Commissie Mensgebonden Onderzoek
DOH: De Ondernemende Huisarts
GP: general practitioner

PDC: percentage of days covered
ProMAS: Probabilistic Medication Adherence Scale

Edited by A Mavragani; submitted 09.09.19; peer-reviewed by M Klein, I Arnet, P Fresco, S Boucher, M Semwal; comments to author 05.09.20; revised version received 30.10.20; accepted 02.02.22; published 07.04.22.

Please cite as:
van Lieshout J, Lacroix J, van Halteren A, Teichert M
Effectiveness of a Pharmacist-Led Web-Based Medication Adherence Tool With Patient-Centered Communication: Results of a Clustered Randomized Controlled Trial
J Med Internet Res 2022;24(4):e16141
URL: https://www.jmir.org/2022/4/e16141
doi:10.2196/16141
PMID:35389359

Original Paper

The Mutual Influence of the World Health Organization (WHO) and Twitter Users During COVID-19: Network Agenda-Setting Analysis

Iman Tahamtan[1], PhD; Devendra Potnis[1], PhD; Ehsan Mohammadi[2], PhD; Vandana Singh[1], PhD; Laura E Miller[3], PhD

[1]School of Information Sciences, The University of Tennessee, Knoxville, TN, United States

[2]School of Information Science, The University of South Carolina, Columbia, SC, United States

[3]School of Communication Studies, The University of Tennessee, Knoxville, TN, United States

Corresponding Author:
Iman Tahamtan, PhD
School of Information Sciences
The University of Tennessee
1345 Circle Park Drive
451 Communications Building
Knoxville, TN, 37996-0332
United States
Phone: 1 865 974 1000
Email: iman.tahamtan@gmail.com

Abstract

Background: Little is known about the role of the World Health Organization (WHO) in communicating with the public on social media during a global health emergency. More specifically, there is no study about the relationship between the agendas of the WHO and Twitter users during the COVID-19 pandemic.

Objective: This study utilizes the network agenda-setting model to investigate the mutual relationship between the agenda of the WHO's official Twitter account and the agenda of 7.5 million of its Twitter followers regarding COVID-19.

Methods: Content analysis was applied to 7090 tweets posted by the WHO on Twitter from January 1, 2020, to July 31, 2020, to identify the topics of tweets. The quadratic assignment procedure (QAP) was used to investigate the relationship between the WHO agenda network and the agenda network of the 6 Twitter user categories, including "health care professionals," "academics," "politicians," "print and electronic media," "legal professionals," and the "private sector." Additionally, 98 Granger causality statistical tests were performed to determine which topic in the WHO agenda had an effect on the corresponding topic in each Twitter user category and vice versa.

Results: Content analysis revealed 7 topics that reflect the WHO agenda related to the COVID-19 pandemic, including "prevention," "solidarity," "charity," "teamwork," "ill-effect," "surveillance," and "credibility." Results of the QAP showed significant and strong correlations between the WHO agenda network and the agenda network of each Twitter user category. These results provide evidence that WHO had an overall effect on different types of Twitter users on the identified topics. For instance, the Granger causality tests indicated that the WHO tweets influenced politicians and print and electronic media about "surveillance." The WHO tweets also influenced academics and the private sector about "credibility" and print and electronic media about "ill-effect." Additionally, Twitter users affected some topics in the WHO. For instance, WHO followers affected "charity" and "prevention" in the WHO.

Conclusions: This paper extends theorizing on agenda setting by providing empirical evidence that agenda-setting effects vary by topic and types of Twitter users. Although prior studies showed that network agenda setting is a "one-way" model, the novel findings of this research confirm a "2-way" or "multiway" effect of agenda setting on social media due to the interactions between the content creators and audiences. The WHO can determine which topics should be promoted on social media during different phases of a pandemic and collaborate with other public health gatekeepers to collectively make them salient in the public.

(J Med Internet Res 2022;24(4):e34321) doi:10.2196/34321

KEYWORDS

COVID-19; agenda setting; network agenda setting; Twitter; social media; public opinion; content analysis; public health; WHO

Introduction

Problem Statement

Social media has changed how online users share and receive news and information on various public health emergencies. A contagious and fatal public health emergency that started in early 2020 was the COVID-19 pandemic, with 263,563,622 confirmed cases, including 5,232,562 deaths worldwide as of December 03, 2021 [1]. Right after the outbreak of the virus, social media became the main technology for sharing and receiving information about various aspects of the pandemic.

During the pandemic, social media users received information from various sources such as the news media, politicians, and celebrities. These sources sometimes disseminate biased information due to their politically biased and partisan stance on public issues, consequently impacting public opinion and behavior [2]. Brennen et al [3] found that approximately 20% of misinformation about COVID-19 on social media with a high engagement was posted by gatekeepers, such as politicians, celebrities, and other influential public figures. Thus, even trusted sources of information are likely to produce biased and irrelevant information [3,4], which can develop biased opinions or a false reality about COVID-19.

The widespread distribution of information on social media resulted in the dissemination of low-quality and unverified stories and facts about COVID-19 [3,5,6], increasing public uncertainty about the situation and what will happen next [7].

Potential Solution

In such situations, trusted sources of public health information such as the World Health Organization (WHO) can develop an agenda to make the public aware of existing challenges and fight against the spread of false information. The WHO agenda can be defined as the topics emphasized and presented to the public at a given time [8].

Research Questions

To understand and explore the role of the WHO in setting a public health agenda regarding COVID-19 for various social media user categories (ie, health care professionals, academics, politicians, print and electronic media, legal professionals, and the private sector), this study investigated the following research questions:

- Research question 1 (RQ1): "Which topics (ie, agenda) related to the COVID-19 pandemic were promoted by the WHO's account on Twitter?"
- Research question 2 (RQ2): "What is the relationship between the WHO agenda regarding COVID-19 and the agenda of different categories of WHO followers on Twitter?"
- Research question 3 (RQ3): "How do the WHO and different categories of WHO followers on Twitter affect each other's agenda regarding COVID-19?"

These research questions help us (1) understand the extent to which the WHO has been successful in setting its agenda about COVID-19 on Twitter and (2) propose how the WHO can more effectively create information-related benefits for different categories of users on Twitter during public health emergencies.

This study adopted the agenda-setting theory [9] as a lens to address the research questions, as it can be used to understand and explain the impact of gatekeepers in shaping public agenda on various issues such as health crises. The importance of agenda setting is that it can be employed by gatekeepers to tell the public what critical issues a country is facing [10] and to impact public opinion about those issues [11]. This study is specifically focused on the network agenda-setting model [12], which states that the salience of interrelationships among attributes (eg, social distancing, handwashing, face covering) of an issue (eg, COVID-19) emphasized by gatekeepers can be transferred to (affect) the public agenda [12].

Literature Review

Agenda Setting

Gatekeepers such as the news media determine which issues are important in society and consequently set the public agenda around those issues [9]. When an issue is being shared frequently and prominently, the public may also come to perceive them as important [13]. For instance, the presentation and repetition of "social distancing" by the WHO on social media could make the majority of people perceive it as an important attribute of COVID-19 that should be given considerable attention.

The agenda-setting theory [9] is about transferring the salience of issues and issue attributes to the public, suggesting that gatekeepers can tell people what to think about and how to think about them [11]. For example, COVID-19 is a public health issue that has a variety of attributes (eg, social distancing, hand hygiene, face covering). The WHO can set an agenda around COVID-19, to bring the importance of COVID-19 and its attributes to the public's attention.

Agenda-Setting Effect

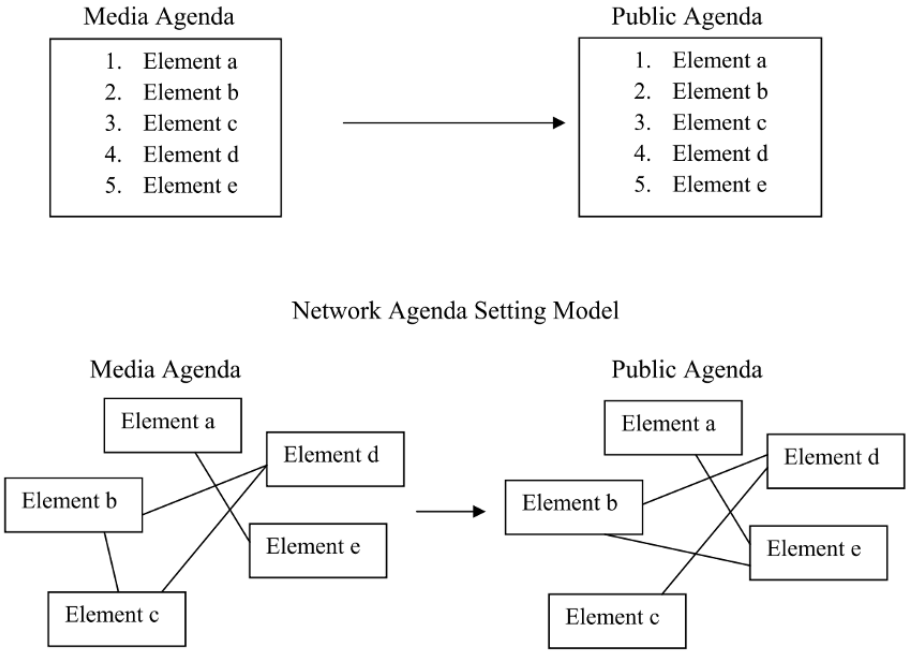
Agenda setting proposes that the repetition of messages about public issues by a gatekeeper can influence people's minds [13]. As a result, people mentally connect to those issues to a similar degree in which they have been emphasized by the gatekeeper [14]. For instance, by emphasizing social distancing, hand hygiene, and face covering, the public is more likely to make links to these 3 attributes in their minds than other less highlighted attributes.

Agenda-setting research uses correlation tests to investigate how the issue and issue attributes presented by a gatekeeper (representing the gatekeeper agenda) correlate with the issue and issue attributes in the public discourse (representing the public agenda). The assumption is that, if there is a positive correlation between both agendas, the gatekeeper, such as the news media, has been able to impact the public agenda [15].

Network Agenda Setting

Guo and McCombs [12] proposed an agenda-setting model, called network agenda setting, according to which gatekeepers like the news media have the power to impact people's cognitive network. A gatekeeper builds "network connections" between the attributes of an issue and transfers those interconnected attributes to the public's minds (see Figure 1). For instance, if the news media consider a country's economic problems to be associated with its foreign policy, audiences are also likely to consider them to be associated with each other [16]. According to network agenda setting, "the more frequently that two elements are associated in the news coverage, the more likely it is that the audience will consider the two interconnected" [16].

Figure 1. Comparison of a traditional agenda-setting approach with a network agenda-setting model [17].



Note: Elements could refer to objects, attributes, or combinations of objects and attributes.

Therefore, the attributes of an issue can be transferred to the public as a network of interconnected attributes. For instance, social distancing, handwashing, and face covering, as 3 attributes related to COVID-19, can be transferred to the public agenda simultaneously as a network of attributes [12]. That said, when people think about social distancing, they would also think about handwashing and face covering as measures for protecting against the COVID-19 virus. Therefore, the public can be told not only what issues and attributes to think about and how to think about them but also how to link those issues and attributes in their minds [15].

The main difference between traditional agenda setting and network agenda setting is that the former assumes an issue and its attributes are separately and discretely transferred to the public agenda, while the latter assumes they are transferred simultaneously as a bundle of networked attributes [12]. Figure 1 [18] compares the traditional and network agenda-setting models.

Network Agenda-Setting Effect

To identify the network agenda-setting effect, just as with the traditional agenda-setting effect, correlation tests are used to identify how well a gatekeeper's agenda is correlated with the public agenda. However, in network agenda setting, the agenda will be presented in a network or co-occurrence square matrix (ie, a matrix with the same number of rows and columns) consisting of issue attributes (see Table 1). The value in each cell in the matrix represents how many times the 2 corresponding attributes have co-occurred in a data set: the higher the value, the stronger the 2 attributes are connected. For instance, in Table 1, attributes 3 and 2, with a co-occurrence of 30, are connected stronger than other attributes [18].

Therefore, a matrix represents the agenda network of a gatekeeper or the public for a given issue in a specific period [15]. Assessing the correlation between the 2 matrices can determine if the former has had any agenda-setting effect on the latter [18].

In network agenda-setting studies, social network analysis is applied to illustrate how issue attributes are interrelated [18]. In a network that demonstrates an agenda, each node represents an attribute of an issue, and each tie between any 2 nodes represents their relationships. The number of times the 2 attributes co-occur in a data set represents the strength of the tie [19].

Table 1. The matrix of an agenda network for a hypothetical gatekeeper.

Attribute	Attribute 1	Attribute 2	Attribute 3	Attribute 4
Attribute 1	—[a]	15	25	5
Attribute 2	15	—	30	15
Attribute 3	25	30	—	12
Attribute 4	5	15	12	—

[a]Not applicable.

Agenda Setting on Social Media

Social media plays an increasingly significant role in agenda setting [20] by making large-scale communication possible and giving voice to different groups of people, such as minorities [21]. The increasing adoption of social media has changed how the news media, politicians, and other influential actors communicate with people and perform their agenda-setting activities [20]. For instance, during elections, candidates and their campaigns make strategic use of social media to mobilize voters by bringing their attention to the issues that are of great public concern [22]. Agenda setting can also be used in health promotion activities on social media [23].

Hemsley [24] stated that the strategic use of social media such as Twitter and its features such as hashtags could establish and promote health, social, political, or environmental agendas. Hashtags and the stories that form around them can become part of people's social reality and inform their worldviews [24]. For instance, Twitter and hashtags can be used to enhance information dissemination, publicize the movement, invite new people to the movement, enhance its visibility, broadcast messages to broader audiences, and attract people's attention [25,26].

Lee and Xu [27] also noted that Twitter could be used during elections to set public agendas. For instance, Donald Trump used Twitter and hashtags during the 2016 US presidential election to develop a variety of public agendas, most importantly, the "media bias" and "Clinton's alleged dishonesty." Lee and Xu [27] showed that Donald Trump was more successful than Hillary Clinton in drawing public attention to the agendas highlighted by his campaign on Twitter. Feezell [17] indicated that being exposed to political information on Facebook increased perceived issue salience and importance, yielding an agenda-setting effect.

The Gap in the Network Agenda-Setting Literature

Past agenda-setting studies have been mostly focused on gatekeepers such as the news media and topics such as political issues (eg, [17,28]). However, it is less studied how global health gatekeepers such as the WHO set an agenda on social media during public health emergencies and what impact they can have on various social media users.

Past studies have suggested that agenda setting can have different effects on various public subgroups [11,19,29]. The network agenda-setting model and the literature in this area does not focus on how different user categories on social media are affected by the gatekeepers they follow on social media.

This study fills in these gaps in the literature, by studying the network agenda-setting effect of the WHO on 6 Twitter user categories (that follow WHO's Twitter account).

Methods

Classifying WHO Followers on Twitter to 6 User Categories

The WHO's followers on Twitter were classified into 6 categories, including (1) health care professionals, (2) academics, (3) politicians, (4) print and electronic media, (5) legal professionals, and (6) the private sector. This classification was done to investigate if and how different categories on social media react to a public health issue.

We adopted and modified the approach used by Toupin et al [30] to classify Twitter users. The main reason for selecting these categories is that they are influential actors in society whose actions can impact citizens during crisis events as explained in the following paragraphs.

In this study, health care professionals include public health workers in academia and industry sectors, such as doctors and nurses. Health care professionals serve citizens during public health emergencies in various ways, for instance, by warning them against self-medicating [31].

Academics were defined as the people who work in academia (except health care professionals), such as professors, researchers, and students who can inform society through their research and writing about various issues related to the COVID-19 pandemic [32].

Politicians include policy and decision makers in the state and federal governments, such as mayors, congressmen, congresswomen, and senators. Politicians can impact people by, for instance, communicating with them about the concerns raised around vaccine safety [33].

Print and electronic media include those responsible for information dissemination and public awareness, such as journalists, press, news agencies, and publishers. The print and electronic media are powerful sources of information for the public as a crisis unfolds [34]. They can, for example, fact-check the information shared on social media related to COVID-19 [35].

Legal professionals include courts, lawyers, and attorneys who provide legal advice and resources, for instance, for an ideal crisis communication strategy [36]. Additionally, they can ensure the compliance of policies with national laws such as human rights [37].

Private sectors include corporations, incorporated organizations, companies, chief executive officers, and for-profits who serve society by satisfying the needs of citizens. For example, companies can take necessary actions to serve their customers by empowering their employees to operate remotely [38].

The classification of followers into 6 categories was done according to the short biographical profiles of Twitter users. The data sets used in this study can be found online [39]. R software was used to collect and analyze the biographies of Twitter users according to the steps in the following paragraphs.

First, the IDs of the approximately 7.5 million Twitter accounts following the WHO in 2020 were retrieved, and the "username" associated with each ID was collected.

Next, the "description" used in biographies of the accounts (ie, usernames) were collected. Table S12 in Multimedia Appendix 1 provides a list of keywords used to classify and identify the Twitter user categories. The list was generated by searching on the Internet and using websites such as the US Bureau of Labor Statistics [40] that provide a list of job titles in various domains. For instance, the keywords used to identify "academics" were as follows: "lecturer," "professor," "phd," "student," "ph.d.," "postdoc," "postdoctoral," "doctoral," "msc," "master," "ms," "bs," "bachelor," "undergrad," "grad," "graduate," "undergraduate," "scientist," "postgrad," "faculty," "chancellor," "university," "college," "school," "provost," and "vice-provost."

To make each category as exclusive as possible and increase the reliability of the classification, general keywords, such as "research," "researching," "teams," "organizations," "institutes," "campus," "professional," "officer," "change," "equity," and "policy," were excluded in classifying Twitter users. The preliminary investigation and analysis of user biographies indicated that some keywords such as "chief" or "boss" appeared in several categories. Such keywords were also excluded.

Data Collection

Collecting WHO Tweets

Twitter was the main source for data collection. Twitter contains features such as hashtags, retweets, replies, mentions, and likes that make it a suitable platform for studying online social interactions [41]. Using the *brandwach* platform, the current study collected 7090 tweets posted by the WHO from January 1, 2020, to July 31, 2020.

Collecting Tweets for the 6 Twitter User Categories

The tweets related to COVID-19 posted by each Twitter user category from January 1, 2020, to July 31, 2020, were collected using the *rtweet* package.

Data Cleaning

Removing Bots

Previous studies have shown that bot accounts are active on Twitter, specifically during political debates, social movements, and public health emergencies. For instance, Ferrara [42] indicated that bots actively promoted political conspiracies during the COVID-19 pandemic. Our study used the *tweetbotornot* package in the R software to classify Twitter accounts into bots and nonbots; the package is 91.78% accurate in identifying bots and 92.61% accurate in identifying nonbots [43]. Accounts that receive a score of at least 50%, or a probability of 0.5, are bots and should be removed from the analysis [44]. Of the 656,805 Twitter accounts in the data set, 441,041 (ie, 67.15%) were classified as bots and consequently excluded from the analysis.

Removing Non-English and Non-COVID-19 Tweets

The WHO tweets that included irrelevant keywords such as Ebola, opioids, cancer, tobacco, and malaria were removed. Only the tweets that contained keywords and hashtags related to COVID-19 (see A1 in Multimedia Appendix 1) were included in the study. This approach removed 2111 (ie, 29.77%) of the tweets, leaving 4979 tweets in the WHO data set.

Overall, 7,965,610 tweets written in non-English languages were removed from the data set of WHO followers, leaving 7,547,019 tweets (48.65% of the total tweets) in the data set written in English (see Table S8 in Multimedia Appendix 1). Only the tweets posted by the 6 Twitter user categories that contained at least a keyword or hashtag related to COVID-19 (see A2 in Multimedia Appendix 1) were included in the analysis. This resulted in including 918,976 tweets (ie, 5.9% of the total tweets) related to COVID-19 and written in English in the data set (see Table S8 in Multimedia Appendix 1).

Content Analysis to Identify the WHO's Agenda

Content analysis was used to identify the main topics discussed by the WHO in the first half of 2020. These topics constitute the WHO's agenda related to COVID-19 on Twitter. An agenda is the topics presented in the public or media or any other medium at a given time [8].

The tweets posted by the WHO (n=7090) were stratified by months (January 2020 to July 2020) in Microsoft Excel. Using stratified sampling, 10.00% (n=709) of WHO tweets were randomly selected. The sample size for each stratum (ie, month) was proportional to the number of tweets in the stratum [45], as indicated in Table S9 in Multimedia Appendix 1.

The content analysis was carried out in 3 steps: In the first step, called the training phase, the 2 coders familiarized themselves with the content of the tweets. They coded a set of randomly selected tweets together inductively to increase their comfort level with and learn about the content of tweets. This helped them come to the same understanding of how the tweets should be coded [46]. The second phase, called the pilot test, included coders coding 100 of the tweets from a separate representative sample and assessing the intercoder reliability, independently [47], and resolving the discrepancies. The final step included coding the total sample size, independently, which resulted in an intercoder reliability agreement of 85%.

The coding was done by assigning 1 to 3 keywords that described the tweet's content to each tweet. For example, "handwash" and "handrub" were the 2 keywords assigned to a tweet by the WHO [48].

In analyzing the WHO tweets, the context of the tweets was taken into consideration: The keywords that appeared in more than one topic or those not contextually meaningful were

removed. For instance, "support" was removed because it appeared in 3 topics. Each topic was identified by a set of exclusive keywords and hashtags (see Table S11 in Multimedia Appendix 1).

This study ensured the keywords within each topic were representative of that category by several rounds of discussions among researchers. We identified 7 topics in the WHO tweets, including "prevention," "solidarity," "charity," "teamwork," "ill-effect," "surveillance," and "credibility." These topics constitute the WHO's agenda about COVID-19 in the first half of 2020. The frequency of the 7 topics in the entire WHO data set was calculated. To identify the relationship between the agenda network of the WHO and its followers, we also identified the frequency of each topic for the 6 Twitter user categories. To do this, the frequency of keywords and hashtags within each topic was calculated and summed.

Constructing Co-occurrence Matrices and Networks

To create the agenda network for the WHO and each Twitter user category, co-occurrence matrices should be created [12,19,49]. Following the literature (eg, [18,49]), a co-occurrence matrix was created for each of the 7 topics for the WHO and the 6 Twitter user categories. Each matrix included 7 columns and 7 rows. Each row and each column represent a topic related to COVID-19. Each cell contains a digit representing how many times the 2 topics have co-occurred. For example, the cell associated with "teamwork" and "charity" in the WHO matrix has a value of 51, which means that they were mentioned 51 times together by the WHO on Twitter in the first half of 2020. The number of times the 2 topics co-occurred in a data set represents the strength of the tie between those topics [19]. For instance, Table 2 demonstrates the co-occurrence matrix for WHO followers (ie, all 6 Twitter user categories together). To provide a better understanding of what the matrices look like, they were visualized in networks using the *quanteda* library in the R software. The networks associated with the matrices are presented in Multimedia Appendices 2-9.

Table 2. The matrix of topics in the data set of World Health Organization (WHO) followers on Twitter.

Topic	Teamwork	Charity	Surveillance	Prevention	Solidarity	Ill-effect	Credibility
Teamwork	0	1303	331	3041	2390	307	208
Charity	1303	0	141	4159	4838	332	86
Surveillance	331	141	0	2849	856	85	82
Prevention	3041	4159	2849	0	13,132	2502	2136
Solidarity	2390	4838	856	13,132	0	1425	682
Ill-effect	307	332	85	2502	1425	0	84
Credibility	208	86	82	2136	682	84	0
Total	7580	10,859	4344	27,819	23,323	4735	3278

Quadratic Assignment Procedure (QAP)

The quadratic assignment procedure (QAP) was used to assess the correlation between the WHO's agenda network (ie, matrix) and the agenda network of each Twitter user category. The QAP is a commonly used statistical test in social network analysis and network agenda-setting studies (eg, [49,50]) to calculate the Pearson correlation coefficient between 2 matrices [15]. The QAP indicates whether the correlation between 2 matrices or networks is statistically significant. Once 2 matrices are significantly correlated, the QAP regression test can be used to assess whether an independent variable can predict a dependent variable [18]. QAP linear regression was used to assess if the WHO agenda network could predict the agenda network of each Twitter user category.

Time Series Modeling (Granger Causality)

Overview

Granger causality has been used in previous agenda-setting studies to examine the relationship between media agenda and public agenda (eg, [51-53]). Granger causality was used in the current study to determine if the changes in 1 variable or time series (a series of data points over time) would impact the changes in another time series [54]. According to Granger [54], Y is said to cause X if the current or lagged values of Y can help to predict the future values of X. It determines whether the future value of a dependent variable can be predicted by the past values of an independent variable. Granger causality determines if there is a correlation between the past values of one variable and the present value of another variable [51].

In this study, Granger causality was used, for instance, to examine if "credibility" as a topic that was promoted by the WHO on Twitter predicted the future values of "credibility" in the tweets posted by the WHO's followers on Twitter. If the *P* value of a Granger causality test is less than .05, the independent variable is said to Granger cause or predict the value of the dependent variable [51]. For instance, it can be said Y Granger caused the values of X. However, it is important to note that Granger causality does not mean causation. In the Results and Discussion sections of this study, "influenced" or similar terms such as "affected" will be used instead of "Granger caused."

To investigate if each time series in the WHO (eg, teamwork) could predict the value of its corresponding time series (ie, teamwork) in any of the Twitter user categories, a vector autoregression (VAR) model was created. Additionally, to investigate if each time series in any of the Twitter user categories (eg, teamwork) could predict the value of its

corresponding time series (ie, teamwork) in the WHO, VAR models were created. Overall, 98 VAR models were created in this study.

To perform Granger causality, first, each variable (ie, topic) should be treated as a time series [53]. This study created time series for each of the 7 topics in the WHO and the 6 Twitter user categories, as explained in the following section.

Creating Time Series

To create time series for each topic, the frequency was calculated over time from January 1, 2020, to July 31, 2020, for both the WHO and the 6 categories of WHO followers on Twitter. For each topic, a time series with 182 records was created. Each record contained the frequency of the topic on a specific date. The initial analysis indicated that each time series had many zero number values, which skewed the data. Each zero number represents a lack of data for a given topic on a given date. The highly skewed data violated the normality assumption (see Testing the Residual Normality).

Additionally, having a wide range of values from zero to several hundred led to violating the heteroscedasticity assumption (see Testing the Residual Heteroskedasticity) in many cases. This study tried to resolve the normality and heteroscedasticity issues by aggregating the data to weekly data to reduce zeros. Therefore, for each topic, a time series of 26 weeks was created. Although the data were aggregated to weekly data, the first 2 weeks for most topics still included zero numbers, which again violated the normality assumption in some cases. Removing the first 2 weeks of June 2020 from the analysis resolved the normality issue. Therefore, the analysis was carried out on times series with 24 records (ie, weeks). As an example, the times series for the topics of WHO followers (ie, all 6 Twitter user categories together) are presented in Table S10 in Multimedia Appendix 1.

To perform Granger causality, VAR models should be created first, as explained below.

VAR Model

Overview

A VAR model is used to determine how 2 or more times series influence each other. In a VAR model, each time series is modeled as a linear combination of past values of itself and the past values of other variables [55]. For instance, a VAR model can be used to determine the relationship between the "teamwork" time series in the WHO (TW) and the "teamwork" time series in one of the Twitter user categories (TU) at time (t). Since there are 2 time series, 2 VAR models should be created: 1 for TW and 1 for TU. The VAR model for TW uses the past values of itself (TW) and the past values of the other variable (TU). In its simple form, the VAR model for TW and TU can be as follows, where TW_{t-1} and TW_{t-2} are the first and second lags of TW (first variable) and TU_{t-1} + TO_{t-2} are the first and second lags of TU (second variable). Each lag is x period ago. For instance, lag one is 1 time period ago or lag two is 2 time periods ago:

$$TW = TW_{t-1} + TU_{t-1} + TW_{t-2} + TU_{t-2}$$

$$TU = TU_{t-1} + TW_{t-1} + TU_{t-2} + TW_{t-2}$$

In a VAR model, a different number of lags can be considered for each variable. Thus, the optimal number of lags should be selected using a criterion. This study used the Akaike information criterion (AIC) to select the optimal lag for each VAR model, as explained in the following sections.

Lag Selection

The most commonly used criterion in lag selection in VAR models is the AIC [56,57]. The AIC specifies the number of lags to be used in a VAR model [57].

Testing the Stationary Nature

A major assumption underlying the VAR model and Granger causality is that time series must be stationary [54]; otherwise, they should be made stationary using the first or higher differences of the variables. A stationary time series has no systematic trend, meaning that its mean and variance do not change over time [58]. Nonstationary time series lead to incorrect inferences [55]. The Augmented Dickey-Fuller Test was used to test whether the variables were stationary [59].

A time series is stationary if the *P* value of the Augmented Dickey-Fuller test is less than .05. In this study, all variables in the WHO and the 6 Twitter user categories were nonstationary, which were transformed to become stationary through differencing (first or higher differences of the variables). Differencing is the process in which the differences between consecutive observations of a variable (ie, time series) are computed [60].

Testing the Residual Autocorrelation

A VAR model should be tested to determine if it "provides an adequate description of the data...In time series models, autocorrelation of the residual values is used to determine the goodness of fit of the model. Autocorrelation of the residuals indicates that there is information that has not been accounted for in the model" [61]. The Portmanteau test was used to check the presence of autocorrelation in the models. If the resulting *P* value in this test is less than .05, autocorrelation exists.

Testing the Residual Normality

Another assumption underlying the VAR model is that its residuals should be normally distributed; otherwise, inferences may be incorrect. The normality of residuals for all models was tested using the multivariate Jarque-Bera test. After performing this test, when the resulting *P* value is larger than .05, the residuals of VAR model are normal. In 4 cases, BoxCox transformation was used to make the VAR model normal (see Tables S22 and S24 in Multimedia Appendix 1).

Testing the Residual Heteroscedasticity

Another assumption underlying the VAR model is that there should be no heteroscedasticity in residuals. Heteroscedasticity refers to a condition in which the variance of the residual in a regression model varies widely. To test heteroscedasticity, the Autoregressive Conditional Heteroscedasticity-Lagrange Multiplier (ARCH-LM) test proposed by Engle [62] was used. Once the test is performed, if the resulting *P* value is larger than

.05, there is no heteroscedasticity in the data. No heteroscedasticity was observed in the VAR models.

Results

Results of Content Analysis

The first study objective was to identify the topics discussed by the WHO on Twitter about COVID-19. Using content analysis, 7 topics were identified inductively, including prevention (n=2430), solidarity (n=717), teamwork (n=601), surveillance (n=276), charity (n=243), ill-effect (n=213), and credibility (n=156). The numbers in the parentheses represent the frequency of each topic.

"Prevention" refers to the tweets posted by the WHO about how to avoid contracting the virus, including content about disinfecting surfaces, hand washing, wearing masks, vaccination, social distancing, isolation, and staying home. "Solidarity" refers to the tweets emphasizing the importance of unity, resilience, kindness, or supporting groups like refugees. "Teamwork" was another topic discussed by the WHO to highlight that countries, governments, organizations, and people should collaborate, coordinate, cooperate, be committed, and be accountable to control the pandemic. "Surveillance" emphasizes that governments should keep tracing, investigating, monitoring, and screening people and regions affected by the virus to take necessary actions. "Charity" refers to the donations, fundraisings, and financial support received by the WHO from governments, organizations, companies, celebrities, or other countries to fight against the virus. "Ill-effect" refers to the WHO's tweets about the consequences of the virus, such as disruptions in the economy, trades, health systems, mental health, abuse, and home violence during the quarantine. "Credibility" includes the tweets that demonstrate the importance of facts and bring people's attention to the rumors, misinformation, and fake information related to COVID-19.

This study also calculated the frequency of these topics for the 6 Twitter user categories and all of them together (hereafter, WHO followers). In the data set of WHO followers, prevention (228,700) had the highest frequency, followed by solidarity (72,192), charity (28,870), teamwork (23,879), ill-effect (15,151), surveillance (13,761), and credibility (11,726). Table S13 in Multimedia Appendix 1 presents the frequency of these topics in the WHO data set, WHO followers data set, and data set of each Twitter user category.

There was a strong tie between "prevention" and "solidarity" in in the matrix of the WHO and WHO followers and all 6 Twitter user categories. The strong tie between "solidarity" and "prevention" could indicate that the former is vital for the prevention of COVID-19 (or even treatment and response to the virus). For instance, to help prevent the spread of the virus, people should be united in following public health guidelines such as staying home, hand washing, and social distancing. These connections could represent social realities constructed around COVID-19 on Twitter and be transferred to the public agenda through network agenda setting [53].

Results of the QAP and QAP Regression

The QAP was used to measure the similarity between agenda matrices (networks). The QAP calculates the Pearson's correlation coefficient between the 2 matrices. The co-occurrence matrices for the WHO followers and WHO are presented in Table 2 and Table 3, respectively. Other matrices are presented in Tables S14 to S19 in Multimedia Appendix 1.

Degree centrality is an important concept in network analysis, which indicates how important a node within the network is and will be calculated by the total number of connections a node has [63]. In this study, nodes are topics. The most central topic on the WHO agenda network is "prevention," with 483 degrees of centrality, followed by "solidarity," with 420 degrees of centrality (see the row that shows the Total in Table 3). The most central topic on the WHO followers' agenda is "prevention," followed by "solidarity" and "charity" (see the row that shows the Total in Table 2). The 2 most frequently linked topics on the WHO agenda network and WHO followers' agenda network are "solidarity" and "prevention" because they have co-occurred 150 times.

Results of the QAP tests indicated a positive and high correlation between the WHO's agenda network and the agenda network of WHO followers (see Table 4). The QAP correlation tests also showed that the WHO agenda network and the agenda network of Twitter user categories are significantly correlated. According to network agenda setting [12], these results provide evidence that the salience of interrelationships among topics can be transferred from the WHO to the agenda of its followers on Twitter.

QAP linear regression was also carried out to assess whether the WHO agenda network (the independent variable) could predict the agenda networks of Twitter user categories (the dependent variables). As evident in Table 5, the WHO agenda network could predict all dependent variables, providing evidence that the agenda network of the WHO can impact the agenda network of Twitter user categories. For instance, the adjusted R-squared for politicians is 0.71, indicating that the WHO can explain 71% of the variance in the politicians' agenda network. The WHO also explains 62% of the variance in the network of WHO followers.

Table 3. The matrix of topics in the World Health Organization (WHO) data set on Twitter.

Topic	Teamwork	Charity	Surveillance	Prevention	Solidarity	Ill-effect	Credibility
Teamwork	0	51	23	110	118	9	6
Charity	51	0	1	37	90	3	2
Surveillance	23	1	0	101	14	2	0
Prevention	110	37	101	0	150	48	37
Solidarity	118	90	14	150	0	28	20
Ill-effect	9	3	2	48	28	0	3
Credibility	6	2	0	37	20	3	0
Total	317	184	141	483	420	93	68

Table 4. The quadratic assignment procedure (QAP) correlations between the World Health Organization (WHO) agenda network and agenda network of WHO followers and Twitter user categories.

Twitter user categories	Correlation (*r*) with the WHO agenda matrix	*P* value
Politicians	0.85	.001
Private sector	0.79	.01
Print and electronic media	0.77	.01
Legal professionals	0.80	.01
Health care professionals	0.79	.001
Academics	0.79	.01
WHO followers	0.80	.01

Table 5. The quadratic assignment procedure (QAP) linear regression for the World Health Organization (WHO) agenda network (independent variable) and networks of Twitter user categories (dependent variables).

Dependent variables	*F* statistic	Coefficient	Adjusted R-squared	*P* value (2-tailed)
Politicians	104.90	5.80	0.71	.002
Private sector	66.22	4.90	0.61	.005
Print and electronic media	60.43	9.50	0.59	.008
Legal professionals	71.95	3.20	0.63	.005
Health care professionals	61.70	13.60	0.60	.001
Academics	67.89	17.00	0.62	.004
WHO followers	70.01	51.20	0.62	.001

Results of Granger Causality

By performing 98 Granger causality tests, this study examined if the 7 topics in the WHO agenda Granger caused or predicted the future values of the topics in the tweets by the 6 types of Twitter users or vice versa. For instance, we tested to see if "teamwork" in the WHO Granger caused (ie, predicted the future values of) "teamwork" in health care professionals. We also investigated if "teamwork" in health care professionals Granger caused "teamwork" in the WHO. This study uses the term "influenced" or "affected" instead of "Granger caused" or "predicted the value of."

Tables 6 and 7 provide summaries of the Granger causality tests for the WHO and WHO followers. The summary of the Granger causality tests for other Twitter user categories is presented in Table S20 to Table S25 in Multimedia Appendix 1. Among the 7 topics, only "surveillance" in the WHO influenced "surveillance" in WHO followers ($F_{5,10}$=4.74, *P*=.01). On the other hand, among the 7 topics in the WHO followers, "charity" influenced "charity" in the WHO ($F_{6,6}$=7.48, *P*=.01), and "prevention" influenced "prevention" in WHO ($F_{5,10}$=4.69, *P*=.01).

The results indicated that the WHO influenced "surveillance" in politicians ($F_{6,6}$=5.13, *P*=.03) and "surveillance" in print and electronic media ($F_{5,10}$=9.33, *P*=.001). Additionally, the WHO influenced "ill-effect" in print and electronic media ($F_{5,10}$=4.02, *P*=.02), "credibility" in the private sector ($F_{5,10}$=7.12, *P*=.001), and "credibility" in academics ($F_{5,10}$=12.5, *P*=.001).

Twitter user categories also influenced several topics in the WHO's agenda: "credibility" in academics ($F_{5,10}$=6.10, *P*=.001),

"credibility" in politicians ($F_{5,8}$=8.06, P=.001), and "credibility" in the private sector influenced "credibility" in the WHO ($F_{5,10}$=4.33, P=.02). "Charity" in health care professionals ($F_{6,6}$=37.93, P=.001) and "charity" in academics influenced "charity" in the WHO ($F_{6,6}$=4.62, P=.04). "Prevention" in politicians ($F_{5,10}$=13.21, P=.001) and "prevention" in print and electronic media ($F_{5,10}$=15.04, P=.001) influenced "prevention" in the WHO. Print and electronic media influenced "surveillance" ($F_{5,10}$=3.7, P=.04) in the WHO. The private sector also influenced "ill-effect" in the WHO ($F_{5,10}$=7.35, P=.001). The WHO and legal professionals did not influence each other in any of the 7 topics.

Table 6. Granger causality test for World Health Organization (WHO) topics as independent variables and the topics of WHO followers as dependent variables.

Topics	*F* test	*P* value
Teamwork	0.04	.99
Charity	0.31	.91
Surveillance	4.74	.01
Prevention	0.27	.92
Solidarity	0.16	.97
Ill-effect	1.49	.29
Credibility	2.57	.11

Table 7. Granger causality test for the topics of World Health Organization (WHO) followers as independent variables and WHO topics as dependent variables.

Topics	*F* test	*P* value
Teamwork	0.49	.78
Charity	7.48	.01
Surveillance	1.70	.22
Prevention	4.69	.01
Solidarity	0.93	.50
Ill-effect	0.66	.66
Credibility	1.23	.38

Discussion

Contribution to Agenda Setting

Who Sets the Agenda?

Rarely any previous network agenda-setting study has investigated the relationship between a gatekeeper's agenda network and the agenda networks of various types of social media users, specifically in the context of public health emergencies.

This study found that, although there was a high correlation between the WHO agenda network and agenda network of Twitter user categories, the WHO influenced only some topics related to COVID-19 in the 6 Twitter user categories and vice versa (see the Granger causality test results). For instance, the WHO only influenced "ill-effect" in print and electronic media. Likewise, different Twitter user categories only influenced some (not all) topics in the WHO agenda.

It is hard to say who is leading the overall trends and topics related to COVID-19 on Twitter, mainly because social media provides unlimited space where various sources can interact with and impact each other [15]. However, it seems that there is less interaction between the WHO and some Twitter user categories during public health emergencies. For instance, neither legal professionals nor the WHO influenced each other. It is possible that legal professionals are not influenced by the WHO because they naturally often perform independently and remain impartial of external resources.

This study was not designed to measure the impact of the WHO on top influencers within each type of Twitter user or vice versa. Future research can explore who sets the agenda and who establishes the agenda-setting effect first. It is possible that the WHO originates the tweets about COVID-19, but the top influencers within each Twitter user category set the agenda or promote the agenda already set by the WHO, through retweeting the WHO's tweets.

Two-Way Agenda-Setting Effect

This study informs the network agenda-setting model by demonstrating that there can be a "2-way" relationship between the agenda of the WHO and its followers on Twitter. Vargo and Guo [53] also stated that the media agenda is reciprocal, in that network agenda setting is more complex than what past traditional agenda-setting studies have suggested. Evaluating the network agenda-setting effects on social media is more

complex than other platforms because, for instance, the agenda of WHO followers about COVID-19 could have been influenced by other resources too, such as the Centers for Disease Control and Prevention (CDC), providing initial evidence for a "multidirectional" network agenda setting effect.

Thus, the WHO does not seem to set the public agenda in a unidirectional nature. Neither the WHO nor Twitter user categories play a leading role on social media because they influence each other's agenda. It is possible that the WHO and different types of Twitter users pay attention to each other's agenda and interact with each other on some topics related to COVID-19 in a bidirectional way (or multidirectional way).

Vargo and Guo [53] also suggested that different news media pay attention to and are impacted by each other's agenda. Therefore, agenda setting is not always a 1-way communication mechanism from the mainstream media to the public. Actually, by making large-scale communication possible through social media, agenda setting is no longer only in the control of certain types of users [21]. An agenda can be created by laypeople on social media and shape the media agenda or vice versa [64].

Two Levels of the Agenda-Setting Effect on Twitter

A limitation in most network agenda-setting studies is that it is not clear if the public agenda is directly impacted by gatekeepers (eg, news media) or if other sources are involved too in influencing public agenda. For instance, the study by Vu et al [15] compared the most prominent issues in the national news media in the United States from 2007 to 2011 (extracted from the Pew Research Center's PEJ) with public opinion extracted from the Gallup Poll results (which has been surveying the public since 1939 about the most important problems facing the United States). In such studies, it is hard to determine if the national news media impact public agenda or whether other sources are also involved. One way to minimize this methodological limitation is to analyze the opinions of the users who follow a gatekeeper's account on social media or to analyze those users subscribed to a news media channel, such as CNN.

To fill this gap, this study investigated 6 Twitter user categories that follow the WHO Twitter account. From the results of this study, it can be concluded that there are 2 levels of the agenda-setting effect on Twitter—one at the aggregate level (see [19]), that is all social media users (eg, WHO followers on Twitter), and another at the Twitter user category level, in which each user category would be influenced differently by the agenda that is set by a gatekeeper. It may also be the case that the interaction among the agenda networks of all Twitter user categories builds the overall Twitter agenda network.

Practical Implications for the WHO

The results of this study can inform policy and be used to prepare for future pandemics in several ways: First, the WHO should define a clear strategy as how to use social media during pandemics to convey its messages to the public. The WHO should have a plan for what topics are more critical for the public during similar public health emergencies and work on transferring them to the public agenda.

Second, how messages are framed and presented to the public is important. The WHO can identify the needs of social media users and provide information-related benefits for them by framing and presenting more effective messages during public health emergencies. The strategic framing of public health messages can help the WHO to have greater impact on social media users. For instance, to design more effective messages and attract more audiences, the WHO can frame its messages by using hashtags that are popular among the public [65].

Third, using hashtags can lead to establishing and promoting important public health, social, political, or environmental agendas [24]. Hashtags have various functions, such as information searching and discovering, information organization, information distribution, information collection, and protecting information [26]. Our analysis of WHO tweets indicated that the WHO had used only a few hashtags related to COVID-19 in its tweets, including #safehands, #togetherathome, #handhygiene, #unitedagainstcoronavirus, and #stayhome. The WHO can use more hashtags during public health emergencies to convey its messages to people and create more engagement with them [65]. For instance, WHO could frame its message about "credibility" of information, using a hashtag like "#WHOFactChecker, as follows: "#WHOFactChecker recommends that you continue to take appropriate actions to protect yourself and those around you in the summer as there is no evidence that warm weather can kill the #COVID-19 virus."

Fourth, this study suggests that there could be a "2-way" or "multiway" agenda-setting effect on social media. For instance, the WHO, the CDC, the US government, and politicians could all interact with and influence each other's agenda. Therefore, in future public health emergencies, the WHO can determine which topics should be promoted on social media during different phases of a pandemic and collaborate with other gatekeepers such as the CDC to collectively make them salient in the public.

Limitations and Future Research

This study investigated only 6 Twitter user categories. There are other types of users following the WHO on Twitter, which can be studied in future work, such as artists and athletes.

Another limitation of using Twitter data is that Twitter users are not representative of the entire population.

The agenda of WHO followers about COVID-19 could have been influenced by other resources too, such as the CDC or any other government agency. Future studies should find ways to also take the effect sources other than the WHO into consideration. Surveys and interviews with WHO followers on Twitter can provide more insights into the WHO's impact on social media users' opinions about COVID-19.

The 7 topics promoted by the WHO on Twitter could possibly also be found in the messages shared on Twitter by other resources such as the US government. It could be the case that the 6 categories of WHO followers on Twitter were also propagating the agenda by other resources and not specifically that of the WHO. Therefore, although this study suggests that there can be a "2-way" agenda-setting effect on social media,

there could be a "multiway agenda-setting effect" on social media.

This study was limited to tweets written in English. Analyzing tweets in other languages may provide new insights on the agenda-setting effect of the WHO on Twitter.

Although this study analyzed a sample of tweets in the first 6 months in 2020, it is possible that the analysis of all the tweets posted by the WHO and its followers in the second half of 2020 could lead to different findings.

We attempted to make the Twitter user categories as exclusive as possible; however, there might be some overlap among different categories. For instance, a Twitter user account could fall into the academics and print and electronic media categories.

Future studies in this area can also determine to which social media user categories and top influencers within those categories the WHO should reach out if it wants to have more impact on social media users.

Future studies in this area can also study whether and to what extent the WHO has made strategic use of hashtags to communicate its messages about COVID-19 to the public. Future research can also investigate the role of hashtags used by the WHO in setting its COVID-19 agenda and the effect of that agenda on public opinion.

Although using social media data provides a legitimate method for studying the agenda-setting effect of WHO, more empirical research, including surveys and interviews, can be used in future studies to understand the role of the WHO in shaping public agenda on social media.

The results show that the WHO and Twitter user categories of different types could influence some of the COVID-19–related topics in each other; however, this study did not investigate whether different Twitter user categories would also impact each other's agenda network. This impact can be examined in future studies by exploring, for instance, the network agenda-setting effect between academics and health care professionals on Twitter.

Conclusions

This is among the first studies that demonstrate the presence of network agenda-setting effects between the WHO and its followers on Twitter, specifically different Twitter user categories. In line with the network agenda-setting model, this study showed that the topics promoted by the WHO about COVID-19, such as "credibility" or "surveillance," were linked together in a network.

This study extends theorizing on agenda setting by providing evidence that agenda–setting effects vary by different Twitter user categories and topics. For instance, the WHO only influenced "surveillance" in politics and print and electronic media, or health care professionals only influenced "charity" in the WHO, while the WHO and legal professionals did influence each other.

This study also extends theorizing on agenda setting by indicating that, while network agenda setting is known as a "1-way" model, there can be "2-way" or "multiway" effects of agenda setting on social media, because the influences between the WHO and Twitter user categories were reciprocal on Twitter.

Acknowledgments

This work is supported by the Open Publishing Support Fund (OPSF) at the University of Tennessee in Knoxville and the Big Data Health Science Center (BDHSC) at the University of South Carolina.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary files.
[DOCX File , 56 KB - jmir_v24i4e34321_app1.docx]

Multimedia Appendix 2
Agenda network of the World Health Organization (WHO).
[PNG File , 165 KB - jmir_v24i4e34321_app2.png]

Multimedia Appendix 3
The agenda network of WHO followers (six categories together).
[PNG File , 153 KB - jmir_v24i4e34321_app3.png]

Multimedia Appendix 4
The agenda network of politicians.
[PNG File , 146 KB - jmir_v24i4e34321_app4.png]

Multimedia Appendix 5
The agenda network of the private sector.
[PNG File , 149 KB - jmir_v24i4e34321_app5.png]

Multimedia Appendix 6
The agenda network of print and electronic media.
[PNG File , 149 KB - jmir_v24i4e34321_app6.png]

Multimedia Appendix 7
The agenda network of legal professionals.
[PNG File , 152 KB - jmir_v24i4e34321_app7.png]

Multimedia Appendix 8
The agenda network of health care professionals.
[PNG File , 138 KB - jmir_v24i4e34321_app8.png]

Multimedia Appendix 9
The agenda network of academics.
[PNG File , 152 KB - jmir_v24i4e34321_app9.png]

References

1. WHO Coronavirus (COVID-19) Dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2022-03-27]
2. Van Scoy LJ, Miller EL, Snyder B, Wasserman E, Chinchilli VM, Zgierska AE, et al. Knowledge, perceptions, and preferred information sources related to COVID-19 among central Pennsylvania adults early in the pandemic: a mixed methods cross-sectional survey. Ann Fam Med 2021 May 13;19(4):293-301 [FREE Full text] [doi: 10.1370/afm.2674] [Medline: 33985977]
3. Brennen JS, Simon F, Howard PN, Nielsen RK. Types, sources, and claims of COVID-19 misinformation. Reuters Institute. 2020 Apr 07. URL: https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation [accessed 2022-03-27]
4. Mohammadi E, Tahamtan I, Mansourian Y, Overton H. Identifying frames of the COVID-19 infodemic: thematic analysis of misinformation stories across media. JMIR Infodemiology 2022 Apr 13;2(1):e33827. [doi: 10.2196/33827]
5. Broniatowski DA, Kerchner D, Farooq F, Huang X, Jamison AM, Dredze M, et al. Debunking the Misinfodemic: Coronavirus Social Media Contains More, Not Less, Credible Content. Cornell University. 2021 Jun 01. URL: https://arxiv.org/abs/2007.09682v2 [accessed 2022-03-27]
6. Cinelli M, Quattrociocchi W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, et al. The COVID-19 social media infodemic. Sci Rep 2020 Oct 06;10(1):16598 [FREE Full text] [doi: 10.1038/s41598-020-73510-5] [Medline: 33024152]
7. Park HW, Park S, Chong M. Conversations and medical news frames on Twitter: infodemiological study on COVID-19 in South Korea. J Med Internet Res 2020 May 05;22(5):e18897 [FREE Full text] [doi: 10.2196/18897] [Medline: 32325426]
8. Christie TB. Framing rationale for the Iraq war. International Communication Gazette 2016 Jul 26;68(5-6):519-532. [doi: 10.1177/1748048506068728]
9. McCombs ME, Shaw DL. The agenda-setting function of mass media. Public Opinion Quarterly 1972;36(2):176-187 [FREE Full text] [doi: 10.1086/267990]
10. Guo L, Vu HT. Media vs. reality. ASJ 2018 Mar 30;2(1):3-24. [doi: 10.1075/asj.16020.guo]
11. McCombs M, Valenzuela S. The agenda-setting theory. Cuadernos.info 2007(20):44-51. [doi: 10.7764/cdi.20.111]
12. Guo L, McCombs M. Network agenda setting: a third level of media effects. Guo Lei. 2011. URL: http://www.leiguo.net/publications/guo_nas_2011_ica.pdf [accessed 2022-03-27]
13. Coleman R, McCombs M, Shaw D, Weaver D. Agenda Setting. In: The Handbook of Journalism Studies. Abingdon-on-Thames, Oxfordshire, England: Routledge; 2009:147-160.
14. Wanta W, Golan G, Lee C. Agenda setting and international news: media influence on public perceptions of foreign nations. Journalism & Mass Communication Quarterly 2016 Jun 25;81(2):364-377. [doi: 10.1177/107769900408100209]
15. Vu HT, Guo L, McCombs ME. Exploring “the world outside and the pictures in our heads”. Journalism & Mass Communication Quarterly 2014 Sep 09;91(4):669-686. [doi: 10.1177/1077699014550090]
16. Guo L, Vargo C. The power of message networks: a big-data analysis of the network agenda setting model and issue ownership. Mass Communication and Society 2015 Jun 26;18(5):557-576. [doi: 10.1080/15205436.2015.1045300]
17. Feezell JT. Agenda setting through social media: the importance of incidental news exposure and social filtering in the digital era. Political Research Quarterly 2017 Dec 26;71(2):482-494. [doi: 10.1177/1065912917744895]
18. Guo L. The application of social network analysis in agenda setting research: a methodological exploration. Journal of Broadcasting & Electronic Media 2012 Oct;56(4):616-631. [doi: 10.1080/08838151.2012.732148]

19. Guo L. Agenda-Setting: Individual-Level Effects Versus Aggregate-Level Effects. In: The International Encyclopedia of Media Effects. Hoboken, NJ: John Wiley & Sons, Ltd; 2017:1-13.
20. Yang X, Chen BC, Maity M, Ferrara E. Social Politics: Agenda Setting and Political Communication on Social Media. In: Spiro E, Ahn YY, editors. Social Informatics. SocInfo 2016. Lecture Notes in Computer Science(), vol 10046. Cham, Switzerland: Springer Publishing Company; 2016:330-344.
21. Boynton G, Richardson GW. Agenda setting in the twenty-first century. New Media & Society 2016 Jul 09;18(9):1916-1934. [doi: 10.1177/1461444815616226]
22. Skogerbø E, Krumsvik AH. Newspapers, Facebook and Twitter. Journalism Practice 2014 Sep 02;9(3):350-366. [doi: 10.1080/17512786.2014.950471]
23. Albalawi Y, Sixsmith J. Agenda setting for health promotion: exploring an adapted model for the social media era. JMIR Public Health Surveill 2015 Nov 25;1(2):e21 [FREE Full text] [doi: 10.2196/publichealth.5014] [Medline: 27227139]
24. Hemsley J. Followers retweet! The influence of middle‐level gatekeepers on the spread of political information on Twitter. Policy & Internet 2019 Mar 08;11(3):280-304. [doi: 10.1002/poi3.202]
25. Wang R, Liu W, Gao S. Hashtags and information virality in networked social movement. Online Information Review 2016 Nov 14;40(7):850-866. [doi: 10.1108/oir-12-2015-0378]
26. Potnis D, Tahamtan I. Hashtags for gatekeeping of information on social media. J Assoc Inf Sci Technol 2021 Mar 11;72(10):1234-1246 [FREE Full text] [doi: 10.1002/asi.24467]
27. Lee J, Xu W. The more attacks, the more retweets: Trump's and Clinton's agenda setting on Twitter. Public Relations Review 2018 Jun;44(2):201-213. [doi: 10.1016/j.pubrev.2017.10.002]
28. Conway BA, Kenski K, Wang D. The rise of Twitter in the political campaign: searching for intermedia agenda-setting effects in the presidential primary. J Comput-Mediat Comm 2015 May 11;20(4):363-380. [doi: 10.1111/jcc4.12124]
29. Blinder AS, Krueger AB. What does the public know about economic policy, and how does it know it? National Bureau of Economic Research. 2004 Sep. URL: https://www.nber.org/papers/w10787 [accessed 2022-03-27]
30. Toupin R, Millerand F, Larivière V. Scholarly communication or public communication of science? Assessing who engage with climate change research on Twitter. 2019 Presented at: 17th International Conference on Scientometrics and Informetrics (ISSI); September 6-9, 2019; Rome, Italy.
31. Hedima EW, Adeyemi MS, Ikunaiye NY. Community Pharmacists: On the frontline of health service against COVID-19 in LMICs. Res Social Adm Pharm 2021 Jan;17(1):1964-1966 [FREE Full text] [doi: 10.1016/j.sapharm.2020.04.013] [Medline: 32317154]
32. Laufer D. Academics engaging through the media-Insights from creating a monthly column on crisis management. Public Relat Rev 2021 Mar;47(1):101932 [FREE Full text] [doi: 10.1016/j.pubrev.2020.101932] [Medline: 32836706]
33. Engel-Rebitzer E, Stokes DC, Buttenheim A, Purtle J, Meisel ZF. Changes in legislator vaccine-engagement on Twitter before and after the arrival of the COVID-19 pandemic. Hum Vaccin Immunother 2021 Sep 02;17(9):2868-2872. [doi: 10.1080/21645515.2021.1911216] [Medline: 33970786]
34. Gruber DA, Smerek RE, Thomas-Hunt MC, James EH. The real-time power of Twitter: Crisis management and leadership in an age of social media. Business Horizons 2015 Mar;58(2):163-172. [doi: 10.1016/j.bushor.2014.10.006]
35. Luengo M, García-Marín D. The performance of truth: politicians, fact-checking journalism, and the struggle to tackle COVID-19 misinformation. Am J Cult Sociol 2020 Sep 28;8(3):1-23 [FREE Full text] [doi: 10.1057/s41290-020-00115-w] [Medline: 33014362]
36. Perry DC, Taylor M, Doerfel ML. Internet-based communication in crisis management. Management Communication Quarterly 2016 Aug 17;17(2):206-232. [doi: 10.1177/0893318903256227]
37. Drennan LT, McConnell A, Stark A. Risk and crisis management in the public sector, 1st edition. Abingdon-on-Thames, Oxfordshire, England: Routledge; 2012.
38. Agrawal S, De Smet A, Lacroix S, Reich A. To emerge stronger from the COVID-19 crisis, companies should start reskilling their workforces now. McKinsey & Company. 2020 May 07. URL: https://www.mckinsey.com/business-functions/organization/our-insights/to-emerge-stronger-from-the-covid-19-crisis-companies-should-start-reskilling-their-workforces-now [accessed 2022-03-27]
39. Tahamtan I. World Health Organization's Agenda For COVID-19 and Its Effects: A Network Agenda-Setting Study with Twitter Users. figshare. 2022 May 01. URL: https://tinyurl.com/2p8fzeen [accessed 2022-03-27]
40. Occupational Employment and Wage Statistics. US Bureau of Labor Statistics. URL: https://www.bls.gov/oes/current/oes_stru.htm [accessed 2022-03-27]
41. Wasike BS. Framing news in 140 characters: how social media editors frame the news and interact with audiences via Twitter. Global Media Journal, Canadian Edition 2013;6(1):5-23 [FREE Full text]
42. Ferrara E. What Types of COVID-19 Conspiracies are Populated by Twitter Bots? Cornell University. 2020 Apr 20. URL: https://arxiv.org/abs/2004.09531 [accessed 2022-03-27]
43. Kearney MW. tweetbotornot. GitHub. 2019 Nov 2. URL: https://github.com/mkearney/tweetbotornot [accessed 2022-03-27]
44. Davis CA, Varol O, Ferrara E, Flammini A, Menczer F. BotOrNot: A system to evaluate social bots. 2016 Presented at: WWW '16 Companion: Proceedings of the 25th International Conference Companion on World Wide Web; April 11-15, 2016; Montréal, Québec, Canada. [doi: 10.1145/2872518.2889302]

45. Krippendorff K. Content analysis: An introduction to its methodology. Thousand Oaks, CA: Sage Publications; 2018.
46. Riffe D, Lacy S, Fico F, Watson B. Analyzing media messages: Using quantitative content analysis in research. Abingdon-on-Thames, Oxfordshire, England: Routledge; 2019.
47. Lombard M, Snyder-Duch J, Bracken CC. Content analysis in mass communication: assessment and reporting of intercoder reliability. Human Comm Res 2002 Oct;28(4):587-604. [doi: 10.1111/j.1468-2958.2002.tb00826.x]
48. How to handwash. World Health Organization. URL: https://www.who.int/gpsc/5may/How_To_HandWash_Poster.pdf [accessed 2022-03-27]
49. Vargo CJ, Guo L, McCombs M, Shaw DL. Network issue agendas on Twitter during the 2012 U.S. Presidential Election. J Commun 2014 Mar 24;64(2):296-316. [doi: 10.1111/jcom.12089]
50. Yüksel E, Dingin A. Does “it” work in Turkey? A network agenda-setting research. Erciyes İletişim Dergisi 2020;7(1):435-450 [FREE Full text]
51. Tan Y, Weaver DH. Agenda-setting effects among the media, the public, and Congress, 1946–2004. Journalism & Mass Communication Quarterly 2016 Jun 25;84(4):729-744. [doi: 10.1177/107769900708400405]
52. Vargo CJ, Guo L, Amazeen MA. The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. New Media & Society 2017 Jun 15;20(5):2028-2049. [doi: 10.1177/1461444817712086]
53. Vargo CJ, Guo L. Networks, big data, and intermedia agenda setting: an analysis of traditional, partisan, and emerging online U.S. news. Journalism & Mass Communication Quarterly 2016 Dec 01;94(4):1031-1055. [doi: 10.1177/1077699016679976]
54. Granger CWJ. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 1969 Aug;37(3):424. [doi: 10.2307/1912791]
55. Rossi B, Wang Y. Vector autoregressive-based Granger causality test in the presence of instabilities. The Stata Journal 2019 Dec 18;19(4):883-899. [doi: 10.1177/1536867x19893631]
56. Ivanov V, Kilian L. A practitioner's guide to lag order selection for VAR impulse response analysis. Studies in Nonlinear Dynamics & Econometrics 2005;9(1):1 [FREE Full text] [doi: 10.2202/1558-3708.1219]
57. Liew VKS. Which lag length selection criteria should we employ? Economics Bulletin 2004;3(33):1-9 [FREE Full text]
58. Chatfield C, Xing H. The analysis of time series: an introduction with R. Boca Raton, FL: CRC Press; 2019.
59. Dickey DA, Fuller WA. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association 1979 Jun;74(366a):427-431. [doi: 10.1080/01621459.1979.10482531]
60. Bessler DA, Kling JL. A note on tests of Granger causality. Applied Economics 2006 May 24;16(3):335-342. [doi: 10.1080/00036848400000041]
61. Bose E, Hravnak M, Sereika S. Vector autoregressive models and Granger causality in time series analysis in nursing research: dynamic changes among vital signs prior to cardiorespiratory instability events as an example. Nurs Res 2017;66(1):12-19 [FREE Full text] [doi: 10.1097/NNR.0000000000000193] [Medline: 27977564]
62. Engle RF. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 1982 Jul;50(4):987. [doi: 10.2307/1912773]
63. Hansen DL, Shneiderman B, Smith MA, Himelboim I. Social network analysis: measuring, mapping, and modeling collections of connections. In: Analyzing Social Media Networks with NodeXL, 2nd Edition. Cambridge, MA: Morgan Kaufmann; 2019:31-51.
64. Russell Neuman W, Guggenheim L, Mo Jang S, Bae SY. The dynamics of public attention: agenda-setting theory meets big data. J Commun 2014 Mar 26;64(2):193-214. [doi: 10.1111/jcom.12088]
65. Tahamtan I, Potnis D, Mohammadi E, Miller LE, Singh V. Framing of and attention to COVID-19 on Twitter: thematic analysis of hashtags. J Med Internet Res 2021 Sep 10;23(9):e30800 [FREE Full text] [doi: 10.2196/30800] [Medline: 34406961]

Abbreviations

AIC: Akaike information criterion
ARCH-LM: Autoregressive Conditional Heteroscedasticity-Lagrange Multiplier
CDC: Centers for Disease Control and Prevention
QAP: quadratic assignment procedure
RQ: research question
VAR: vector autoregression
WHO: World Health Organization

Edited by C Basch; submitted 17.10.21; peer-reviewed by T Ndabu, M Ahmed Kamal, R Gore; comments to author 10.11.21; revised version received 17.01.22; accepted 08.02.22; published 26.04.22.

Please cite as:
Tahamtan I, Potnis D, Mohammadi E, Singh V, Miller LE
The Mutual Influence of the World Health Organization (WHO) and Twitter Users During COVID-19: Network Agenda-Setting Analysis
J Med Internet Res 2022;24(4):e34321
URL: https://www.jmir.org/2022/4/e34321
doi:10.2196/34321
PMID:35275836

Viewpoint

Digital Visual Communication for Public Health: Design Proposal for a Vaccinated Emoji

Tamara Sonia Boender[1], BSc, MSc, PhD; Noah Louis-Ferdinand[2], BA; Gideon Duschek[3], BA

[1]Surveillance Unit, Department of Infectious Disease Epidemiology, Robert Koch Institute, Berlin, Germany

[2]Voices for Vaccines, Decatur, GA, United States

[3]Ommado, Amsterdam Area, Netherlands

Corresponding Author:
Tamara Sonia Boender, BSc, MSc, PhD
Surveillance Unit
Department of Infectious Disease Epidemiology
Robert Koch Institute
Seestrasse 10
Berlin, 13353
Germany
Phone: 49 3018754
Email: BoenderS@rki.de

Abstract

In the 21st century, the internet and particularly social media have become essential platforms for the spread of health information (including misinformation and disinformation). One of the distinguishing features of communication on these platforms is the widespread use of emojis. Though seemingly trivial emojis are now used by many if not most public health figures and organizations alongside important health updates. Much of that information has had to do with vaccination. Vaccines are a critical public health tool but one surrounded by falsehoods, phobias, and misinformation fueling vaccine hesitancy. Part of that has to do with their lack of positive representation on social media (eg, the syringe emoji is a plain needle, which for many people is an uncomfortable image). We thus argue that vaccination deserves an entirely new emoji to communicate vaccine confidence and discuss a design proposal for a vaccinated emoji that has gained traction in the global public health community.

(J Med Internet Res 2022;24(4):e35786) doi:10.2196/35786

KEYWORDS

vaccination; emoji; design; science communication; infodemic management; vaccine confidence; digital communication

Digital Media and Vaccine Hesitancy

Communicating online has reached an all-time high for both scientists and the public, and has been catalyzed by public health and social measures in the context of the COVID-19 pandemic. The potential value of emojis in the context of medical communication and care has recently come into the limelight [1], given that their use is a defining feature of many of the online platforms. For instance, a recent survey of 68 million tweets found that one in five contained an emoji [2]. These emojis are often used to discuss current events, and indeed in the early phases of the COVID-19 pandemic, there was a clear increase on the *microbe* emoji and *face with medical mask* emoji [3]. As the BioNTech/Pfizer and Moderna mRNA vaccines began to roll out, there was then a subsequent increase in the use of the *syringe* emoji [4]. However, for many people, it did not feel representative since the syringe is meant to represent blood donation; influential medical doctor Jeremy Faust referred to it on Twitter as an “inaccurate bloody syringe” [3,5,6]. Several platforms responded by removing the blood from the *syringe* emoji. Yet this does not address the fact that for many people *needles themselves* come with a host of negative associations. Needle phobia affects some 25% of adults [7], often causing them to delay or turndown vaccination. Even for those without a phobia, the symbol is still tied to a painful experience. This is despite vaccination being a profoundly protective intervention central to public health. We therefore argue that there is currently no accurate emoji for vaccination.

This is no small problem. In 2019, the World Health Organization (WHO) listed vaccine hesitancy, or the reluctance to vaccinate despite availability, among the top 10 threats to

global health [8]. This was before the pandemic, when the lifesaving power of vaccines had come to be taken for granted, despite them being one of the best tools we have to prevent moderate and severe disease. Take for example the human papillomavirus vaccine that has the remarkable ability to prevent cancer. It has all but eliminated cervical cancer in women in the United Kingdom born after the mid-1990s. Yet, the WHO/United Nations Children's Fund (UNICEF) estimate that average performance coverage for the final dose among member states is only about 54% [9]. Lack of uptake year after year has surely caused a great deal of unnecessary suffering.

Vaccine hesitancy becomes an especially pressing problem in the context of pandemics, such as the 2009-2010 H1N1 (hemagglutinin type 1 and neuraminidase type 1) novel influenza A pandemic and the current COVID-19 pandemic. This is especially costly given that the COVID-19 vaccines offer an extraordinary degree of protection against severe clinical outcomes, including death. In 33 countries across the WHO European Region, an estimated 470,000 lives have been saved among those aged ≥60 years in 2021 alone [10]. Yet, only about one-third of the region was fully vaccinated by October 2021, with hesitancy being a major barrier to uptake [11].

Improving vaccine confidence, whether in Europe or elsewhere, will require more effective messaging. In the social media age [12], this means coming up with new tools for digital communication, supporting online risk communication and community engagement. We would specifically highlight the use of digital symbols to create shifts in attitude toward real-world phenomena. For instance, Plan International ran a campaign back in 2017 to create a "period emoji" [13]. Their stated intent was to reduce stigma and push back against harmful beliefs by coming up with a more modern way to communicate about menstruation. The product (designed alongside National Health Services Blood and Transplant) was the *drop of blood* emoji. In another case, Apple decided to change its *pistol* emoji to a *water pistol* emoji [14,15]. This was widely perceived to be an effort to reduce progun sentiment given that the company also vetoed a rifle emoji.

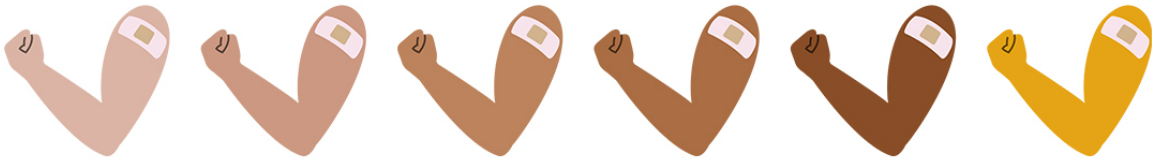
With regard to public health, the creation of emojis is not a new idea. For example, the *mosquito* emoji was proposed in 2016 in the context of the Zika virus outbreak [16]. The mosquito is the deadliest animal on earth due to spreading diseases like Zika, malaria, dengue, and yellow fever. The emoji helps health professionals to communicate with the public about the presence of mosquitoes and allows researchers to promote their work around mosquito-borne diseases more easily via social media and other digital platforms. Of note, the original *mosquito* emoji design has been added to the permanent collection of the Victoria and Albert Museum in London [17].

In this viewpoint, we build the case for a *vaccinated* emoji to meet the current public health need for better digital communication on vaccine confidence.

The Design Proposal: The Vaccinated Emoji

We present a design for a new emoji that accurately reflects the benefits of vaccination: the *vaccinated* emoji (Figure 1). Following the request for an alternative to the *syringe* emoji as posted on Twitter by author TSB on January 1, 2021 [18], she contacted author GD for his ideas for the design of a vaccinated emoji that conveys the message of both the act and sentiment of vaccination: the protection of being vaccinated. The design process unites the strengths of health sciences and epidemiology (TSB) with design (GD). As a first step, it made sense to remove the blood from the syringe and leave it clear [3]. However, the emoji would still contain a sharp needle, which for many people is an uncomfortable image. We wanted to create a symbol of hope, which represents protection against once-deadly diseases.

Figure 1. The proposed design of the vaccinated emoji.



We aimed for a generic arm as a base for the emoji, representing all ages, genders, and skin tones, based on the *flexed biceps* emoji. The *flexed biceps* emoji is a top-ranked emoji in its class (body parts) and often used to indicate strength, success, overcoming a struggle, or rolling up your sleeve to receive a vaccine [19]. We then added the *adhesive bandage* emoji over the deltoid to couple this symbol of strength with one of protection and care. Their combination enables the user to communicate the benefits of vaccination in an intuitive, playful way. The initial design of the *vaccinated* emoji by author GD, introducing the relatable and protective image of an arm with a bandage, was posted on Twitter by January 6, 2021 [20]. Of note, accurate placement of the bandage on the deltoid muscle has been challenging, and the bandage itself is also disproportionally large. However, disproportionate scale in the relationship between parts is typical for emojis. For example, the *crying face* emoji has a relatively large tear because it would not be visible otherwise. The emoji was designed using the graphics editor and design programs Photoshop (Adobe Inc) and Illustrator (Adobe Inc).

To enable universal and international access to the *vaccinated* emoji, TSB and GD have submitted the design proposal to the

Unicode Consortium—who formally manages emoji approval and implementation across devices, platforms, and languages—in 2021. The Unicode Standard can ensure global access to the emojis across all platforms through the associated license agreement [21,22]:

> *Submitter hereby acknowledges and agrees that they are submitting the Proposed Emoji for inclusion in the Standards, and further acknowledges and agrees that, if included in the Standards, the Proposed Emoji will be freely licensed by Unicode to all users around the world under the Unicode Licenses.*

The Unicode Emoji Subcommittee only selects a limited number of emojis that can be encoded and unfortunately did not select the vaccinated emoji proposal to move forward in 2021. Nonetheless, we are currently revising the emoji proposal for resubmission in 2022. It is not uncommon for emojis to be proposed multiple times before successful implementation, which was also the case for the *mosquito* emoji [23]. To ensure optimal advocacy for the vaccinated emoji, we also reached out to Emojination, a grassroots organization that has previously been successful at advocating for inclusive and representative emojis, including the *hijab* (person headscarf) emoji, as well as the infectious disease prevention–related emojis for *soap* and *microbe* [24].

To increase uptake where vaccines are available and affordable, thoughts, feelings, and social processes can motivate getting vaccinated [25]. People are more likely to get vaccinated if those in their social circle do so as well [26]. Sending an emoji to friends and family or adding one to a tweet is literally a matter of two clicks or phone taps. The vaccinated emoji is a low-key way of both seeing and signaling others about their support of vaccination and sharing the news that they just got vaccinated. The vaccinated emoji pairs well with ongoing trends such as the vaccine selfie, where social media users post photos of themselves at a clinic with a bandage over their arm [27]. Perhaps most importantly, the emoji can be used alongside personal stories, posts, and chats or group chats about vaccination to mark the event and reinforce one's reasons for getting vaccinated. This is important both because antivaccine messages are much more likely to make use of emotionally compelling stories [28] and users are more likely to remember and be deterred by stories of harm [29].

Communicating Vaccine Confidence

The vaccine emoji was created by TSB, a health scientist and field epidemiologist, and GD, a designer. After posting the design proposal to Twitter on January 1, 2021 [18], it quickly gained 40,000 impressions (user views). This prompted NRC—a Dutch newspaper with 4.1 million monthly readers—to pick up and report on the trend the following week [30]. The emoji has been encouraged by campaigns such as Voices for Vaccines *#WhyIVax* on Twitter where trusted doctors and public health professionals gave positive examples, often with a photo of a loved one they were protecting through vaccination. The idea behind the campaign—which author NLF played a role in—was to give other users the courage to speak up about vaccines. The powerful flex of the vaccine emoji conveys confidence in one's convictions and helps animate the important yet dry facts that comprise many provaccine posts [28]. People are more likely to engage in social media trends that take minimal effort. The intuitive nature of the action—though itself small—can have compounding effects due to the nature of virality and the fact that vaccination in one season predicts vaccination in the following seasons [26]. An emoji is the most organic, least heavy-handed way of getting that point across.

Later, NLF, who was a social media specialist for the Task Force for Global Health, began to help popularize the emoji. We reached out to a number of people through Twitter and rallied the global and public health community for *#WorldEmojiDay*, July 17. International attention skyrocketed, and within a day, the proposal had been endorsed by many, including senior leaders at the WHO including Dr Tedros Adhanom Ghebreyesus (Director-General) and Dr Maria van Kerkhove (COVID-19 technical lead) [31,32]. The Task Force for Global Health metrics show that the emoji had received 300,000 impressions on their page alone in a matter of 3 days, and one can see many tweets from ordinary users endorsing the emoji and posting selfies with flexed arms. Much of this support seemed to come from the fact that the emoji is an explicitly positive symbol. Hashtags, which can also be used to rally support for a cause, have in the past year often been a site of misinformation as they can be "hijacked" easily. They are also less intuitive with respect to what they mean or how they are to be used. The success of the *vaccinated* emoji is driven by its accessibility, ease of use, and universal appeal, winning praise from commentators as diverse as Heidi Murkoff (the author of the What to Expect When You're Expecting series of pregnancy guides) and Toyin Saraki (global health advocate, health care philanthropist, and the Founder-President of Wellbeing Foundation Africa). The emoji going viral on Twitter spilled over to media coverage in the United States [6] and the Netherlands [33]. We see this kind of organic support as crucial to the success of any public health symbol. We also see scientific communication as an underleveraged component of public health. This means we should follow organic interest to create public investment in health campaigns. The vaccine emoji does precisely that.

Since World Emoji Day 2021, the *vaccinated* emoji has become a fixture in the space of digital health advocacy. By August 2021, the emoji was selected for presentation at the Stanford Infodemic Conference on Social Media and COVID-19 Misinformation [34]. In October 2021, GD was commissioned by the German Federal Government to design an icon for their national contact-tracing app (Corona-Warn-App, Version 2.13 [35]) to visually strengthen the call to get vaccinated, drawing heavily upon the vaccine emoji. Both events speak to a growing recognition that the public health community must adopt symbols that meet the public where they are at. Though still considered trivial by some, our need to shift tactics is more than urgent in a time when vaccine hesitancy has hardened into refusal by some and a sense of defeat by the rest. Now more than ever, an intuitive participatory approach to health communications is needed to support vaccine confidence. At times of physical distancing and increased use of digital

communications, such as app groups and social media, emojis add to the sentiment of conversation around vaccination. Furthermore, the visual language of the *vaccinated* emoji (ie, showing an arm with a bandage) has been integrated into health promotion materials by national and international public health institutes with large reach, such as the Robert Koch Institute (Germany) [36], the US Centers for Disease Control and Prevention [37,38], and European Centre for Disease Prevention and Control [39], as well as the Google Doodle with the message "Get vaccinated. Wear a mask. Save lives" [40].

One year after its initial design, the proposal to add an emoji for vaccination gained international support from public health and medical professionals, national and international nongovernmental organizations, the infodemic management community, and the public. We endorse previous advocacy for expanding the available set of medical emojis and wish to amplify the need for emojis for public health by our proposed vaccinated emoji (Figure 1). Big tech companies have taken the first steps to provide appropriate visual digital communication tools for public health on the topic of vaccination, such as the *Vaccines for All* WhatsApp Sticker Pack in collaboration with the WHO [41] and the Twitter hashflag for the hashtag *#vaccinated* [42]. Furthermore, UNICEF is running a major social media campaign using the hashtag #vaccinated to help spread the word that COVID-19 vaccines are safe and effective by encouraging people to post pictures of their bandage [43]. These major social media campaigns highlight the need for a universal icon supporting this hashtag (potentially also used as a hashflag). The utility of a symbol cannot be divorced from its popularity, and from a communications perspective, the latter is a primary determinant of value.

Emojis are here to stay in our societies and have played a role not only in the fields of computer science and communication but also in marketing, behavioral science, linguistics, psychology, medicine, and education [44,45]. Importantly, social development is linked to emojis, such as the availability for emojis with different skin tones and genders, menstruation, and pregnancy. Encoding the vaccine emoji in the global Unicode Standard would help to make this symbol available for direct use for all to facilitate swift adaptation to ongoing (online) narratives regarding vaccination, which is much needed in the context of the pandemic and infodemic. Ultimately, the vaccinated emoji aims to be more than just an icon: a tool for public health.

Disclaimer

All rights of the *vaccinated* emoji (Figure 1 in this work) are preserved to the author GD. For the illustrative purpose of publication for this article, JMIR is granted permission to use this image (Figure 1) with open access. The *vaccinated* emoji design proposal will be resubmitted to Unicode in 2022, in agreement with the Emoji Proposal Agreement & License. In this legal agreement, we warrant that the proposed emoji is available for free and open licensing, and grant to the Consortium broad rights, specifically a nonexclusive, irrevocable, perpetual, worldwide, royalty-free license to encode the proposed emoji and to sublicense it under the Consortium's various open-source licenses.

Conflicts of Interest

TSB (@SoniaBoender) is a health scientist and field epidemiologist at the Robert Koch Institute (Germany's national public health institute), an institute within the portfolio of the Federal Ministry of Health. The development of the vaccinated emoji (@VaccineEmoji) and the publication of this viewpoint are not related to the author's employment at the Robert Koch Institute. NLF (@Michigan_Noah) is the communications coordinator at Voices for Vaccines, formerly interned at the Task Force For Global Health. GD (@GideonDuschek) is a freelance designer and has designed vaccination icons based on the vaccine emoji design for the Corona-Warn-App. TSB and GD plan to resubmit the described vaccinated emoji design to Unicode in 2022 and have high hopes that the committee will ultimately proceed with the proposal. If the emoji eventually will be encoded by the Unicode Standard, the authors will gain nothing from this apart from immortal geek fame.

References

1. Lai D, Lee J, He S. Emoji for the medical community-challenges and opportunities. JAMA 2021 Sep 07;326(9):795-796. [doi: 10.1001/jama.2021.8409] [Medline: 34547099]
2. Broni K. Emoji use in the new normal. Emojipedia. 2020 May 1. URL: https://blog.emojipedia.org/emoji-use-in-the-new-normal/ [accessed 2021-11-21]
3. Broni K. Vaccine emoji comes to life. Emojipedia. 2021 Feb 16. URL: https://blog.emojipedia.org/vaccine-emoji-comes-to-life/ [accessed 2021-11-21]
4. Coronavirus / COVID19 emoji list Updated 2021. Emojipedia. 2021. URL: https://emojipedia.org/coronavirus/ [accessed 2021-11-21]
5. Faust J. BREAKING: they finally got us a vaccine emoji that looks right!!! No more inaccurate bloody syringe! @TFGH @WHO @DrTedros @RickABright @KizzyPhD @NIHDirector @VaccineEmoji @ashishkjha @uche_blackstock @meganranney @Cleavon_MD @Craig_A_Spencer. Twitter. 2021 Jul 19. URL: https://twitter.com/jeremyfaust/status/1417136828654792708 [accessed 2021-07-19]

6. D'Ambrosio A. No more bloody syringe: new vaccine emoji gains key support - icon promotes positive image of COVID vaccine on social media, docs and public health experts say. Medpage Today. 2021 Jul 23. URL: https://www.medpagetoday.com/special-reports/exclusives/93728 [accessed 2021-07-23]
7. McLenon J, Rogers MAM. The fear of needles: a systematic review and meta-analysis. J Adv Nurs 2019 Jan;75(1):30-42. [doi: 10.1111/jan.13818] [Medline: 30109720]
8. Ten threats to global health in 2019. World Health Organization. 2019 Jan 10. URL: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019 [accessed 2021-11-21]
9. Falcaro M, Castañon A, Ndlela B, Checchi M, Soldan K, Lopez-Bernal J, et al. The effects of the national HPV vaccination programme in England, UK, on cervical cancer and grade 3 cervical intraepithelial neoplasia incidence: a register-based observational study. Lancet 2021 Dec 04;398(10316):2084-2092. [doi: 10.1016/S0140-6736(21)02178-4] [Medline: 34741816]
10. Meslé MM, Brown J, Mook P, Hagan J, Pastore R, Bundle N, et al. Estimated number of deaths directly averted in people 60 years and older as a result of COVID-19 vaccination in the WHO European Region, December 2020 to November 2021. Eurosurveillance 2021 Nov;26(47):2101021 [FREE Full text] [doi: 10.2807/1560-7917.ES.2021.26.47.2101021] [Medline: 34823641]
11. One billion COVID-19 vaccine doses administered in the WHO European Region – but risks for the unvaccinated leave no room for complacency. World Health Organization Regional Office for Europe. Copenhagen: WHO/Europe Press Office; 2021 Oct 21. URL: https://www.euro.who.int/en/media-centre/sections/press-releases/2021/one-billion-covid-19-vaccine-doses-administered-in-the-who-european-region-but-risks-for-the-unvaccinated-leave-no-room-for-complacency [accessed 2021-12-28]
12. Hammer CC, Boender TS, Thomas DR. Social media for field epidemiologists (#SoMe4epi): how to use Twitter during the #COVID19 pandemic. Int J Infect Dis 2021 Oct;110 Suppl 1:S11-S16 [FREE Full text] [doi: 10.1016/j.ijid.2021.05.035] [Medline: 34022332]
13. Our period emoji is appearing on phones!. Plan International UK. 2019. URL: https://plan-uk.org/act-for-girls/join-our-campaign-for-a-period-emoji [accessed 2022-07-07]
14. Water pistol. Emojipedia. URL: https://emojipedia.org/pistol/ [accessed 2022-02-03]
15. Kelly H. Apple replaces the pistol emoji with a water gun. CNN Business. 2016 Aug 2. URL: https://money.cnn.com/2016/08/01/technology/apple-pistol-emoji/index.html [accessed 2022-02-10]
16. Mackay IM. World's most dangerous animals set free… digitally. Virology Down Under. 2018 Mar 1. URL: https://virologydownunder.com/worlds-most-dangerous-animals-set-free-digitally/ [accessed 2021-12-28]
17. Messer A. Mosquito emoji in .png format. Victoria and Albert Museum. 2018. URL: https://collections.vam.ac.uk/item/O1432402/mosquito-emoji-in-aphelandra-messer/ [accessed 2022-01-28]
18. Boender TS. Instead of a blood dripping syringe, can we please have an accurate vaccin emoji? 2021 needs it! ❤ https://emojipedia.org/search/?q=Vaccin @Emojipedia #VaccinEmoji #vaccination #hope #COVID19 #SciComm. Twitter. 2021 Jan 21. URL: https://twitter.com/SoniaBoender/status/1344960382482456576 [accessed 2021-11-21]
19. Daniel J. The most frequently used emoji of 2021. Unicode. 2021. URL: https://home.unicode.org/emoji/emoji-frequency/ [accessed 2021-12-28]
20. Boender TS. Twitter. 2021 Jan 6. URL: https://twitter.com/SoniaBoender/status/1346749706458816512 [accessed 2022-01-28]
21. Guidelines for submitting Unicode® emoji proposals. Unicode. URL: https://www.unicode.org/emoji/proposals.html [accessed 2022-01-26]
22. Emoji proposal agreement and license. Unicode. URL: https://www.unicode.org/emoji/emoji-proposal-agreement.pdf [accessed 2022-01-26]
23. Emoji requests. Unicode. URL: https://unicode.org/emoji/emoji-requests.html [accessed 2022-01-29]
24. Lee J. Hijab and interracial couple emoji images acquired by Cooper Hewitt, Smithsonian Design Museum For Permanent Collection. Emojination. 2021. URL: http://www.emojination.org/press-releases [accessed 2022-01-29]
25. Brewer NT, Chapman GB, Rothman AJ, Leask J, Kempe A. Increasing vaccination: putting psychological science into action. Psychol Sci Public Interest 2017 Dec;18(3):149-207. [doi: 10.1177/1529100618760521] [Medline: 29611455]
26. Bruine de Bruin W, Parker AM, Galesic M, Vardavas R. Reports of social circles' and own vaccination behavior: a national longitudinal survey. Health Psychol 2019 Dec;38(11):975-983 [FREE Full text] [doi: 10.1037/hea0000771] [Medline: 31259597]
27. Ittefaq M, Ahmad Kamboh S, Abwao M. COVID-19 vaccine selfie: a modest endeavor to increase vaccine acceptance. Psychol Health 2021 Aug 02:1-5. [doi: 10.1080/08870446.2021.1957888] [Medline: 34339323]
28. Teoh D. The power of social media for HPV vaccination-not fake news!. Am Soc Clin Oncol Educ Book 2019 Jan;39:75-78 [FREE Full text] [doi: 10.1200/EDBK_239363] [Medline: 31099637]
29. Margolis MA, Brewer NT, Shah PD, Calo WA, Gilkey MB. Stories about HPV vaccine in social media, traditional media, and conversations. Prev Med 2019 Jan;118:251-256. [doi: 10.1016/j.ypmed.2018.11.005] [Medline: 30414396]
30. van der Stap S. Emoji voor een prik. NRC Handelsblad 2021 Jan 9:a4026867 [FREE Full text]

31. Adhanom Ghebreyesus T. This is a great idea, @TFGH! I hope next #WorldEmojiDay we will have a #VaccineEmoji and that people in all countries will be able to use it after getting their #COVID19 vaccine. #VaccinEquity. Twitter. 2021 Jul 18. URL: https://twitter.com/DrTedros/status/1416520738987728897?s=20 [accessed 2021-07-18]
32. van Kerkhove M. #VaccinesWork. Twitter. 2021 Jul 18. URL: https://twitter.com/mvankerkhove/status/1416817412058619904 [accessed 2021-07-18]
33. van Molenbeek M. Artsen houden pleidooi voor '#VaccineEmoji'. Nieuws Nieuws. 2021 Jul 19. URL: https://www.nieuwnieuws.nl/5154048/internet-houdt-pleidooi-voor-vaccineemoji/ [accessed 2021-07-19]
34. Boender TS, Duschek G. The #VaccineEmoji: a design proposal for a vaccinated emoji. Stanford, CA; 2021 Presented at: INFODEMIC: A Stanford Conference on Social Media & COVID-19 Misinformation; August 26, 2021; Online URL: https://www.stanfordinfodemic.org/
35. Corona-Warn-App Open-Source Project. URL: https://www.coronawarn.app/en/ [accessed 2021-12-28]
36. Robert Koch Institute. @RKI_fuer_Euch. Instagram. 2022 Jan 12. URL: https://www.instagram.com/p/CYonzl1L8Cv/?utm_source=ig_web_copy_link [accessed 2022-01-28]
37. CDC. New @CDCMMWR found when #Omicron first began circulating, adults who got a #COVID19 booster dose had 5x the protection against infection compared w/ adults who were not vaccinated. All eligible people should stay up to date w/ COVID-19 vaccination. Twitter. 2022 Jan 21. URL: https://twitter.com/CDCgov/status/1484589124392828934?s=20 [accessed 2022-01-21]
38. CDC digital media metrics. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/metrics/ [accessed 2022-01-26]
39. ECDC. What can YOU do to reduce the spread of #COVID19? Get vaccinated! Wear a #FaceMask Keep your distance Wash your hands Cough/sneeze into a tissue Ventilate the room We are all in this together - remember - no one is safe until all of us are safe. Twitter. URL: https://twitter.com/ECDC_EU/status/1484173807745073161?s=20 [accessed 2022-01-28]
40. Google. Get vaccinated. Wear a mask. Save lives. Google. URL: https://www.google.com/doodles/get-vaccinated-wear-a-mask-save-lives-january-21 [accessed 2022-01-28]
41. 'Vaccines for All' sticker pack + vaccine helplines. WhatsApp Blog. 2021 Apr 6. URL: https://blog.whatsapp.com/vaccines-for-all-sticker-pack-vaccine-helplines/ [accessed 2021-12-28]
42. Hay D. #vaccinated. Talk Forums. URL: https://talk.tf/hashflags/hf-vaccinated.html [accessed 2021-12-28]
43. #Vaccinated. UNICEF. URL: https://www.unicef.org/immunization/vaccinated [accessed 2022-01-25]
44. Bai Q, Dan Q, Mu Z, Yang M. A systematic review of emoji: current research and future perspectives. Front Psychol 2019;10:2221. [doi: 10.3389/fpsyg.2019.02221] [Medline: 31681068]
45. Cohn N, Engelen J, Schilperoord J. The grammar of emoji? Constraints on communicative pictorial sequencing. Cogn Res Princ Implic 2019 Aug 30;4(1):33 [FREE Full text] [doi: 10.1186/s41235-019-0177-0] [Medline: 31471857]

Abbreviations

H1N1: hemagglutinin type 1 and neuraminidase type 1
UNICEF: United Nations Children's Fund
WHO: World Health Organization

Edited by M Gisondi; submitted 29.12.21; peer-reviewed by R Rastmanesh, SS Amritphale; comments to author 11.01.22; revised version received 11.02.22; accepted 06.03.22; published 07.04.22.

Please cite as:
Boender TS, Louis-Ferdinand N, Duschek G
Digital Visual Communication for Public Health: Design Proposal for a Vaccinated Emoji
J Med Internet Res 2022;24(4):e35786
URL: https://www.jmir.org/2022/4/e35786
doi:10.2196/35786
PMID:35389363

Review

The Impact and Applications of Social Media Platforms for Public Health Responses Before and During the COVID-19 Pandemic: Systematic Literature Review

Dinesh Visva Gunasekeran[1*], MBBS; Alton Chew[2*], MBBS; Eeshwar K Chandrasekar[3], MD, MPH; Priyanka Rajendram[4], MBBS, MPH; Vasundhara Kandarpa[1]; Mallika Rajendram[1], MEd; Audrey Chia[1], BA, PhD; Helen Smith[5], BMedSci, BMBS, MD; Choon Kit Leong[1,6], MBBS

[1]National University of Singapore, Singapore, Singapore

[2]University College London, London, United Kingdom

[3]University of Rochester Medical Center, New York, NY, United States

[4]Ministry of Health Holdings, Singapore, Singapore

[5]Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, Singapore

[6]Mission Medical Clinic, Singapore, Singapore

[*]these authors contributed equally

Corresponding Author:
Dinesh Visva Gunasekeran, MBBS
National University of Singapore
10 Medical Drive
Singapore, 117597
Singapore
Phone: 65 67723737
Email: dineshvg@hotmail.sg

Abstract

Background: Social media platforms have numerous potential benefits and drawbacks on public health, which have been described in the literature. The COVID-19 pandemic has exposed our limited knowledge regarding the potential health impact of these platforms, which have been detrimental to public health responses in many regions.

Objective: This review aims to highlight a brief history of social media in health care and report its potential negative and positive public health impacts, which have been characterized in the literature.

Methods: We searched electronic bibliographic databases including PubMed, including Medline and Institute of Electrical and Electronics Engineers Xplore, from December 10, 2015, to December 10, 2020. We screened the title and abstracts and selected relevant reports for review of full text and reference lists. These were analyzed thematically and consolidated into applications of social media platforms for public health.

Results: The positive and negative impact of social media platforms on public health are catalogued on the basis of recent research in this report. These findings are discussed in the context of improving future public health responses and incorporating other emerging digital technology domains such as artificial intelligence. However, there is a need for more research with pragmatic methodology that evaluates the impact of specific digital interventions to inform future health policy.

Conclusions: Recent research has highlighted the potential negative impact of social media platforms on population health, as well as potentially useful applications for public health communication, monitoring, and predictions. More research is needed to objectively investigate measures to mitigate against its negative impact while harnessing effective applications for the benefit of public health.

(J Med Internet Res 2022;24(4):e33680) doi:10.2196/33680

KEYWORDS

digital health; social media; big data; population health; blockchain; COVID-19; review; benefit; challenge; public health

Introduction

Humans are an inherently social species, and the evolutionary and health benefits of this trait are well documented [1]. This predilection to form and live in groups is deeply rooted in human psychology. It follows that the fourth industrial revolution of digitization has brought with it social platforms as a technological embodiment of human interconnectedness and communication. Social media platforms bring content sharing and entertainment to the masses. They superficially bridge time and space to enable friendship, intimacy, and a sense of connection, consuming the time and attention of most individuals across all ages on a daily basis [2,3]. However, the COVID-19 pandemic has revealed the downside of this "online closeness," as with the greater ease of infectious disease transmission from physical closeness [4,5].

Social media platforms have drawn criticism for propagating misinformation and crowding out of public health communication [6,7]. As the pandemic rages on, it has exposed our limited knowledge regarding the potential health impact of these platforms, which have been a medium to propagate false information and widespread population anxiety [8,9]. It is timely, therefore, to investigate the benefits and drawbacks of social media on population health [10]. In this review, we aim to highlight a brief history of social media in health care, its negative public health impact that has marred outbreak responses, and its potential positive impact.

Methods

We searched electronic bibliographic databases, including PubMed, including Medline and Institute of Electrical and Electronics Engineers Xplore, from December 10, 2015, to December 10, 2020, with the following search terms: "((Social media[Title/Abstract]) OR (Social network[Title/Abstract]) OR (TikTok[Title/Abstract]) OR (Facebook[Title/Abstract]) OR (Instagram[Title/Abstract]) OR (Twitter[Title/Abstract]) OR (Baidu[Title/Abstract]) OR (Weibo[Title/Abstract])) AND ((Public health[Title/Abstract]) OR (Infectious Disease[Title/Abstract]) OR (Outbreak[Title/Abstract]) OR (Pandemic[Title/Abstract]) OR (COVID[Title/Abstract])) AND ((Intervention[Title/Abstract]) OR (Content analysis[Title/Abstract]) OR (Trial [Title/Abstract]) OR (Application[Title/Abstract]) OR (Health Promotion [Title/Abstract])) AND (English[Language])."

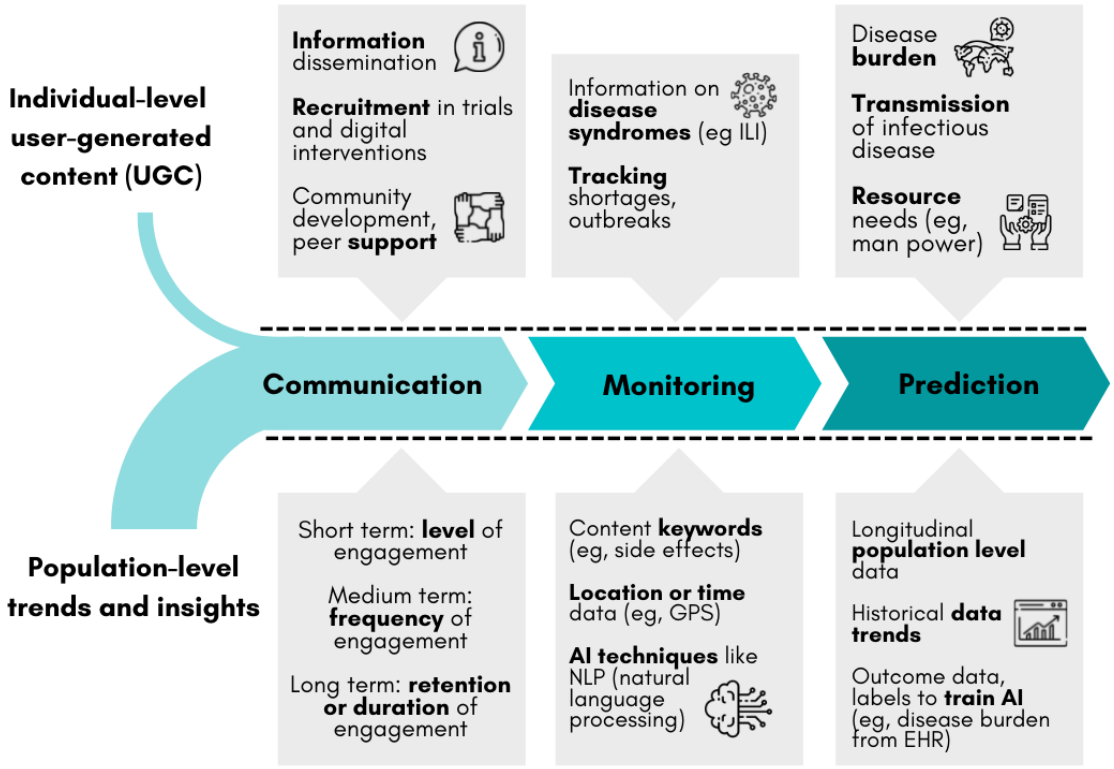
A total of 678 reports were identified. We screened the title and abstract of these reports to identify relevant English-language manuscripts for this review. The full text of selected manuscripts and their reference lists were analyzed thematically on the basis of the thematic paradigm of social media apps for public health communication, monitoring, and predictions. This analysis was conducted by a multidisciplinary panel of clinicians, researchers, public health specialists, and professors from business and medical schools to provide a holistic assessment of the published literature. The findings of this panel based on the reviewed studies are described using a narrative review approach in accordance with specific issues described in the literature, which have a positive or negative impact on population health, in order to inform future public health responses.

Results

Social Media: A Brief History Before the COVID-19 Pandemic

Prior to the COVID-19 pandemic, the possibilities of scalable public health promotion through leveraging the network effects of social media had garnered praise from the academic community. The effectiveness of these platforms for dissemination of information, conduct of digital interventions, or individual campaigns can be evaluated at three levels of chronology. These include the short-term using level of engagement (frequency or duration a platform is accessed each day, number of reactions or shares to content, etc), medium-term with frequency of engagement (daily or monthly active users, etc), and long-term based on retention or duration of engagement (adherence to or compliance with digital interventions) [11,12]. The ubiquity of social media platforms enables many public health applications including the communication of public health messages, real-time monitoring of population health, and potential predictions such as infectious disease outbreaks [13]. Descriptions of these applications in existing literature are summarized thematically and depicted in Figure 1.

Figure 1. Evaluating the impact and applications of social media in public health. AI: artificial intelligence, EHR: electronic health record, ILI: influenza-like illness.



Communication: Digital Public Health Promotion

The exponential potential of social media platforms for information dissemination has been strategically used for positive impact in the past [3]. They can be applied to reinvigorate public health promotion efforts and raise awareness about diseases, as exemplified by the "ALS Ice Bucket Challenge" in 2014 [14]. Other such health promotion campaigns include smoking cessation campaigns such as Tweet2Quit [12], and the #smearforsmear campaign to raise awareness about cervical cancer screening [15]. These initiatives capitalize on the network effects of social media to amplify the impact of web-based public health interventions. This is achieved by leveraging visibility (through search or content), peer-to-peer advocacy ("word of mouth"), or contextual paid advertising, the fundamental pillars of marketing digital initiatives [16,17].

The Tweet2Quit initiative attracted considerable attention to public health promotion using social media following a randomized controlled trial of a digital intervention using Twitter to help smokers abstain from their habit. It recruited users into twitter groups of 17-20 participants and encouraged smoking cessation by seeding conversation topics for users using automated messages to each group. These messages were aligned with clinical practice smoking cessation guidelines. The messages served as a conversation starter for users to provide encouragement for others, forging camaraderie as they embarked on their arduous smoking cessation journeys. The digital intervention was found to be more effective than Nicotine patches and a quit smoking website in this study [18].

Monitoring: Precision Public Health

The epidemiological value of social media applications includes surveillance of information, disease syndromes, and events (outbreak tracing, needs or shortages during disasters) [19]. The benefit of social media is that it provides real-time big data rapidly to epidemiologists from millions of users worldwide. The utility of epidemiological monitoring using social media during public health emergencies was well illustrated in the H7N9 avian influenza A virus outbreak using user-generated content (UGC) in the Sina microblog and the daily Baidu Activity Index [20]. This facilitated network monitoring for rapid information collection, allowing officials to disseminate public health communication in a relevant and timely manner when information-seeking behavior was at its highest. These methods were reproduced following subsequent outbreaks, highlighting potential rapid surveillance of population reactions to outbreaks and informing public health responses [21,22].

During the initial onset of an outbreak, uncertainty promotes fear among members of the public, who become desperate for more information. During the H1N1 outbreak: internet attention peaked in the first 3 days before dwindling as information saturation set in [23]. Public attention was also positively correlated with the case fatality rate and geographical advancement of the outbreak. This suggests that public health communication should use such critical features and time points in an outbreak to draw attention to accurate information. To achieve this, social media seems to present a potential tool for governments to (1) rapidly assess public reaction to an outbreak, (2) identify critical time points and topics that need to be

addressed, and (3) rapidly disseminate vital public health communication during outbreaks.

During the 2009 H1N1 outbreak, real-time monitoring using Twitter was clearly demonstrated [23]. Tweets containing relevant keywords such as "flu," "swine," and "Tamiflu" among others were extracted along with geolocations and time stamps. These were largely found to be posted by users in Twitter "live" (ie, real-time) and often contained information about the users' condition or symptoms. The researchers then applied a machine learning method to create a real-time model for the estimation of disease activity from the data, and demonstrated positive correlation with the national and regional prevalence of influenza-like illness reported by the US Centre for Disease Control and Prevention's (CDC's). They proposed other potential applications of social media using similar techniques, including surveillance for treatment side effects and shortages in medical supplies.

Predictions: Public Health Forecasting and Planning

Infodemiology (ie, information epidemiology) entails methods which analyze trends in web-based health data for applications, such as policy making [24,25]. On the other hand, infoveillance (ie, information surveillance) is the detection of events using web-based data, which can be faster than traditional surveillance methods [26,27]. Earlier studies have successfully illustrated the use of social media microblogs and geographical locations to track infectious disease outbreaks in many countries [28]. The authors searched Twitter for keywords such as "headache," "fever," and "runny nose" among others, mapping the locations of these tweets against the results of CDC's surveillance system FluView. They then modeled the potential spread of influenza on the basis of airline traffic and demonstrated predictions of influenza outbreaks a week in advance. The technical demonstration of these capabilities was a prelude to their future application during the current pandemic, which are discussed in a later section on social media and the COVID-19 pandemic.

Social Media and Infodemics

Although social media has the potential for positive public health utility, it can also amplify poor quality content [3]. Public fear and anxiety are known to be heightened by sensational reporting in the media during outbreaks, a phenomenon heightened by the ease of sharing on social media. These trends were described during previous outbreaks, such as the prominence of risk-elevating messages in American media during the Ebola outbreak [29]. During the COVID-19 pandemic, cross-sectional surveys in Russia, Bangladesh, and Iraq found elevated baseline levels of anxiety in individuals with higher levels of consumption of COVID-19–related news [8,30,31].

Similar associations between media consumption and mental health disorders have been reported during the COVID-19 pandemic and were worsened by poor quality of information dissemination among quarantined undergraduates in France [32]. The sharing of poor-quality information during outbreaks was also highlighted in an earlier infodemiological analysis of public reactions to the Zika epidemic between 2015 and 2016. Reliable sources, including the World Health Organization, had accounted for less than 0.1% of all highest-ranking content by dissemination, while over a quarter originated from social media like Facebook and Twitter. A similar study conducted during the COVID-19 pandemic suggested improved prominence of reliable sources [33], which may be driven by technology and public health partnerships for education campaigns [3,34].

The Impact of Social Media During the COVID-19 Pandemic

Despite the negative impact of social media in propagating "infodemics," it also provides a reservoir of UGC as individuals share a range of topics from emotions to symptoms [35]. The COVID-19 pandemic has shed light on various public health applications of social media that were developed and piloted post hoc using retrospective data such as trends in UGC following earlier outbreaks [36]. However, the potential real-world impact of several applications of social media platforms as digital health interventions have been hindered by a lack of stakeholder engagement and barriers to adoption [37]. The following section summarizes descriptions of social media apps for public health communication, monitoring, and predictions during the COVID-19 pandemic.

Communication During the COVID-19 Pandemic

The volumes of fear-driven information sharing at the beginning of the pandemic overwhelmed individuals and the capacity of regulators in many regions [7]. Some distributors capitalized on public anxiety, using fear mongering and predatory sales tactics for fraudulent medical products, a situation worsened by high-profile figures touting baseless claims [6,38]. Moreover, professionals were bombarded with rapidly evolving advisories as public health organizations and academia scrambled to process the flood of scientific reports of variable quality [39]. These challenges have presented an unprecedented need for scalable tools that detect trends in information sharing, develop targeted public health communications, and facilitate their dissemination—both to members of the public as well as frontline health care workers [5].

Fortunately, new methods using topical modeling and engagement metrics in social media were available to allay the concerns of the public and provide updated information to health care professionals. These leveraged application programming interfaces (APIs) of platforms such as Twitter or Weibo to identify trends in content sharing to inform public health communications [34,40]. These methods used such APIs to filter UGC on the basis of predetermined hashtags as well as identify temporal or geographical trends in information sharing, topical modeling, and engagement using the natural language processing branch of artificial intelligence (AI) [33]. Reports have also described pairing these techniques with sentiment analysis using Python textblob library or Valence Aware Dictionary and sEntiment Reasoner to provide a barometer of public sentiment [41,42].

Finally, social media has also been applied as a tool for grassroots health promotion initiatives [3,43]. For example, many US physicians actively used Twitter for health promotion during the COVID-19 pandemic [44], and in Singapore, the Government applied various social media platforms for health

promotion initiatives [5]. In Italy, these platforms were even used by health care professionals to share practice updates and scientific information among one another [45]. Despite these benefits, the relative lack of legitimate voices has been thought to enable the propagation of misinformation in social media, such as the purported association between the COVID-19 pandemic and 5G networks in the United Kingdom [46]. Nonetheless, sometimes misinformation has been aggravated by academics or health care providers stepping outside their areas of expertise in well-meaning attempts to help educate the public [47].

Monitoring Applications During the COVID-19 Pandemic

Comprehensive surveillance is vital during infectious disease outbreaks to monitor compliance and effectiveness of measures such as social distancing [48]. This allows the extensiveness of these measures to be tailored, to balance competing individual freedoms, health, and economic priorities for public benefit [48]. During the COVID-19 pandemic, methods leveraging social media, which were validated in previous outbreaks, were applied prospectively to inform advisories for both health care practitioners and public health administrators. Social reactions on Instagram and Twitter can be used as proxies for outbreak monitoring and assessment of public health measures for outcomes such as reduction in the basic reproductive number of COVID-19 with social distancing, as demonstrated in the United States [49].

This underscores the importance of investigating the relationship between web-based and offline behavior for translatable population health benefits. Digital data from social media platforms has also been used to detect predatory sellers, counterfeit health products, and unapproved products with questionable claims [50]. These emerging applications can provide governments and health authorities effective tools for real-time monitoring of public health measures, targeted law enforcement activities, and developing protective measures for public safety.

Prediction Techniques Applied During the COVID-19 Pandemic

The AI and regression techniques applied for the abovementioned real-time monitoring applications were based on cross-sectional data that became available during the pandemic. However, increases in computational power and availability of large, longitudinal data sets have paved the way for applications of big data from social media for future outbreak forecasting among other predictions. Applications that predict the potential number of cases during the COVID-19 outbreak used social media search indexes (SMSI) for keywords such as “dry cough,” “fever,” “coronavirus,” and “pneumonia” on platforms such as Baidu, where a significant correlation between new COVID-19 cases and SMSI findings have been reported [51].

Researchers have even developed and demonstrated such capabilities during the COVID-19 pandemic to accurately predict the burden of incident cases 2 weeks ahead of official sources [52]. This was achieved by applying the machine learning (ML) branch of AI to the social media posts of over 250 million users of the Weibo social media platform, based on self-reported symptoms and illness in UGC. The scale of big data and predictive value of novel such approaches represent a paradigm shift for public health capabilities, enabling anticipatory strategies and agile infection control responses driven by real-world data during an evolving threat [9].

However, it is worth noting that the CDC’s prediction initiative COVID-19 forecast hub has indicated that methods using social media big data have underperformed traditional methods such as the Susceptible-Exposed-Infectious Removed (SEIR) model when applied for forecasting. Ultimately, further research is necessary to fine-tune these novel techniques, and researchers may find that applying social media apps together with existing traditional modeling paradigms such as the SEIR may yield superior results. Limitations of existing modeling approaches include their primary focus on human-human transmission, along with difficulties modelling environmental transmission from fomites as well as variations in transmissibility. The latter is particularly important in a new public health emergency with growing awareness over time and public health communication such as that to encourage the adoption of hygiene measures. Public health organizations may also consider funding this research for capacity building to evaluate how these tools can be applied to enhance resource allocation during future health crises.

Discussion

Principal Findings

The COVID-19 pandemic has exposed the public health risks of unchecked health information–sharing on social media. It has also highlighted the pivotal role of human behavior in epidemic risk, prevention, and control [53]. This review has highlighted the potential negative impact of social media platforms on population health, as well as their useful public health applications for communication, monitoring, and predictions. Strategic planning for outbreaks should specifically explore leveraging the benefits of social media as potential tools for public health responses, as well as specific measures to mitigate against its potential drawbacks [9]. This includes planned behavioral and social communication to mitigate the infodemic, along with monitoring and predictive applications identified in this review. To be most effective, this needs to be developed using a participatory approach involving members of target populations [53,54].

The literature regarding social media apps for public health communication before the COVID-19 pandemic highlighted that digital behavioral modifications can be less time-consuming and less costlier than traditional approaches implemented using offline channels, such as patient support groups [55]. Through the rapidity and ease of recruitment facilitated by social media, these studies have shed light on the potential for social media to be applied in a scalable manner for behavioral modifications through peer support and networks. Other benefits include ease of monitoring and withdrawing these trials, as well as low inherent risks to participants given the use of platforms that are already widely used. Nonetheless, as with any patient-directed

digital intervention, risk mitigation measures such as methodology and ethics review are critical to ensure participants understand the intervention, potential benefits, and its risks [56].

Although similar applications of social media for communication were effectively applied to amplify public health messages during the COVID-19 pandemic, they were also used by some to perpetuate the spread of misinformation, thus marring its positive impact [47]. Therefore, researchers are now calling for novel approaches such as provider-moderated online health communities (OHCs), to leverage the utility of social networks as scalable channels for the dissemination of information, with added controls such as expert peer-verification to amplify benefits over its risks [57]. OHCs have been developed by social entrepreneurs to connect stakeholders such as health experts, providers, caregivers, and patients on a common platform. Besides OHCs, social entrepreneurs have also devised frugal solutions to successfully address a range of public health problems such as last-mile health, sanitation, and capacity building of health care workers [58]. Other relevant applications for health include disease-focused virtual communities, such as ParkinsonsNet, topical forums within Reddit, and the Psoriasis MSN or Google groups [59,60]. General health forums such as WebMD communities have also been previously described [61]. New OHCs have since been created for clinical practice updates, ranging from individual clinical discussions to entire virtual conferences. These include the inaugural virtual Primary Care Grand Conference (AGC) launched on the internet in Singapore in 2020 by an OHC, with participation of stakeholders from various sectors—clinical, allied health, political, and social sectors—to provide comprehensive updates on trends in disease presentation and administration [62].

New trends in personal content creation are constantly emerging, such as video logging (“vlogging”) using platforms such as TikTok or modules of established platforms such as “Stories” in Instagram [63]. These present new challenges for regulation. Content moderation is especially challenging given the size of video content and configurations with automated purging after a brief interval during which many impressions can be formed. This can happen rapidly at scale, creating a narrow window for enforcement. Users of these platforms are disproportionately represented by youth, and their demand for health-related content producers exceed their supply, exposing vulnerable users to content from sources of uncertain reliability [63].

Nonetheless, various reports of these public health responses to the COVID-19 pandemic, which applied social media for positive impact signal a future in which these platforms can be used to address new public health threats. Social media data can be combined with other sources of publicly available and digital behavioral data to improve the accuracy of existing approaches for various public health applications. These include analysis of UGC in open social media platforms such as Twitter, as well as internet search trends in search engines using ML. This has been demonstrated for applications such as monitoring for influenza surveillance [64]. Data from social media applications for public health monitoring in this manner can be used for operational planning. This is particularly useful when triangulated with other sources of data pertaining to web-based behavior, such as search trends, which have been used to predict future requirements for telehealth capacity [65]. These were also introduced during the COVID-19 pandemic for the monitoring of social distancing measures [49] and predictions of outbreaks [66].

However, the effectiveness of applying these tools at a population level has yet to be formally evaluated [36]. Moreover, given the heterogeneity between these social media platforms and within them as they evolve over time, another future area for further research would be to evaluate the public health implications of specific modules or social media functions. These remain key priorities for future research to improve our understanding of this new digital domain within public health. New variables of interest in outbreaks, such as attention saturation with temporal and topical variation, were already described in one such multinational study corroborating multimodal content from Reddit, Wikipedia, and news media with epidemic progression [67]. Furthermore, data from electronic medical records (EMRs) have been previously applied through global disease registries to improve our understanding of infectious diseases [68,69]. The incorporation of data from EMRs to triangulate with publicly available data can also improve the validity of these tools [70]. However, each additional source of data carries privacy concerns that must be addressed [71].

Future research is needed to develop scalable methods to mitigate against the risks of “online closeness.” Fortunately, solutions such as provider-moderated OHCs have emerged as potential tools to counter web-based medical misinformation, with applications described in fields such as psychiatry [57] and anesthesia for chronic pain management during the COVID-19 pandemic [72]. Moreover, promising solutions to automate the processing of big data using the deep learning branch of AI, such as long short-term memory or gated recurrent unit neural networks are emerging [60,73]. Finally, progress in cryptography has resulted in the successful incorporation of blockchain in digital platforms [74]. These are distributed databases with security configurations accommodating smart contracts, and programmable permissions for data access [75,76]. These configurations help address privacy concerns with data accession for the applications of social media big data by only revealing aggregated trends for public health planning while concealing individual identity through cryptography, or enabling the incorporation of individual consent for data accession transparently using smart contracts. Relevant applications of blockchain in health care during the COVID-19 pandemic include real-time tracking of drug delivery following telemedicine services [74]. The field of blockchain is still evolving with new potential applications such as social money solutions that provide financial incentives for content creation [77], and could be applied to reward creators of reliable content to combat web-based medical misinformation. Other potential health care applications include programmable patient consent or compute-to-data solutions for privacy-preserving data applications [78,79], data storage in medical devices [80], and queries from pharmacogenomics databases [81].

The incorporation of these various technologies with social media platforms may eventually contribute to a “learning” digital public health system in future, that can scale up and

improve existing methods for targeted communication, monitoring, and predictions [60]. However, improved study design and more empirical investigations of specific digital interventions using social media platforms are needed to develop and validate targeted strategies for key responses. For instance, recent reports described the use of programmed reminders to prompt individuals to consider accuracy of UGC [82]. Others forged partnerships between formal news media and social media influencers for targeted health promotion campaigns [43].

Finally, strategies for implementing these tools in health care macrosystems will also need to be developed. Examples of these include implementation of these tools using the lighthouse and safety net operational models for remote monitoring solutions [37]. Ultimately, more research on the link between health and human behavior is urgently required.

Conclusions

The pandemic has had a massive human toll and economic impact [83]. However, even with the availability of vaccines, new challenges remain including problems of logistics, distribution to low-income nations, and antivaccine activism [6,84]. Fortunately, social media platforms have emerged as new digital tools for public health professionals and providers. This review has highlighted existing and developing applications of social media for public health communication, monitoring, and predictions. These tools were sharpened by our experience with the COVID-19 pandemic and will likely have increasing prominence in responses to future public health threats. However, we also identified a need for greater pragmatic research for these applications of social media in order to better inform public health responses.

Conflicts of Interest

DVG reports equity investment in AskDr, Doctorbell (acquired by MaNaDr Mobile Health), VISRE, and Shyfts. A Chia reports equity investment in Bodhi Health Education. The remaining authors have no relevant financial declarations. A Chew and VK are medical students on research attachment with DVG. DVG is a senior lecturer (Medical Innovation) at the National University of Singapore (NUS), and physician leader (Telemedicine) in Raffles Medical Group (RMG). EKC and PR are actively practicing clinicians trained in Public Health. MR is a tutor of Academic English at the Center for English Language and Communication, NUS. A Chia is associate professor of Management and Organisation at the NUS Business School with joint appointment at the Yong Loo Lin School of Medicine, NUS. HS is dually accredited in General Practice and Public Health, has practiced extensively in Canada and the United Kingdom, and is presently appointed as a professor at Lee Kong Chian School of Medicine (LKCMedicine), Singapore. CKL is trained in Family Medicine and Public Health and has practiced extensively at the Asia and Mission Medical Clinic, and contributes to the development of health policies and infectious disease guidelines in Singapore. CKL is also appointed as an adjunct assistant professor at Duke-NUS and the Yong Loo Lin school of Medicine, NUS, and adjunct clinical instructor at LKCMedicine.

References

1. Oxford University Press. Interlend Doc Supply 2002 Dec;30(4). [doi: 10.1108/ilds.2002.12230dab.004]
2. Marcum CS. Age Differences in Daily Social Activities. Res Aging 2013 Sep;35(5):612-640 [FREE Full text] [doi: 10.1177/0164027512453468] [Medline: 25190898]
3. Chew AMK, Gunasekeran DV. Social Media Big Data: The Good, The Bad, and the Ugly (Un)truths. Front Big Data 2021;4:623794 [FREE Full text] [doi: 10.3389/fdata.2021.623794] [Medline: 34142082]
4. Anderson RM, May RM. Population biology of infectious diseases: Part I. Nature 1979 Aug 02;280(5721):361-367. [doi: 10.1038/280361a0] [Medline: 460412]
5. Wong JEL, Leo YS, Tan CC. COVID-19 in Singapore-Current Experience: Critical Global Issues That Require Attention and Action. JAMA 2020 Apr 07;323(13):1243-1244. [doi: 10.1001/jama.2020.2467] [Medline: 32077901]
6. The Lancet Infectious Diseases. The COVID-19 infodemic. Lancet Infect Dis 2020 Aug;20(8):875 [FREE Full text] [doi: 10.1016/S1473-3099(20)30565-X] [Medline: 32687807]
7. 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]
8. Ahmad AR, Murad HR. The Impact of Social Media on Panic During the COVID-19 Pandemic in Iraqi Kurdistan: Online Questionnaire Study. J Med Internet Res 2020 May 19;22(5):e19556 [FREE Full text] [doi: 10.2196/19556] [Medline: 32369026]
9. Merchant RM, Lurie N. Social Media and Emergency Preparedness in Response to Novel Coronavirus. JAMA 2020 May 26;323(20):2011-2012. [doi: 10.1001/jama.2020.4469] [Medline: 32202611]
10. Operational planning guidance to support country preparedness and response. COVID-19 strategic preparedness and response. World Health Organization. 2020 May 22. URL: https://www.who.int/publications/i/item/draft-operational-planning-guidance-for-un-country-teams(Accessed [accessed 2020-07-11]
11. Edney S, Bogomolova S, Ryan J, Olds T, Sanders I, Maher C. Creating Engaging Health Promotion Campaigns on Social Media: Observations and Lessons From Fitbit and Garmin. J Med Internet Res 2018 Dec 10;20(12):e10911 [FREE Full text] [doi: 10.2196/10911] [Medline: 30530449]

12. Naslund JA, Kim SJ, Aschbrenner KA, McCulloch LJ, Brunette MF, Dallery J, et al. Systematic review of social media interventions for smoking cessation. Addict Behav 2017 Oct;73:81-93 [FREE Full text] [doi: 10.1016/j.addbeh.2017.05.002] [Medline: 28499259]
13. Abroms LC. Public Health in the Era of Social Media. Am J Public Health 2019 Feb;109(S2):S130-S131. [doi: 10.2105/AJPH.2018.304947] [Medline: 30785795]
14. Koohy H, Koohy B. A lesson from the ice bucket challenge: using social networks to publicize science. Front Genet 2014;5:430 [FREE Full text] [doi: 10.3389/fgene.2014.00430] [Medline: 25566317]
15. Lenoir P, Moulahi B, Azé J, Bringay S, Mercier G, Carbonnel F. Raising Awareness About Cervical Cancer Using Twitter: Content Analysis of the 2015 #SmearForSmear Campaign. J Med Internet Res 2017 Oct 16;19(10):e344 [FREE Full text] [doi: 10.2196/jmir.8421] [Medline: 29038096]
16. Plummer J, Rappaport S, Hall T, Barocci R. The Online Advertising Playbook: Proven Strategies and Tested Tactics from the Advertising Research Foundation. Hoboken, NJ: John Wiley & Sons; 2007.
17. Petrescu M, Korgaonkar P. Viral Advertising: Definitional Review and Synthesis. J Internet Commer 2011 Jul;10(3):208-226. [doi: 10.1080/15332861.2011.596007]
18. Pechmann C, Pan L, Delucchi K, Lakon CM, Prochaska JJ. Development of a Twitter-based intervention for smoking cessation that encourages high-quality social media interactions via automessages. J Med Internet Res 2015 Feb 23;17(2):e50 [FREE Full text] [doi: 10.2196/jmir.3772] [Medline: 25707037]
19. Fung IC, Tse ZTH, Fu K. The use of social media in public health surveillance. Western Pac Surveill Response J 2015;6(2):3-6 [FREE Full text] [doi: 10.5365/WPSAR.2015.6.1.019] [Medline: 26306208]
20. Gu H, Chen B, Zhu H, Jiang T, Wang X, Chen L, et al. Importance of Internet surveillance in public health emergency control and prevention: evidence from a digital epidemiologic study during avian influenza A H7N9 outbreaks. J Med Internet Res 2014 Jan 17;16(1):e20 [FREE Full text] [doi: 10.2196/jmir.2911] [Medline: 24440770]
21. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
22. Huang C, Chan E, Hyder AA. Web 2.0 and internet social networking: a new tool for disaster management?-lessons from Taiwan. BMC Med Inform Decis Mak 2010 Oct 06;10:57 [FREE Full text] [doi: 10.1186/1472-6947-10-57] [Medline: 20925944]
23. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PLoS One 2011 May 04;6(5):e19467 [FREE Full text] [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]
24. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc 2006:244-248 [FREE Full text] [Medline: 17238340]
25. Velardi P, Stilo G, Tozzi AE, Gesualdo F. Twitter mining for fine-grained syndromic surveillance. Artif Intell Med 2014 Jul;61(3):153-163. [doi: 10.1016/j.artmed.2014.01.002] [Medline: 24613716]
26. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection--harnessing the Web for public health surveillance. N Engl J Med 2009 May 21;360(21):2153-2157 [FREE Full text] [doi: 10.1056/NEJMp0900702] [Medline: 19423867]
27. 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]
28. Elkin LS, Topal K, Bebek G. Network based model of social media big data predicts contagious disease diffusion. Inf Discov Deliv 2017;45(3):110-120 [FREE Full text] [doi: 10.1108/IDD-05-2017-0046] [Medline: 31179401]
29. Sell TK, Boddie C, McGinty EE, Pollack K, Smith KC, Burke TA, et al. Media Messages and Perception of Risk for Ebola Virus Infection, United States. Emerg Infect Dis 2017 Jan;23(1):108-111 [FREE Full text] [doi: 10.3201/eid2301.160589] [Medline: 27983495]
30. Nekliudov NA, Blyuss O, Cheung KY, Petrou L, Genuneit J, Sushentsev N, et al. Excessive Media Consumption About COVID-19 is Associated With Increased State Anxiety: Outcomes of a Large Online Survey in Russia. J Med Internet Res 2020 Sep 11;22(9):e20955 [FREE Full text] [doi: 10.2196/20955] [Medline: 32788143]
31. 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]
32. Wathelet M, Duhem S, Vaiva G, Baubet T, Habran E, Veerapa E, et al. Factors Associated With Mental Health Disorders Among University Students in France Confined During the COVID-19 Pandemic. JAMA Netw Open 2020 Oct 01;3(10):e2025591 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.25591] [Medline: 33095252]
33. Pobiruchin M, Zowalla R, Wiesner M. Temporal and Location Variations, and Link Categories for the Dissemination of COVID-19-Related Information on Twitter During the SARS-CoV-2 Outbreak in Europe: Infoveillance Study. J Med Internet Res 2020 Aug 28;22(8):e19629 [FREE Full text] [doi: 10.2196/19629] [Medline: 32790641]
34. Zarocostas J. How to fight an infodemic. Lancet 2020 Feb 29;395(10225):676 [FREE Full text] [doi: 10.1016/S0140-6736(20)30461-X] [Medline: 32113495]

35. Berkovic D, Ackerman IN, Briggs AM, Ayton D. Tweets by People With Arthritis During the COVID-19 Pandemic: Content and Sentiment Analysis. J Med Internet Res 2020 Dec 03;22(12):e24550 [FREE Full text] [doi: 10.2196/24550] [Medline: 33170802]
36. Gunasekeran DV, Tseng RMWW, Tham Y, Wong TY. Applications of digital health for public health responses to COVID-19: a systematic scoping review of artificial intelligence, telehealth and related technologies. NPJ Digit Med 2021 Feb 26;4(1):40 [FREE Full text] [doi: 10.1038/s41746-021-00412-9] [Medline: 33637833]
37. Gunasekeran DV, Tham Y, Ting DSW, Tan GSW, Wong TY. Digital health during COVID-19: lessons from operationalising new models of care in ophthalmology. Lancet Digit Health 2021 Feb;3(2):e124-e134 [FREE Full text] [doi: 10.1016/S2589-7500(20)30287-9] [Medline: 33509383]
38. Liu M, Caputi TL, Dredze M, Kesselheim AS, Ayers JW. Internet Searches for Unproven COVID-19 Therapies in the United States. JAMA Intern Med 2020 Aug 01;180(8):1116-1118 [FREE Full text] [doi: 10.1001/jamainternmed.2020.1764] [Medline: 32347895]
39. The Editors Of The Lancet Group. Learning from a retraction. Lancet 2020 Oct 10;396(10257):1056 [FREE Full text] [doi: 10.1016/S0140-6736(20)31958-9] [Medline: 32950071]
40. Wang J, Zhou Y, Zhang W, Evans R, Zhu C. Concerns Expressed by Chinese Social Media Users During the COVID-19 Pandemic: Content Analysis of Sina Weibo Microblogging Data. J Med Internet Res 2020 Nov 26;22(11):e22152 [FREE Full text] [doi: 10.2196/22152] [Medline: 33151894]
41. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
42. Chandrasekaran R, Mehta V, Valkunde T, Moustakas E. Topics, Trends, and Sentiments of Tweets About the COVID-19 Pandemic: Temporal Infoveillance Study. J Med Internet Res 2020 Oct 23;22(10):e22624 [FREE Full text] [doi: 10.2196/22624] [Medline: 33006937]
43. Yousuf H, Corbin J, Sweep G, Hofstra M, Scherder E, van Gorp E, et al. Association of a Public Health Campaign About Coronavirus Disease 2019 Promoted by News Media and a Social Influencer With Self-reported Personal Hygiene and Physical Distancing in the Netherlands. JAMA Netw Open 2020 Jul 01;3(7):e2014323 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.14323] [Medline: 32639569]
44. Wahbeh A, Nasralah T, Al-Ramahi M, El-Gayar O. Mining Physicians' Opinions on Social Media to Obtain Insights Into COVID-19: Mixed Methods Analysis. JMIR Public Health Surveill 2020 Jun 18;6(2):e19276 [FREE Full text] [doi: 10.2196/19276] [Medline: 32421686]
45. Murri R, Segala FV, Del Vecchio P, Cingolani A, Taddei E, Micheli G, COVID II Columbus Group. Social media as a tool for scientific updating at the time of COVID pandemic: Results from a national survey in Italy. PLoS One 2020;15(9):e0238414 [FREE Full text] [doi: 10.1371/journal.pone.0238414] [Medline: 32881933]
46. Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data. J Med Internet Res 2020 May 06;22(5):e19458 [FREE Full text] [doi: 10.2196/19458] [Medline: 32352383]
47. Llewellyn S. Covid-19: how to be careful with trust and expertise on social media. BMJ 2020 Mar 25;368:m1160. [doi: 10.1136/bmj.m1160] [Medline: 32213480]
48. Reddy K, Shebl F, Foote J, Harling G, Scott JA, Panella C, et al. Cost-effectiveness of public health strategies for COVID-19 epidemic control in South Africa: a microsimulation modelling study. Lancet Glob Health 2021 Feb;9(2):e120-e129 [FREE Full text] [doi: 10.1016/S2214-109X(20)30452-6] [Medline: 33188729]
49. Younis J, Freitag H, Ruthberg JS, Romanes JP, Nielsen C, Mehta N. Social Media as an Early Proxy for Social Distancing Indicated by the COVID-19 Reproduction Number: Observational Study. JMIR Public Health Surveill 2020 Oct 20;6(4):e21340 [FREE Full text] [doi: 10.2196/21340] [Medline: 33001831]
50. Mackey TK, Li J, Purushothaman V, Nali M, Shah N, Bardier C, et al. Big Data, Natural Language Processing, and Deep Learning to Detect and Characterize Illicit COVID-19 Product Sales: Infoveillance Study on Twitter and Instagram. JMIR Public Health Surveill 2020 Aug 25;6(3):e20794 [FREE Full text] [doi: 10.2196/20794] [Medline: 32750006]
51. Qin L, Sun Q, Wang Y, Wu K, Chen M, Shia B, et al. Prediction of Number of Cases of 2019 Novel Coronavirus (COVID-19) Using Social Media Search Index. Int J Environ Res Public Health 2020 Mar 31;17(7):2365 [FREE Full text] [doi: 10.3390/ijerph17072365] [Medline: 32244425]
52. Shen C, Chen A, Luo C, Zhang J, Feng B, Liao W. Using Reports of Symptoms and Diagnoses on Social Media to Predict COVID-19 Case Counts in Mainland China: Observational Infoveillance Study. J Med Internet Res 2020 May 28;22(5):e19421 [FREE Full text] [doi: 10.2196/19421] [Medline: 32452804]
53. Communication for Behavioural Impact (COMBI): A toolkit for behavioural and social communication in outbreak response. Zika Communication Network. URL: https://tinyurl.com/kpryvxst [accessed 2022-03-09]
54. Adewuyi E, Adefemi K. Behavior change communication using social media: A review. Int J Commun Health 2016(9):109-116 [FREE Full text]
55. Lindson-Hawley N, Thompson TP, Begh R. Motivational interviewing for smoking cessation. Cochrane Database Syst Rev 2015 Mar 02(3):CD006936. [doi: 10.1002/14651858.CD006936.pub3] [Medline: 25726920]

56. Gunasekeran DV. Regulations for the development of deep technology applications in healthcare urgently needed to prevent abuse of vulnerable patients. BMJ Innov 2018 Jan 27;4(2):111-112. [doi: 10.1136/bmjinnov-2017-000242]
57. Chew AMK, Ong R, Lei H, Rajendram M, Verma SK, Fung DSS, et al. Digital Health Solutions for Mental Health Disorders During COVID-19. Front Psychiatry 2020;11:582007 [FREE Full text] [doi: 10.3389/fpsyt.2020.582007] [Medline: 33033487]
58. Lim YW, Chia A. Social Entrepreneurship: Improving Global Health. JAMA 2016 Jun 14;315(22):2393-2394. [doi: 10.1001/jama.2016.4400] [Medline: 27299615]
59. Idriss SZ, Kvedar JC, Watson AJ. The role of online support communities: benefits of expanded social networks to patients with psoriasis. Arch Dermatol 2009 Jan;145(1):46-51. [doi: 10.1001/archdermatol.2008.529] [Medline: 19153342]
60. Jelodar H, Wang Y, Orji R, Huang S. Deep Sentiment Classification and Topic Discovery on Novel Coronavirus or COVID-19 Online Discussions: NLP Using LSTM Recurrent Neural Network Approach. IEEE J Biomed Health Inform 2020 Oct;24(10):2733-2742. [doi: 10.1109/JBHI.2020.3001216] [Medline: 32750931]
61. Huh J, Marmor R, Jiang X. Lessons Learned for Online Health Community Moderator Roles: A Mixed-Methods Study of Moderators Resigning From WebMD Communities. J Med Internet Res 2016 Sep 08;18(9):e247 [FREE Full text] [doi: 10.2196/jmir.6331] [Medline: 27608721]
62. PCN GP Annual Grand Conference 2020. 2020. URL: https://www.askdr.co/pcn-conference-2020?week=2#schedules [accessed 2020-11-28]
63. Ostrovsky AM, Chen JR. TikTok and Its Role in COVID-19 Information Propagation. J Adolesc Health 2020 Nov;67(5):730 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.07.039] [Medline: 32873499]
64. Clemente L, Lu F, Santillana M. Improved Real-Time Influenza Surveillance: Using Internet Search Data in Eight Latin American Countries. JMIR Public Health Surveill 2019 Apr 04;5(2):e12214 [FREE Full text] [doi: 10.2196/12214] [Medline: 30946017]
65. Wong MYZ, Gunasekeran DV, Nusinovici S, Sabanayagam C, Yeo KK, Cheng C, et al. Telehealth Demand Trends During the COVID-19 Pandemic in the Top 50 Most Affected Countries: Infodemiological Evaluation. JMIR Public Health Surveill 2021 Feb 19;7(2):e24445 [FREE Full text] [doi: 10.2196/24445] [Medline: 33605883]
66. Liu D, Clemente L, Poirier C, Ding X, Chinazzi M, Davis J, et al. Real-Time Forecasting of the COVID-19 Outbreak in Chinese Provinces: Machine Learning Approach Using Novel Digital Data and Estimates From Mechanistic Models. J Med Internet Res 2020 Aug 17;22(8):e20285 [FREE Full text] [doi: 10.2196/20285] [Medline: 32730217]
67. Gozzi N, Tizzani M, Starnini M, Ciulla F, Paolotti D, Panisson A, et al. Collective Response to Media Coverage of the COVID-19 Pandemic on Reddit and Wikipedia: Mixed-Methods Analysis. J Med Internet Res 2020 Oct 12;22(10):e21597 [FREE Full text] [doi: 10.2196/21597] [Medline: 32960775]
68. Agrawal R, Gunasekeran DV, Grant R, Agarwal A, Kon OM, Nguyen QD, Collaborative Ocular Tuberculosis Study (COTS)–1 Study Group. Clinical Features and Outcomes of Patients With Tubercular Uveitis Treated With Antitubercular Therapy in the Collaborative Ocular Tuberculosis Study (COTS)-1. JAMA Ophthalmol 2017 Dec 01;135(12):1318-1327 [FREE Full text] [doi: 10.1001/jamaophthalmol.2017.4485] [Medline: 29075752]
69. Gunasekeran DV, Agrawal R, Testi I, Agarwal A, Mahajan S, Nguyen QD, Collaborative Ocular Tuberculosis Study (COTS) Group. Lessons in Digital Epidemiology from COTS-1: Coordinating Multicentre Research across 10 Countries Using Operational and Technology Innovation to Overcome Funding Deficiencies. Ocul Immunol Inflamm 2020 Apr 22:1-7. [doi: 10.1080/09273948.2020.1744669] [Medline: 32320326]
70. Yang S, Santillana M, Brownstein JS, Gray J, Richardson S, Kou SC. Using electronic health records and Internet search information for accurate influenza forecasting. BMC Infect Dis 2017 May 08;17(1):332 [FREE Full text] [doi: 10.1186/s12879-017-2424-7] [Medline: 28482810]
71. Sustained suppression. Nat Biomed Eng 2020 May 13;4(5):479-480 [FREE Full text] [doi: 10.1038/s41551-020-0567-0] [Medline: 32405031]
72. Li LW, Chew AM, Gunasekeran DV. Digital health for patients with chronic pain during the COVID-19 pandemic. Br J Anaesth 2020 Nov;125(5):657-660 [FREE Full text] [doi: 10.1016/j.bja.2020.08.003] [Medline: 32863018]
73. Sesagiri Raamkumar A, Tan SG, Wee HL. Use of Health Belief Model-Based Deep Learning Classifiers for COVID-19 Social Media Content to Examine Public Perceptions of Physical Distancing: Model Development and Case Study. JMIR Public Health Surveill 2020 Jul 14;6(3):e20493 [FREE Full text] [doi: 10.2196/20493] [Medline: 32540840]
74. Ting DSW, Carin L, Dzau V, Wong TY. Digital technology and COVID-19. Nat Med 2020 Apr;26(4):459-461 [FREE Full text] [doi: 10.1038/s41591-020-0824-5] [Medline: 32284618]
75. 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]
76. Ghafur S, Grass E, Jennings NR, Darzi A. The challenges of cybersecurity in health care: the UK National Health Service as a case study. Lancet Digit Health 2019 May;1(1):e10-e12 [FREE Full text] [doi: 10.1016/S2589-7500(19)30005-6] [Medline: 33323235]
77. Swartz L. New Money: How Payment Became Social Media. London: Yale University Press; 2020.
78. Kakarlapudi PV, Mahmoud QH. A Systematic Review of Blockchain for Consent Management. Healthcare (Basel) 2021 Feb 01;9(2):137 [FREE Full text] [doi: 10.3390/healthcare9020137] [Medline: 33535465]

79. Ocean Protocol: Tools for the Web3 Data Economy. Ocean Protocol Foundation Ltd. URL: https://oceanprotocol.com/tech-whitepaper.pdf [accessed 2020-03-09]
80. Nawaz A, Peña Queralta J, Guan J, Awais M, Gia TN, Bashir AK, et al. Edge Computing to Secure IoT Data Ownership and Trade with the Ethereum Blockchain. Sensors (Basel) 2020 Jul 16;20(14):3965 [FREE Full text] [doi: 10.3390/s20143965] [Medline: 32708807]
81. Gürsoy G, Brannon CM, Gerstein M. Using Ethereum blockchain to store and query pharmacogenomics data via smart contracts. BMC Med Genomics 2020 Jun 01;13(1):74 [FREE Full text] [doi: 10.1186/s12920-020-00732-x] [Medline: 32487214]
82. Pennycook G, McPhetres J, Zhang Y, Lu JG, Rand DG. Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. Psychol Sci 2020 Jul 30;31(7):770-780 [FREE Full text] [doi: 10.1177/0956797620939054] [Medline: 32603243]
83. Cutler DM, Summers LH. The COVID-19 Pandemic and the $16 Trillion Virus. JAMA 2020 Oct 20;324(15):1495-1496 [FREE Full text] [doi: 10.1001/jama.2020.19759] [Medline: 33044484]
84. Wilson SL, Wiysonge C. Social media and vaccine hesitancy. BMJ Glob Health 2020 Oct;5(10):e004206 [FREE Full text] [doi: 10.1136/bmjgh-2020-004206] [Medline: 33097547]

Abbreviations

AI: artificial intelligence
API: application programming interface
CDC: US Centers for Disease Control and Prevention
EMR: electronic medical record
LKCMedicine: Lee Kong Chian School of Medicine
ML: machine learning
NUS: National University of Singapore
OHC: online health community
SIER: Susceptible-Exposed-Infectious Removed
SMSI: social media search indexes
UGC: user-generated content

Edited by M Gisondi; submitted 19.09.21; peer-reviewed by Y Liu, S Yang; comments to author 08.10.21; revised version received 27.01.22; accepted 04.02.22; published 11.04.22.

Please cite as:
Gunasekeran DV, Chew A, Chandrasekar EK, Rajendram P, Kandarpa V, Rajendram M, Chia A, Smith H, Leong CK
The Impact and Applications of Social Media Platforms for Public Health Responses Before and During the COVID-19 Pandemic: Systematic Literature Review
J Med Internet Res 2022;24(4):e33680
URL: https://www.jmir.org/2022/4/e33680
doi:10.2196/33680
PMID:35129456

Original Paper

The Prevalence and Impact of Fake News on COVID-19 Vaccination in Taiwan: Retrospective Study of Digital Media

Yen-Pin Chen[1,2], MD; Yi-Ying Chen[2], MD; Kai-Chou Yang[3], MSc; Feipei Lai[1,4], PhD; Chien-Hua Huang[2*], MD, PhD; Yun-Nung Chen[4*], PhD; Yi-Chin Tu[3], MSc

[1]Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei, Taiwan

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

[3]Taiwan AI Labs, Taipei, Taiwan

[4]Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan

[*]these authors contributed equally

Corresponding Author:
Yun-Nung Chen, PhD
Department of Computer Science and Information Engineering
National Taiwan University
No 1, Sec 4, Roosevelt Rd
Taipei, 106
Taiwan
Phone: 886 2 3366 3366
Email: yvchen@csie.ntu.edu.tw

Abstract

Background: Vaccination is an important intervention to prevent the incidence and spread of serious diseases. Many factors including information obtained from the internet influence individuals' decisions to vaccinate. Misinformation is a critical issue and can be hard to detect, although it can change people's minds, opinions, and decisions. The impact of misinformation on public health and vaccination hesitancy is well documented, but little research has been conducted on the relationship between the size of the population reached by misinformation and the vaccination decisions made by that population. A number of fact-checking services are available on the web, including the Islander news analysis system, a free web service that provides individuals with real-time judgment on web news. In this study, we used such services to estimate the amount of fake news available and used Google Trends levels to model the spread of fake news. We quantified this relationship using official public data on COVID-19 vaccination in Taiwan.

Objective: In this study, we aimed to quantify the impact of the magnitude of the propagation of fake news on vaccination decisions.

Methods: We collected public data about COVID-19 infections and vaccination from Taiwan's official website and estimated the popularity of searches using Google Trends. We indirectly collected news from 26 digital media sources, using the news database of the Islander system. This system crawls the internet in real time, analyzes the news, and stores it. The incitement and suspicion scores of the Islander system were used to objectively judge news, and a fake news percentage variable was produced. We used multivariable linear regression, chi-square tests, and the Johnson-Neyman procedure to analyze this relationship, using weekly data.

Results: A total of 791,183 news items were obtained over 43 weeks in 2021. There was a significant increase in the proportion of fake news in 11 of the 26 media sources during the public vaccination stage. The regression model revealed a positive adjusted coefficient (β=0.98, *P*=.002) of vaccine availability on the following week's vaccination doses, and a negative adjusted coefficient (β=–3.21, *P*=.04) of the interaction term on the fake news percentage with the Google Trends level. The Johnson-Neiman plot of the adjusted effect for the interaction term showed that the Google Trends level had a significant negative adjustment effect on vaccination doses for the following week when the proportion of fake news exceeded 39.3%.

Conclusions: There was a significant relationship between the amount of fake news to which the population was exposed and the number of vaccination doses administered. Reducing the amount of fake news and increasing public immunity to misinformation will be critical to maintain public health in the internet age.

(J Med Internet Res 2022;24(4):e36830) doi:10.2196/36830

KEYWORDS

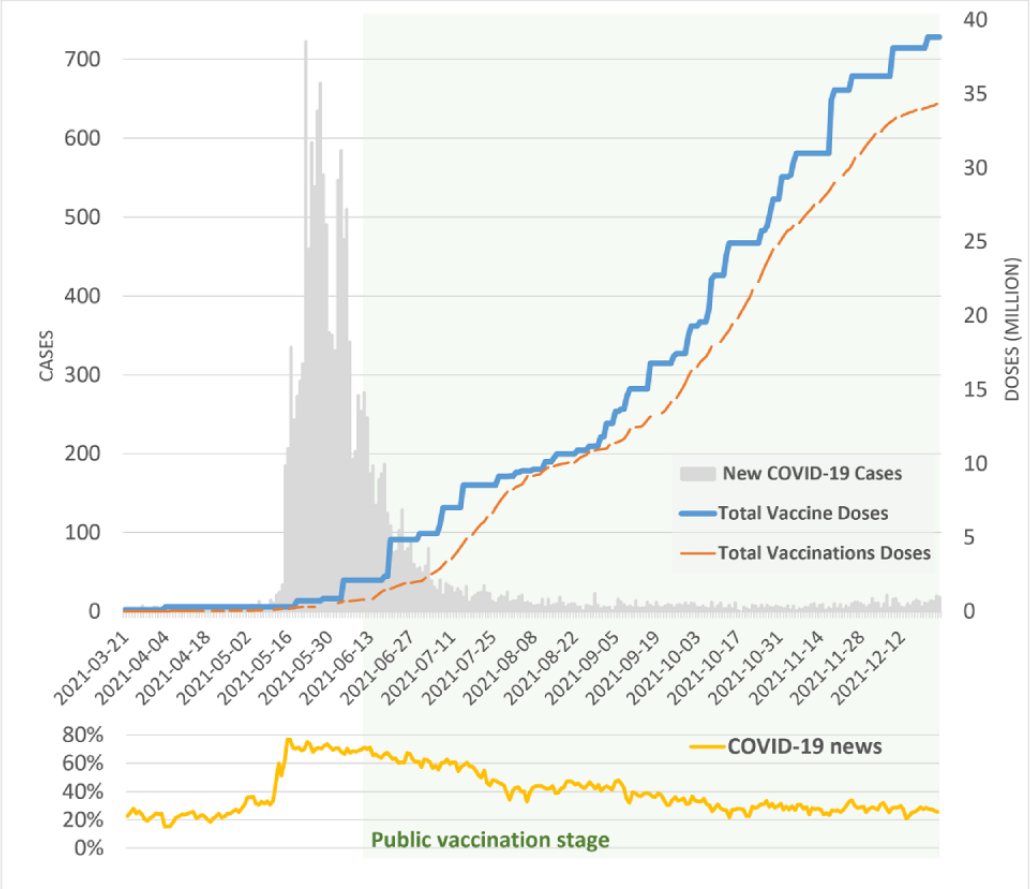
misinformation; vaccine hesitancy; vaccination; infodemic; infodemiology; COVID-19; public immunity; social media; fake news

Introduction

To take the blue pill or the red pill: decisions are made every day in our lives. As expressed in the 1999 film The Matrix, “You take the blue pill—the story ends, you wake up in your bed and believe whatever you want to believe. You take the red pill—you stay in Wonderland, and I show you how deep the rabbit hole goes” [1]. Every decision may have critical or trivial effects on our future and may be influenced by our environment. Decisions about whether to accept or reject vaccination can be influenced by a variety of factors [2-6] including personal lifestyle, disease severity, vaccine effectiveness, side effects, peer decisions, and internet information. The internet has brought everyone together over the last decades, and misinformation on the internet can spread like a plague and affect public positions [7-13], even encouraging individuals to make potentially self-harming health decisions [14,15].

The COVID-19 pandemic spread around the world from about mid-2020, and vaccines were authorized for emergency use in early 2021 [16]. Taiwan, located in East Asia, with a population of 23 million (population density of 646 people/km²), received its first batch of COVID-19 vaccines on March 3, 2021, and started vaccination on March 22, 2021 [17]. Given the initially limited number of vaccine doses available, and the policy to vaccinate health care workers first, public vaccination started on June 12, 2021 [17]. During the vaccination period, Taiwan experienced its first wave of large-scale community infections, and the internet was flooded with news about COVID-19 and vaccines (Figure 1). Considerable research has indicated that misinformation about diseases and potential vaccine side effects have adverse effects on vaccination rates [15,18,19]. Some researchers have designed questionnaire-based studies to investigate this association [20-22]. One such study quantified the rise in the number of antivaccine tweets during the pandemic [23], and several studies investigated factors affecting the spread of misinformation [24,25]. Building upon this previous research, we hypothesized that a higher prevalence of misinformation might have a greater adverse effect on vaccination decisions.

Figure 1. Data about COVID-19 infection cases, total vaccine doses, vaccine uptake (vaccination doses), and the percentage of COVID-19 news in Taiwan. The data covers a period ranging from March 2021 to December 2021, and the orange dotted line represents vaccinations in Taiwan, with missing values on weekends and holidays. The public vaccination stage began on June 12, 2021, as indicated by the green background.



Detecting misinformation or fake news from big data on the internet is challenging [13]. In this decade, deep learning for natural language processing (NLP) has been developed to help address this problem, and many news analysis services are already available on the web [26]. These services use machine learning algorithms or manual detection methods to provide online fact-checking covering multiple topics [26,27]. However, these services were difficult to use in this study due to language differences. In this study, we focused on digital media news in Taiwan and used the Islander news analysis system [28], which uses an innovative language model to automatically screen and score internet news.

There is no consistent definition of fake news; its identification is complex and can sometimes be difficult to determine [12,27,29-31]. The definition of fake news can be as broad as improper information or stories [18,27,32], or as narrow as verifiably false articles deliberately published by the media [11,12,27], and anything in between [13,33]. Experts or the wisdom of crowds can detect false information manually [27,34], but efficiency can be an issue when news may have spread before a judgment was made. An automatic detection method could involve knowledge base retrieval systems [27], but breakthrough knowledge may be considered misinformation. Content style analysis is another automated method, based on the assumption that there is a certain pattern in intentional news [31,35-37], but outlets may evade detection by manipulating their writing style [27]. In this study, we employed a style-based approach to fake news detection. Generally speaking, the typical characteristics of fake news are associated with the writing style, quantity of subjective language, and sentiment lexical or incited discourse [26,27,31,35-37]. We adopted the scores of suspicion and incitement provided by the Islander news analysis system [28] in which a language model, RoBERTa [38], was trained using a supervised learning approach to analyze and score news (Figure 2). This news analysis language model was trained on the Chinese valence-arousal text data set (CVAT) [39], and 198 random news items from mid-2019, labeled by 2 journalism experts. These 2 experts labeled the bias of the title and objective statements or subjective claims, and crossvalidated them. CVAT includes 720 texts tagged with affective words, and each sentence was scored according to valence and arousal, which were used to train the incitement judgment of the Islander system. This quantifiable domain knowledge, combined with the writing style and incited score, constitutes the Islander system's fake news discriminator.

Individuals obtain internet information by passively accepting pushes from web services or by actively searching for specific terms. Searches reflect user interests [40,41], and many web news services have adopted a recommendation system to push information to potentially interested people using data gathered from personal surfing behavior or search histories [13,42-44]. Some studies have indicated that search trends can reflect the amount of information dissemination [45,46]. We used Google Trends as a metric for the amount of information propagated by web news, due to its up to 85% market share [6,47].

Few studies have investigated the interplay among the quantity of misinformation, information propagation, and its impact on decision-making [13]. In this study, we retrospectively analyzed the relationship between vaccination acceptance and digital news dissemination in Taiwan and aimed to quantify the effect of the propagation of fake news on COVID-19 vaccination decisions (Figure 3).

Figure 2. The Islander news analysis system. This system has 3 components: a web crawler to collect web news in real time, a news analysis model to judge the news objectively, and a website that provides a user interface.

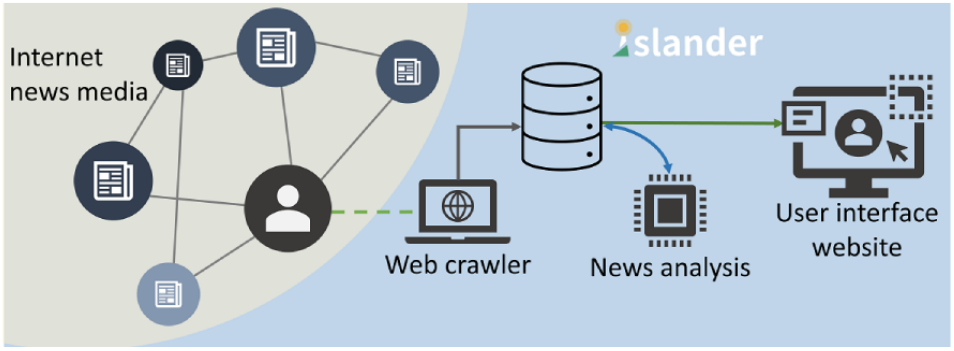
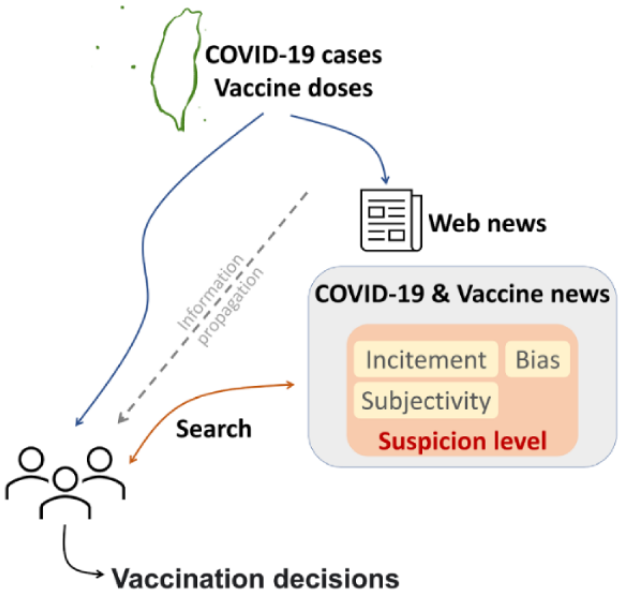


Figure 3. Graphical summary of this study. Taiwanese officials publicly release COVID-19 and vaccination information, and the media post news about this information on the internet. The public may obtain relevant information using searches or pushes from a recommendation service. This information will help individuals make vaccination decisions. In this study, we investigated the relationship between the quality of news, its dissemination, and vaccination decisions.



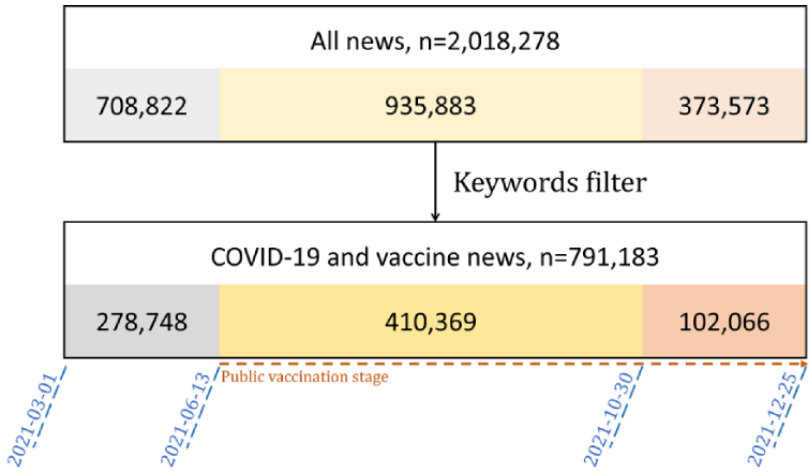
Methods

Study Design and Setting

The study population was the population of Taiwan. We conducted a retrospective study using publicly available data from March 1, 2021, to December 25, 2021, starting from when Taiwan first obtained the vaccine. The government publicly releases information about COVID-19, vaccines, and vaccination numbers, and we collected information on the COVID-19 pandemic from the Taiwan Ministry of Health and Welfare [17] and the Our World In Data [48] website. A total of 5 variables were used, including the number of COVID-19 infection cases, the number of COVID-19 deaths, total vaccine doses available, total vaccinations, and the number of vaccinated individuals. The web news we collected came from the Islander system news database in which news is crawled and stored in real time. Each news item included the title, content, source, publishing time, suspicion score, and incited score. We obtained data on daily trends through a Google Trends news subgroup search for "疫苗" (vaccine) in Taiwan within the date range.

To investigate the relationship between internet news and vaccination acceptance by the public, we set the analysis interval from June 13, 2021, to December 25, 2021, according to the timing of public vaccination. We divided the time interval into training and validation parts, with a ratio of 70 to 30. Data from before October 30, 2021, were analyzed separately, and the other data were used for validation (Figure 4).

Figure 4. The news collected in this study. A total of 2,018,278 items were included and filtered by keywords for COVID-19 and vaccine news, leaving 791,183 news items for research. A study interval of June 13, 2021, to December 25, 2021, was used to investigate decisions by the public about vaccination. We used data from October 31, 2021, to December 25, 2021, for validation.



Variables and Outcome

We resampled daily to weekly data and obtained the following information: the number of available vaccine doses, calculated as the difference between the number of vaccine doses available and the number of vaccinations; the number of new COVID-19 cases per week; the number of new COVID-19 deaths per week; the number of new vaccinations administered per week; the number of newly vaccinated people per week; and the average Google Trends score each week. Individuals will be interested in the issue and search for it, and relevant information will be provided; thus, we selected COVID-19 and vaccine keywords to filter the news data set. We filtered news related to COVID-19 and vaccination using the following keywords limited to Chinese news: “破口,” “病例,” “polymerase chain reaction (PCR),” “放寬,” “疫,” “隔離,” “確診,” “COVID,” “新冠,” “新型冠狀病毒,” “肺炎,” “疾管,” “疫苗,” “BioNTech (BNT),” “AstraZeneca (AZ),” “高端,” “默德納,” “Moderna,” “vaccine,” “接種,” “vaccinate,” “vaccination.” Multimedia Appendix 1 presents the meaning and English translation of the Chinese search keywords. Subgroups of digital news with different subsets of keywords were also employed in the study to investigate their relationship with vaccination doses. We counted the weekly number of news and the percentage of fake news. In this study, fake news was set as news with a suspicion score greater than zero. Suspicion scores ranged from 0 to 1000; lower scores indicate greater objectivity, and zero scoring was predominant in the data, which looked like a Poisson distribution. We also selected the weekly average incitement score as a variable. Incitement scores ranged from 0 to 1000 and presented as a Gaussian distribution; lower scores indicate less incitement (Multimedia Appendix 2).

The outcomes of this study were the number of new vaccination doses and newly vaccinated people for the following week. We investigated the factors affecting vaccination decisions using the following variables available: vaccine doses, new COVID-19 cases, average Google Trends score, fake news percentage, average incitement score, and the interaction term of the average Google Trends score with the fake news percentage.

Statistical Analysis

We used chi-square tests for the analysis of fake news percentages, and multivariable linear regression with the stepwise method was used for variable selection. The variance inflation factor was used to detect multicollinearity among variables and to remove probable linear combinations of variables. The Johnson-Neyman procedure was used to generate plots of the interaction effects with 95% CIs. The final models were validated using the validation data.

Data were normalized and then analyzed using the R (version 4.1.1; R Core Team), statistical packages interactions (version 1.1.5), R commander (version 2.7-1), and RStudio (version 1.3.1093). All *P* values in this study were 2-sided and were considered statistically significant when less than .05.

Results

Using the settings described, 791,183 COVID-19 and vaccine news items were collected from 26 internet news media sources. A higher percentage of fake news (193,188/512,435, 37.7%; 95% CI 37.6%-37.8%) was found during the public vaccination stage, than during the nonpublic vaccination stage (99,791/278,748, 35.8%; 95% CI 35.6%-36.0%); and 11 of the 26 news media sources had significantly increased fake news percentages during the public vaccination stage (Figure 5). This study involved 28 weeks of data for the regression analysis (details on variables and outcomes are shown in Table 1). Every week, about 3 million vaccine doses were available in Taiwan, and about 1 million doses were administered to the public.

Figure 5. Fake news percentages, with 95% CI, of each media. Multimedia Appendix 1 provides the sources of digital media.

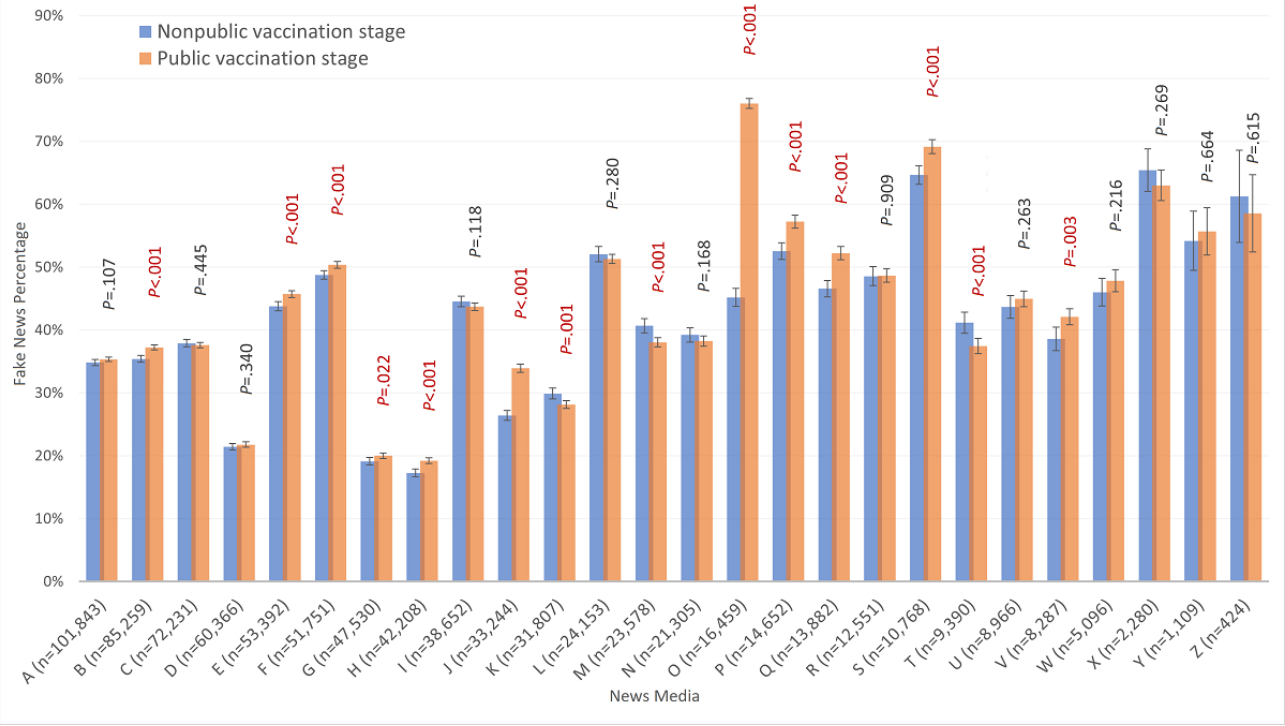


Table 1. Summary statistics of the variables used in the study.

	Mean (SD)	Minimum	Median	Maximum
Variables				
Available vaccine doses	3,129,315.5 (1,684,054.2)	351,662	3,291,468	6,263,838
New COVID-19 cases	148 (238)	28	65.5	1150
New COVID-19 death cases	15.7 (30.4)	0	2	127
Incitement score	488.7 (2.4)	483.9	488.5	492.4
Fake news (%)	37.4 (1.9)	33.7	37.6	41.3
Google Trends	22.4 (14.9)	4.3	19	54
Outcomes				
Following week's vaccination doses	1,194,379.4 (632,178.6)	308,400	1,090,186.5	2,764,054
Following week's newly vaccinated people	633,134.8 (490,868.2)	52,519	460,499.5	1,590,232

Multivariate analysis revealed a statistically significant relationship between the number of vaccine doses administered and the number of available vaccine doses, as well as an interaction term for the percentage of fake news and Google Trends levels. These significances persisted even when analyzed together with the validation data (Table 2). These coefficients suggested that there may be a positive relationship between the number of vaccine doses available and the number of vaccine doses administered during the following week, and that the incitement score might adversely affect vaccination doses in the following week. There also appeared to be an interaction between the fake news percentage and the Google Trends level, due to the opposite sign of the interaction term.

The interaction effects for fake news percentage and Google Trends levels in the multiple regression revealed that as the fake news percentage increased, the slope of the Google Trends level moved from positive to negative (Figure 6). The Johnson-Neyman procedure suggested that when the fake news percentage exceeded 39.3%, the Google Trends level had a significantly negative adjusted effect on the following week's vaccination doses (Figure 7).

Table 2. A multivariable linear regression model of factors associated with vaccination doses for the following week. The variance inflation factor (VIF) for each factor was less than 10.

	June 13 to October 30, 2021[a]				June 13 to December 25, 2021[b]			
	Estimate	SE	*P* value	VIF	Estimate	SE	*P* value	VIF
Coefficients								
Intercept	–0.1482	0.4805	.76	—[c]	–0.0450	0.3721	.90	—
Available vaccine doses	0.9799	0.2637	.002[d]	1.96	0.4510	0.1774	.02[d]	1.43
Incitement score	–0.4725	0.2953	.13	3.31	–0.5222	0.2279	.03[d]	2.40
Fake news (%)	3.8286	1.9884	.07	4.72	1.6420	1.1771	.18	2.53
Google Trends	0.8257	0.5208	.14	8.14	1.0382	0.3970	.02[d]	6.64
Fake news: Google Trends	–3.2121	1.3796	.04[d]	9.95	–2.5846	0.9058	.009[d]	5.23

[a]Multiple R^2=0.647, adjusted R^2=0.521, $F_{5,14}$=5.133; P=.007.

[b]Multiple R^2=0.507, adjusted R^2=0.395, $F_{5,22}$=7.714; P<.001.

[c]Not applicable.

[d]Indicates significant values.

Figure 6. Interaction plot with 95% confidence bands. This plot demonstrates the interaction of the following week's vaccination doses with the Google Trends levels for those with 1 SD above and below the average for the fake news percentage.

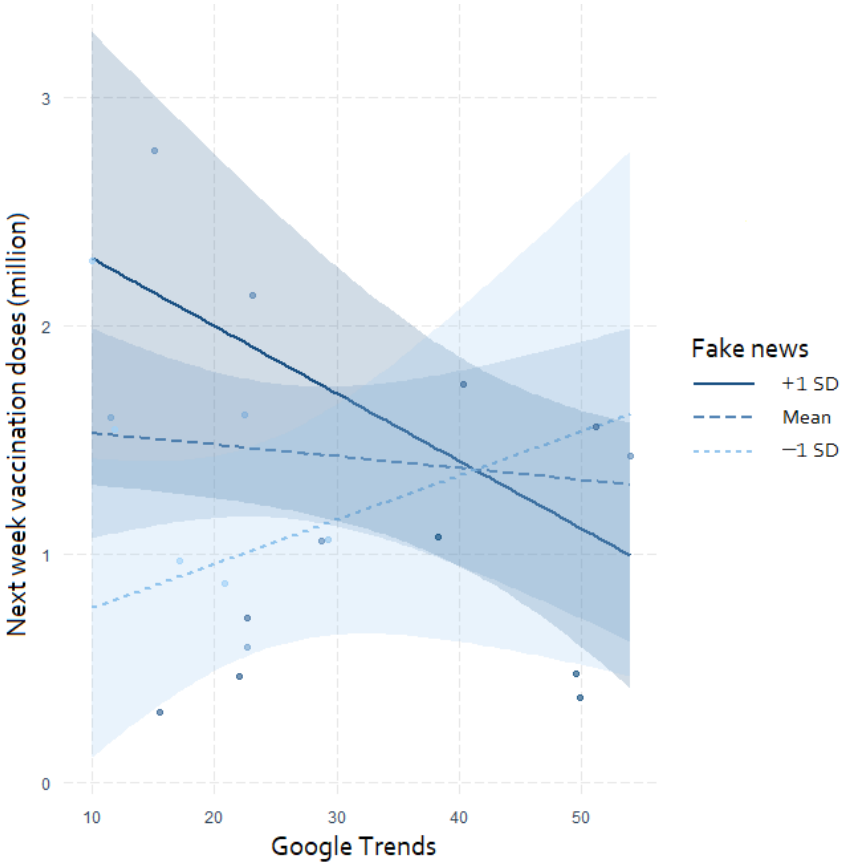
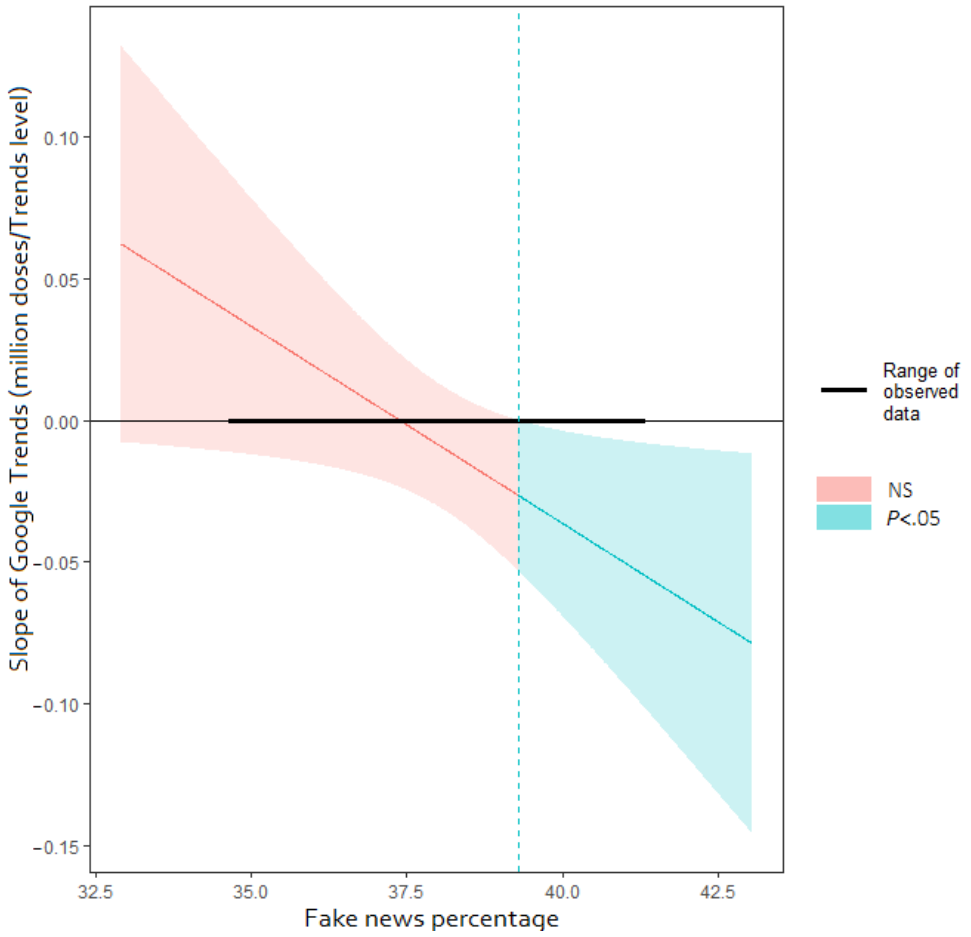


Figure 7. Johnson-Neyman plot with 95% confidence bands. This plot shows the Google Trends level coefficient adjusted for different percentages of fake news. NS: not significant.



Discussion

Principal Findings

In this study, we quantified the relationship between the proportion of fake news, its propagation, and vaccination decisions in Taiwan, using multivariable linear regression and interaction analysis. A higher percentage of fake news about COVID-19 and vaccines on the internet and greater search volumes predicted more adverse effects on vaccination doses administered in the following week. During the study interval, the fake news percentage threshold was 37.4%, which was the zero-crossing coefficient of the Google Trends level and was statistically significant when it reached 39.3%. This number may vary with study intervals, but this trend existed even in the unseen validation data. The exposure of populations to more than a specific amount of fake news about diseases and vaccines can negatively impact public health. Public health work on vaccination should strengthen public immunity to fake news and encourage balance and objectivity among news media outlets.

The overall percentage of fake news rose by 2 points during the public vaccination stage. One reason for this increase might be the official announcement of the community spread of COVID-19 in Taiwan on May 15, 2021, although there was no specifically significant increase in the fake news percentage for the following 2 weeks (26,447/73,669, 35.9%; 95% CI 35.6%-36.3%). The percentage increased significantly during the first 10 days of June 2021 (19,969/52,276, 38.2%; 95% CI 37.8%-38.6%). At the same time, Taiwan was facing its second peak of infection and received Japan's donation of the first batch of vaccines. The number of infections then ebbed, but some media outlets seemed to still overreact during the public vaccination stage (Multimedia Appendix 3). News media have different news styles based on their culture, which might relate to varying levels of suspicion and incitement. Figure 5 shows the different fake news percentages for each form of media, some of which maintained a consistent style in both stages, but some of which increased significantly in the second stage. The greatest increase was 1.7 times and the second largest was a 34% increase. Lazer et al [13] indicated that the internet accelerated the news media's move toward biased and affective reporting. Internet news outlets are commercial, and click-through rates reflect revenue and sometimes share prices. Using attractive discourse and sentimental titles will be the preference of some media companies, and sometimes the content is subjective and lacks fact-checking. It may be reasonable to change styles in the pursuit of click-through rates, but this approach might undermine the credibility of the media and public trust.

The number of vaccine doses available had a positive adjusted effect on the number of vaccine doses administered in the following week. For most people seeking vaccinations in Taiwan, it is necessary to reserve a vaccination day and then

visit. As when booking a flight, the number of seats on an aircraft determines the number of bookings available. Although no-shows happen, overselling is prohibited when it comes to vaccination, as limited resources could lead to a "bank run" phenomenon on vaccination, especially if the masses panic. In August 2021, fewer than 400,000 vaccine doses were available every week, and the rate of vaccination was slow without the "vaccine run" effect (Figure 1). In that month, the percentage of fake news increased 1 point to 38.5% (95% CI 38.2%-38.8%), exceeding the threshold, but not reaching a significant level, which may be a decelerating factor.

In the regression model analysis, we factored out infection and death cases, because COVID-19 was gradually brought under control over the interval analyzed, and the number of deaths was correlated with the number of infections. We found multicollinearity between the number of infections and the percentage of fake news, the Google Trends level, and their interaction term. The values for infection cases could be almost linear combinations of these factors, potentially undermining the reliability of the model. The coefficients for these factors were 1.0, –0.7, and 2.9 respectively (R^2=0.896; P<.001).

In this study, we used the Google Trends level to represent the magnitude of the spread of COVID-19 and vaccine news. We believe this approach is justifiable because Google's dominance of the market share of searches makes them a good proxy for the overall data. The trend for declining search levels within the study interval may be related to the ebb of COVID-19 infections and the public's attention shifting to other issues. These tendencies might reflect a link between information dissemination and the Google Trends level. It is a caveat to note that the Google Trends tool does not provide consistent results; specifically, the Google Trends level varies based on the selected time interval and is relative over time rather than being a fixed score. In the regression analysis, normalization was used to counteract this variation in the data. The effects of the subgroups of COVID-19 and vaccine news on vaccination were also analyzed, but only the entire set had statistically significant results. When people search for information about vaccines, relevant information will be available to the public through associated links, search engines, or recommender systems. Everyone is faced with an overwhelming amount of information on the internet, few people read every news item, and sometimes people skim them. Also, attention may shift to another related topic rather than the original one during a search [49]. These sources of noise might lead to a lack of statistical significance when using news subgroups with only COVID-19 or vaccines.

The interaction between the fake news percentage and the Google Trends level is an important factor in this regression analysis, without which no statistical significance can be observed for the individual variables. This observation may suggest that no matter what the media has to offer, it cannot influence public opinion without human contact. However, this lack of access is not possible unless the internet collapses. In this study, we found that there is a threshold above which the fake news percentage had a negative impact, which might be regarded as the point at which the resistance of the public to misinformation was overcome. As more media outlets adopt attractive journalism styles and more inciting discourse, it may be practical to strengthen our resistance rather than restrict the freedom of expression of the media, but the media should reflect and consider returning to the essence of journalism.

Comparison With Prior Works

Lazer et al [13] points out that little is known about the prevalence of misinformation or the scale of its spread and impact. To the best of our knowledge, studies to date have not explicitly addressed these gaps. Loomba et al [22] designed a prospective study to examine vaccine intent before and after exposure to misinformation and confirmed that misinformation has adverse effects on vaccination rates. Questionnaire studies have demonstrated the impact of misinformation on vaccine hesitancy [20-22], but this approach does not quantify how much misinformation is needed to change the public's perspective. King and Wang [24] retrospectively collected 42 million tweets and found that messages containing misinformation or emotional content spread quickly. Infodemic research involving social media data is common [23,24], and information about user interactions can be used to analyze the dissemination of information. The amount of misinformation can be estimated from public postings, but this approach may lead to an underestimation of the extent of the misinformation because the data do not include information from private communities or groups on social media.

This study used big news data, and the target population was the population of Taiwan. The results were consistent with those from previous studies [20-22], which found that misinformation can lower vaccine intent. We further quantified the effect of varying amounts of fake news on the public vaccination rate. By accessing almost every news outlet in Taiwan, we estimated the prevalence of fake news using an automatic style-based detection method. Although we adopted a broad definition of fake news, the results of this study provided an estimate of the extent of fake news in Taiwan. However, the best way to directly estimate the spread of misinformation remains a challenge.

Implications

The internet connects the world, shortening the distance between people by the rapid transfer of information. Computers have shrunk to the size of a palm, and in the information society, most people can surf the internet anytime and anywhere. During the last few decades, many economic activities and startups have flourished with the benefit of the internet. These organizations provide as much information as we can imagine for free or very cheaply. Much knowledge and information are open source and can enhance our abilities or interfere with our decision-making based on the way we use it. As more and more well-designed open-source generative language models become available, large amounts of unverified information may shortly be packaged by bots as attractive news on the web. Sometimes bots are designed for a specific issue [13] and might have malicious intent. The growth of biased, intentional, or extremist public opinion in the news is sometimes difficult to detect, but it potentially impacts our thinking [25,26]. Understanding the potential media framing is a vital personal ability in the internet age of massive information floods.

Some online resources are available for fact-checking [26], providing the public with access to media literacy. While the Islander system cannot directly detect false information, it can monitor the media in real time and provide objective scores. These scores help us think critically; identify the opinions, roles, and goals of the media; and determine whether an item of information is credible. The news analysis systems work like an attenuated vaccine, reducing the toxicity of malicious information, increasing our immunity to misinformation, and preventing the spread of fake news. Future work on this issue should focus on providing a progressively more robust information judgment system that can grow with fake news generators even under adversarial attacks.

Limitations

One limitation of this study is the lack of detailed demographic information about vaccination recipients, as a result of which we could not investigate further factors that influence vaccination decisions. The scope of the study was to investigate the relationship between digital news and vaccination decisions, and some demographic characteristics that may be relevant for accessing web news. The lack of such detailed information makes it challenging to explore consumer engagement with digital media. Another limitation is that this study was conducted in an Asian society, and the news judgment system is only applicable to Chinese news, which makes it difficult to adapt the results and web applications to another region or society. Nevertheless, in recent years, dubiousness in digital news has become an important global issue, and the results of this study revealed its implications for vaccination in Asian societies. In future works, such news analysis systems may be established in different regions to help enhance the media literacy of the public, while collecting news data in different areas and conducting extended analyses.

Conclusions

In this study, we retrospectively analyzed an Asian society of 23 million people, using deep learning NLP methods to analyze 0.7 million digital news items over a half-year period, and identified a correlation between the percentage of fake digital news and COVID-19 vaccination doses. A higher prevalence of fake news had a significantly more adverse effect on vaccination decisions. Public health policy efforts to increase vaccination coverage might focus on reducing the impact of fake news on the public, and the use of news analysis systems may help to improve the public's media literacy.

Acknowledgments

We would like to thank Japan, the United States, Poland, Lithuania, Slovakia, and Czech Republic for donating vaccines to Taiwan. We would also like to thank the Ministry of Science and Technology, in Taiwan, for financially supporting this research (grant MOST 110-2634-F-002 -046, and MOST 110-2634-F-002 -034).

Conflicts of Interest

The Islander news analysis system is a free web service from Taiwan AI Labs.

Multimedia Appendix 1
Meaning and English translation of search keywords, and information on digital media sources.
[PDF File (Adobe PDF File), 195 KB - jmir_v24i4e36830_app1.pdf]

Multimedia Appendix 2
Distribution of scores. The left side is the suspicion score distribution; the dotted line indicates a Poisson distribution. On the right is the incitement score distribution; the dotted line represents a Gaussian distribution.
[PNG File , 51 KB - jmir_v24i4e36830_app2.png]

Multimedia Appendix 3
Percentage trend of suspicious news in some media sources.
[PNG File , 107 KB - jmir_v24i4e36830_app3.png]

References

1. The Matrix (1999). IMDb. URL: https://www.imdb.com/title/tt0133093/ [accessed 2022-04-13]
2. Harmsen I, Ruiter R, Paulussen T, Mollema L, Kok G, de Melker HE. Factors that influence vaccination decision-making by parents who visit an anthroposophical child welfare center: a focus group study. Adv Prev Med 2012;2012:175694 [FREE Full text] [doi: 10.1155/2012/175694] [Medline: 23209917]
3. Hoogink J, Verelst F, Kessels R, van Hoek AJ, Timen A, Willem L, et al. Preferential differences in vaccination decision-making for oneself or one's child in The Netherlands: a discrete choice experiment. BMC Public Health 2020 Jun 01;20(1):828 [FREE Full text] [doi: 10.1186/s12889-020-08844-w] [Medline: 32487041]
4. Courbage C, Peter R. On the effect of uncertainty on personal vaccination decisions. SSRN :1-12 Preprint posted online Apr 20, 2021. [doi: 10.2139/ssrn.3821166]

5. Fridman A, Gershon R, Gneezy A. COVID-19 and vaccine hesitancy: a longitudinal study. PLoS One 2021 Apr 16;16(4):e0250123 [FREE Full text] [doi: 10.1371/journal.pone.0250123] [Medline: 33861765]
6. Pullan S, Dey M. Vaccine hesitancy and anti-vaccination in the time of COVID-19: a Google Trends analysis. Vaccine 2021 Apr 01;39(14):1877-1881 [FREE Full text] [doi: 10.1016/j.vaccine.2021.03.019] [Medline: 33715904]
7. Porter E, Wood TJ, Bahador B. Can presidential misinformation on climate change be corrected? Evidence from internet and phone experiments. Res Politics 2019 Aug 07;6(3). [doi: 10.1177/2053168019864784]
8. Polak M. The misinformation effect in financial markets: an emerging issue in behavioural finance. Financ Internet Quart 2012;8(3):55-61 [FREE Full text]
9. Cook J, Ecker U, Lewandowsky S. Misinformation and how to correct it. Wiley Online Library 2015 May 15:e2015-e2017. [doi: 10.1002/9781118900772.etrds0222]
10. Berriche M, Altay S. Internet users engage more with phatic posts than with health misinformation on Facebook. Palgrave Commun 2020 Apr 28;6(1):71. [doi: 10.1057/s41599-020-0452-1]
11. Endsley MR. Combating information attacks in the age of the internet: new challenges for cognitive engineering. Hum Factors 2018 Dec 30;60(8):1081-1094. [doi: 10.1177/0018720818807357] [Medline: 30376429]
12. Shu K, Sliva A, Wang S, Tang J, Liu H. Fake news detection on social media: a data mining perspective. SIGKDD Explor Newsl 2017 Sep;19(1):22-36. [doi: 10.1145/3137597.3137600]
13. Lazer DMJ, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, et al. The science of fake news. Science 2018 Dec 09;359(6380):1094-1096. [doi: 10.1126/science.aao2998] [Medline: 29590025]
14. Ghenai A. Health misinformation in search and social media. In: Proceedings of the 2017 International Conference on Digital Health. 2017 Presented at: DH '17; July 2-5; New York, NY p. 235-236. [doi: 10.1145/3079452.3079483]
15. MacDonald NE. Fake news and science denier attacks on vaccines. What can you do? Can Commun Dis Rep 2020 Nov 05;46(1112):432-435 [FREE Full text] [doi: 10.14745/ccdr.v46i1112a11] [Medline: 33447164]
16. World Health Organization. URL: https://www.who.int/ [accessed 2022-04-13]
17. Ministry of Health and Labor. URL: https://www.mohw.gov.tw/mp-1.html [accessed 2022-04-13]
18. Mills MC, Sivelä J. Should spreading anti-vaccine misinformation be criminalised? BMJ 2021 Feb 17;372:n272 [FREE Full text] [doi: 10.1136/bmj.n272] [Medline: 33597153]
19. Garett R, Young S. Online misinformation and vaccine hesitancy. Transl Behav Med 2021 Dec 14;11(12):2194-2199 [FREE Full text] [doi: 10.1093/tbm/ibab128] [Medline: 34529080]
20. Roozenbeek J, Schneider CR, Dryhurst S, Kerr J, Freeman ALJ, Recchia G, et al. Susceptibility to misinformation about COVID-19 around the world. R Soc Open Sci 2020 Oct 14;7(10):201199 [FREE Full text] [doi: 10.1098/rsos.201199] [Medline: 33204475]
21. Montagni I, Ouazzani-Touhami K, Mebarki A, Texier N, Schück S, Tzourio C, CONFINS group. Acceptance of a Covid-19 vaccine is associated with ability to detect fake news and health literacy. J Public Health (Oxf) 2021 Dec 10;43(4):695-702 [FREE Full text] [doi: 10.1093/pubmed/fdab028] [Medline: 33693905]
22. Loomba S, de Figueiredo A, Piatek SJ, de Graaf K, Larson HJ. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. Nat Hum Behav 2021 Mar 05;5(3):337-348. [doi: 10.1038/s41562-021-01056-1] [Medline: 33547453]
23. Bonnevie E, Gallegos-Jeffrey A, Goldbarg J, Byrd B, Smyser J. Quantifying the rise of vaccine opposition on Twitter during the COVID-19 pandemic. J Healthc Commun 2020 Dec 15;14(1):12-19. [doi: 10.1080/17538068.2020.1858222]
24. King KK, Wang B. Diffusion of real versus misinformation during a crisis event: a big data-driven approach. Int J Inf Manage 2021 Jul:102390. [doi: 10.1016/j.ijinfomgt.2021.102390]
25. Saling LL, Mallal D, Scholer F, Skelton R, Spina D. No one is immune to misinformation: an investigation of misinformation sharing by subscribers to a fact-checking newsletter. PLoS One 2021 Aug 10;16(8):e0255702 [FREE Full text] [doi: 10.1371/journal.pone.0255702] [Medline: 34375356]
26. Zhang X, Ghorbani AA. An overview of online fake news: characterization, detection, and discussion. Inf Process Manag 2020 Mar;57(2):102025. [doi: 10.1016/j.ipm.2019.03.004]
27. Zhou X, Zafarani R. A survey of fake news: fundamental theories, detection methods, and opportunities. ACM Comput Surv 2020 Oct 15;53(5):1-40. [doi: 10.1145/3395046]
28. Islander. URL: https://islander.cc/about [accessed 2020-01-01]
29. Zafarani R, Zhou X, Shu K, Liu H. Fake news research: theories, detection strategies, and open problems. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019 Presented at: KDD '19; Aug 4-8, 2019; Anchorage, AK p. 3207-3208. [doi: 10.1145/3292500.3332287]
30. Collins B, Hoang DT, Nguyen NT, Hwang D. Trends in combating fake news on social media – a survey. J Inf Syst Telecommun 2020 Nov 27;5(2):247-266. [doi: 10.1080/24751839.2020.1847379]
31. Zhou X, Jain A, Phoha VV, Zafarani R. Fake news early detection. DTRAP 2020 Jun 30;1(2):1-25. [doi: 10.1145/3377478]
32. Bermes A. Information overload and fake news sharing: a transactional stress perspective exploring the mitigating role of consumers' resilience during COVID-19. J Retail Consum Serv 2021 Jul;61:102555. [doi: 10.1016/j.jretconser.2021.102555]
33. Apuke OD, Omar B. Fake news and COVID-19: modelling the predictors of fake news sharing among social media users. Telemat Inform 2021 Jan;56:101475 [FREE Full text] [doi: 10.1016/j.tele.2020.101475] [Medline: 34887612]

34. Allen J, Arechar AA, Pennycook G, Rand DG. Scaling up fact-checking using the wisdom of crowds. Sci Adv 2021 Sep 03;7(36):eabf4393 [FREE Full text] [doi: 10.1126/sciadv.abf4393] [Medline: 34516925]
35. Vieira L, Jeronimo C, Campelo C, Marinho L. Analysis of the subjectivity level in fake news fragments. In: Proceedings of the Brazilian Symposium on Multimedia and the Web. 2020 Presented at: WebMedia '20; Nov 30 - Dec 4, 2020; São Luís, Brazil p. 233-240. [doi: 10.1145/3428658.3430978]
36. Jeronimo C, Marinho L, Campelo C, Veloso A, da Costa Melo CMA. Fake news classification based on subjective language. In: Proceedings of the 21st International Conference on Information Integration and Web-based Applications & Services. 2019 Presented at: iiWAS2019; Dec 2-4; Munich, Germany p. 15-24. [doi: 10.1145/3366030.3366039]
37. Volkova S, Shaffer K, Jang J, Hodas N. Separating facts from fiction: linguistic models to classify suspicious and trusted news posts on twitter. 2017 Presented at: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers); July 2017; Vancouver, Canada p. 647-653. [doi: 10.18653/v1/p17-2102]
38. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D. RoBERTa: a robustly optimized bert pretraining approach. arXiv Preprint posted online on 19 Jul 2019 [FREE Full text]
39. Yu L, Lee L, Hao S, Wang J, He Y, Hu J. Building Chinese affective resources in valence-arousal dimensions. 2016 Presented at: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; June 2016; San Diego, CA p. 540-545. [doi: 10.18653/v1/n16-1066]
40. Qiu F, Cho J. Automatic identification of user interest for personalized search. In: Proceedings of the 15th International Conference on World Wide Web. 2006 Presented at: WWW '06; May 23-26; Edinburgh, Scotland p. 727-736. [doi: 10.1145/1135777.1135883]
41. Harb H, Khalifa A, Ishkewy H. Personal search engine based on user interests and modified page rank. 2009 Presented at: 2009 International Conference on Computer Engineering & Systems; Dec 14-16; Cairo, Egypt. [doi: 10.1109/icces.2009.5383228]
42. Feng C, Khan M, Rahman AU, Ahmad A. News recommendation systems-accomplishments, challenges and future directions. IEEE Access 2020;8:16702-16725. [doi: 10.1109/access.2020.2967792]
43. Zhu Z, Li D, Liang J, Liu G, Yu H. A dynamic personalized news recommendation system based on bap user profiling method. IEEE Access 2018;6:41068-41078. [doi: 10.1109/access.2018.2858564]
44. Liu S, Dong Y, Chai J. Research of personalized news recommendation system based on hybrid collaborative filtering algorithm. 2016 Presented at: 2nd IEEE International Conference on Computer and Communications (ICCC); Oct 14-17; Chengdu, China. [doi: 10.1109/CompComm.2016.7924826]
45. Jamnadass E, Aboumarzouk O, Kallidonis P, Emiliani E, Tailly T, Hruby S, et al. The role of social media and internet search engines in information provision and dissemination to patients with kidney stone disease: a systematic review from European association of urologists young academic urologists. J Endourol 2018 Aug;32(8):673-684. [doi: 10.1089/end.2018.0319] [Medline: 29926740]
46. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of google trends in health care research: a systematic review. PLoS One 2014 Oct;9(10):e109583 [FREE Full text] [doi: 10.1371/journal.pone.0109583] [Medline: 25337815]
47. Worldwide desktop market share of leading search engines from January 2010 to January 2022. Statista. 2022 Mar. URL: https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/ [accessed 2022-04-13]
48. Our World in Data. URL: https://ourworldindata.org/ [accessed 2022-01-01]
49. Jiang J, He D, Allan J. Searching, browsing, and clicking in a search session: changes in user behavior by task and over time. In: Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval. 2014 Presented at: SIGIR '14; July 6-11; Gold Coast, Queensland, Australia. [doi: 10.1145/2600428.2609633]

Abbreviations

CVAT: Chinese valence-arousal text data set
NLP: natural language processing

Edited by M Gisondi; submitted 27.01.22; peer-reviewed by X Zhou, A Chang, W Ceron; comments to author 17.02.22; revised version received 02.03.22; accepted 04.04.22; published 26.04.22.

Please cite as:
Chen YP, Chen YY, Yang KC, Lai F, Huang CH, Chen YN, Tu YC
The Prevalence and Impact of Fake News on COVID-19 Vaccination in Taiwan: Retrospective Study of Digital Media
J Med Internet Res 2022;24(4):e36830
URL: https://www.jmir.org/2022/4/e36830
doi:10.2196/36830
PMID:35380546

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org